2019

Computational model of negotiation skills in virtual artificial agents

Pacella, Daniela

Computational model of negotiation skills in virtual artificial agents

Daniela Pacella

School of Computing, Electronics and Mathematics

University of Plymouth

PhD

September 2018

Acknowledgements

FIRST and foremost, I wish to express my gratitude to my supervisor Prof. Davide Marocco, for his guidance, support, patience and dedication. Thanks to his knowledge and experience he allowed me, through hard and less hard lessons, to evolve, be passionate about research and to learn how to be responsible, how to improve my skills, and, most importantly, how to make the best decisions for my future.

I also wish to thank my second supervisor Prof. Steve Furnell, for his support throughout the PhD programme and his insightful comments that allowed a great improvement of the results and overall quality of the present work.

Another huge thanks goes to Dr. Elena Dell'Aquila, for her great experience and competence, but also for her incredible kindness and help. It was a pleasure to work with her and I will always be grateful for her academic and personal support.

I also wish to thank all the people behind and within the ENACT Project which I had the chance and the pleasure to work with during these years. In particular, a very special thanks goes to Prof. Orazio Miglino, for sharing his outstanding knowledge, creativity and inspiration and for his infinite support.

A big thanks to all those who supported me and helped me improve my skills especially during my visit at the CNR in Rome, in particular Dr. Andrea Di Ferdinando, Dr. Massimiliano Schembri and Dr. Raffaele di Fuccio.

Also a huge thanks goes to all the people from the NAC Laboratory at the University of Naples Federico II, and in particular to Prof. Onofrio Gigliotta, Dr. Michela Ponticorvo and Franco Rubinacci, for their special and sincere support, guidance and comments.

Finally, I want to express gratitude to my family, who supported me throughout the years and never left me alone despite the distance, and to my boyfriend, who since the beginning of this PhD programme has witnessed a substantial increase of my average stress level and yet managed to make me smile every single day.

Abstract

NEGOTIATION skills represent crucial abilities for engaging in effective social interactions in formal and informal settings. Serious games, intelligent systems and virtual agents can provide solid tools upon which one-to-one training and assessment can be reliably made available. The aim of the present work is to fill the gap between the recent growing interest towards soft skills, and the lack of a robust and modern methodology for supporting their investigation. A computational model for the development of Enact, a 3D virtual intelligent platform for training and testing negotiation skills, will be presented. The serious game allows users to interact with simulated peers in scenarios depicting daily life situations and receive a psychological assessment and adaptive training reflecting their negotiation abilities. To pursue this goal, this work has gone through different research stages, each with a unique methodology, results and discussion described in its specific section. In the first phase, the platform was designed to operationalize the examined negotiation theory, developed and assessed. The negotiation styles considered, consistently with previous findings, have been found not to correlate with personality traits, coping strategies and perceived self-efficacy. The serious game has been widely tested for its usability and underwent two development and release stages aimed at improving its accuracy, usability and likeability. The variables measured by the platform have been found to predict in all cases at least two of the negotiation styles considered. Concerning the user feedback, the game has been judged as useful, more pleasant than the traditional test, and the perceived time spent on the game resulted significantly lower than the real time spent. In the second stage of this research, the game scenarios were used to collect a dataset of documents containing natural language negotiations between users and the virtual agents. The dataset was used to assess the correlations between the personal pronouns' use and the negotiation styles. Results showed that more engaged styles generally used pronouns with a significantly higher frequency than less engaged styles. Styles with a high concern for self showed a higher frequency of singular personal pronouns while styles with a high concern for others used significantly more relational pronouns. The corpus of documents was also used to perform multiclass classification on the negotiation styles using machine learning. Both linear (SVM) and non-linear models (MNB, CNN) performed reliably with a state-of-the-art accuracy.

Authors declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award. Work submitted for this research degree at Plymouth University has not formed part of any other degree either at Plymouth University or at another establishment. Relevant scientific seminars and conferences were regularly attended at which work was often presented. Three papers have been accepted for publication.

Word count for the main body of this thesis: **59134**

Signed: _____

Date: _____

Papers and conference presentations:

Pacella, D., Di Ferdinando, A., Dell'Aquila, E., and Marocco, D. (2015). *Online assessment of negotiation skills through 3D role play simulation.* In Artificial Intelligence in Education Conference proceedings, Madrid. Vol 9112, (pp. 921-922), Springer International Publishing.

Marocco, D., Pacella, D., Dell'Aquila, E., and Di Ferdinando, A. (2015). *Grounding serious game design on scientific findings: the case of ENACT on soft skills training and assessment.* In Design for Teaching and Learning in a Networked World (pp. 441-446). Springer, Cham.

Pacella, D., Dell'Aquila, E., Marocco, D., and Furnell, S. (2017, August). *Toward an Automatic Classification of Negotiation Styles using Natural Language Processing.* In International Conference on Intelligent Virtual Agents (pp. 339-342). Springer, Cham.

Contents

List of Figures

Chapter 1

Introduction

Serious games and intelligent virtual platform represent reliable tools for learning, training and assessing hard competencies. The field of soft skills, an ill-defined domain toward which the interest is steadily growing, currently lacks a standard methodology to provide e-learning and virtual training grounded on cognitive and psychological theories. This study revolves around the general aim of designing an innovative methodology for the implementation of a serious game to train and test negotiation skills, viewed as the interpersonal ability to handle conflict. In this chapter, the general scope of this thesis is presented, including the methodological and theoretical contributions that it provides to the current state of knowledge. In the last section, an overview of the chapters of the thesis is outlined, summarising the steps that were undertaken and the key results reached at the end of all research stages.

THE scope of this thesis is to introduce a new computational model and methodology for the development of a mobile, free and accessible software for training and testing negotiation skills that provides the user with a flexible, adaptive and intelligent experience relating with real life situations and based on 3D games and role-play simulations. In particular the focus of the work lies in the investigation and design of a system that revolves around the players, adapts to them, provides many means of interactions - among which the possibility of using natural language for the user assessment - and constantly learns from the collected data. In particular, the main objectives of the thesis are:

- to provide a novel approach to standardised psychological assessment that makes use of e-learning technologies and gamification principles;
- to integrate teaching and assessment into a single, mobile application that provides personalised training and an adaptive environment;

- to ground soft skill assessment and training, in particular negotiation skills, on validated psychological theories;
- to investigate the advantages of using a traditional paper and pencil test methodology against technology-enhanced approaches for psychological assessment;
- to create a natural language dataset of negotiation sentences collected through the game platform that can be used by the research community;
- to analyse the dataset of negotiation sentences using a text mining approach and research the differences among the use and distribution of words across the negotiation profiles;
- to propose a state-of-the-art machine learning approach to the assessment of negotiation and conflict management skills on the basis of natural language and is comparable to paper and pencil psychological tests.

To pursue these objectives, this work has gone through different stages, each of which aimed at accomplishing a special research goal and develop new features for the software, test them and evaluate the outcomes on the field. This research also partly contributes to the European Project ENACT (Enhancing Negotiation skills through online Assessment of Competencies and interactive mobile Training) funded by EACEA under the Lifelong Learning Programme, and its initial stage represents a collaboration on the involvement with the design, development, validation and improvement of the final release of the Enact platform, conducted in cooperation with the other project partners. This work continues beyond the boundaries of the project in modelling the different aspects and features that aim at enhancing the user experience and the naturalness of the interaction between the human and the virtual agents in the platform. The assessment environment of the software was also developed in collaboration with the other partners of the ENACT project, whose principal research objective was to prove the reliability and the validity of the evaluation provided to the users in comparison with standard psychological tests for negotiation skills, in particular the Rahim Organiza-

tional Conflict Inventory II (ROCI II). The ROCI II is a test to assess negotiation and conflict management skills consisting of 28 items that allows the categorization of the users into 5 styles, as according to Rahim and Bonoma's theory (Integrating, Compromising, Avoiding, Dominating, Obliging) (Rahim 1983, Afzalur Rahim et al. 2001). While a mutual cooperation with the other ENACT partners (Aidvanced, Italy; Mondo Digitale, Italy; University of Naples Federico II, Italy; Fundetec, Spain; The Turkish Ministry of Youth and Sport, Turkey) was agreed for testing and for the achievement of this validation, during and after the project's conclusion this research goes on in its attempt to enhance and expand the features of the software. For clarity and in order to stress the difference between the two aspects, the ENACT Project will be referred to with capital letters, while the software object of this work and whose assessment environment is developed within the scope the Project will be referred to as "Enact", with small letters. The beginning steps of the present research, as the Training Need Analysis, the research conducted upon the user interface and interactivity of the game platform, the implementation of profile adaptive 3D avatars and the first validation of the platform, have been conducted in the framework of the ENACT Project and in collaboration with project partners. The research upon the intelligent tutoring system, the comparison between the platform and the pencil and paper psychological test, the second validation of the platform, the implicit and explicit pleasantness evaluation of the platform and the studies involving natural language sentences, text mining and machine learning for natural language processing applied negotiation skill theories have been conducted as independent personal work. Additional details about the work conducted within and outside the scope of the ENACT project can be found in Appendix L.

1.1 Research focus

1.1.1 Scope of the thesis

The general aim of this research is to develop a modern technology for psychological assessment and training in the emerging field of soft skills. The project focuses on a fundamental daily life soft skill, particularly requested in the job market, that concerns

the way individuals manage interpersonal conflicts, with a special focus for organizations. Traditionally, such negotiation abilities are evaluated with the use of pencil and paper psychological tests. This kind of evaluation, however, relies on self-report and declarative information provided by individuals. The serious game Enact, by contrast, proposes to measure such styles using simulated scenarios that are able to enact conflicts during the interaction itself rather than simply asking direct questions to assess the memory of the subject's last conflicting situation - as in the case of Rahim's Inventory. The former solution, in fact, presents several main advantages:

- it allows to test the users during their actions rather than perceptions (subjects are requested to act, not to declare how they acted);
- it does not rely on the user memory and therefore does not suffer from possible recall biases;
- considering its gaming components, Enact represents a more engaging and challenging assessment environment;
- it allows an automated and immediate classification that does not involve the intervention of an expert;
- it allows to provide an adaptive and tailored training environment based on the outcome of the assessment;
- it is highly portable and free to use.

The current thesis has the aim to present the steps undertaken to develop the Enact game platform based on Rahim's theory of negotiation, validate it as a reliable assessment tool against the Rahim's test and other relevant psychological tests, evaluate its usability and user-friendliness, evaluate its implicit and explicit pleasantness and, lastly, introduce a natural language analysis of the utterances of the users to allow an automatic categorization. These steps will be summarised in the following subsection.

1.1.2 Methodological and experimental framework

The development and extensive test of the Enact software represents an ambitious attempt to go beyond the way we usually conceive videogames, psychological assessment and training. In particular, the research focuses on the importance of possessing a confident level of soft skills in order to better cope with social and environmental demands. School first and academic education later on, provide students with the so called hard skills: a set of cognitive skills that can be learnt in a rigorous manner and methodology and are not subject to situational and individual changes. Mathematics, biology, literature, geography, computer programming, and the like, all constitute examples of hard skills. They well describe exactly what we know or what we are supposed to know about them. On the other hand soft skills are described as non cognitive skills that do not depend exclusively on acquired knowledge, at least on solely declarative one. Soft skills entail, among them, effective communication, critical thinking, problem solving, creativity and leadership. The details of the different stages of this research will be now briefly summarised.

1.1.2.1 Software development

In the first stage of this study, the research has been conducted in parallel with the framework of the ENACT Project, and therefore the main focus was brought on the implementation of the software interface and of the expert system behind the platform, which is mainly composed of two different environments: assessment and training. The principal research aim that drove this stage was the challenge to reproduce a psychological test and translate it into a 3D world, operationalizing each detail of Rahim's theory in sentences, animations, gestures and an attractive environment for the first time, so to create a novel methodology.

The graphic interface and the assessment environment has been developed in collaboration with the other project partners. In particular, my research contribution to the development of the platform and prior to the beginning of testing was to choose the suitable gestures for the virtual game characters, develop non-verbal aspects of the

game that could help deliver the tone of the simulated speech happening within the platform, to design the user interface of the introduction to the scenarios, to design the tutoring system, to develop the content within each scenario and to ensure all of the specified content met the psychological theory behind the game. In detail, the assessment environment introduces the user to a series of scenarios preceded by a description that includes useful details and information about the situation that the user is playing in. In each scenario, the user is asked to interact with virtual characters and try to reach a deal about specific interpersonal conflicts involving daily life relationships with partners, siblings or friends. The 3D characters are able to express a range of basic emotions in an effective way using verbal and non-verbal indicators. The answers chosen by the user from a predefined list are analysed and all the variables gathered during the interaction are evaluated in order to provide a customised profile. An Intelligent Tutoring System, integrated in the second and following stage, was also able, as explained below, to evaluate the history of the choices of the participant and suggest possible strategies to improve or raise awareness on the style enacted. This introduces a pedagogical aspect in the platform that allows to deliver important information to the participant throughout a tailored training.

1.1.2.2 Software validation

The second, subsequent stage aimed at exploring the validity of the newly developed videogame software against the negotiation test provided by Rahim and Bonoma and investigate the possibility to find correlations between Rahim's Inventory, the Enact platform and possibly other psychological tests. The main hypothesis was to prove the possibility to map the scores obtained by participants during the interaction with the first released version of the Enact game with the negotiation styles assessment provided by Rahim's Inventory. Experiments were carried out to validate the platform as an alternative tool to investigate the negotiation skills as provided by the model by Rahim and Bonoma. The results indicated the first directions to follow for an improvement of the platform, and, after the modifications and further development stages, the platform

underwent a second round of validity assessment. The profiles of the subjects which undertake the two different types of assessment tools for negotiation (Enact platform and the paper questionnaire by Rahim, ROCI II) were compared and data was extensively analysed to outline the differences between the groups. The platform was then modified in its length, interface and content according to the results, integrating the findings that showed the importance of including a unique model of the Compromising profile - that was previously modelled as a combination of the other four profiles, the need of shortening the number and length of the scenarios according to the reliability obtained in the validation and the need to remodel the aspect and graphics of the virtual agents.

The model of the training environment is also developed and improved in the framework of this research and independently, and is designed for practising negotiation skills and be guided throughout the scenarios with the help of an intelligent tutoring system. In this context, intelligent tutoring system refers to an adaptive system that makes use of technology and algorithms to personalise the experience of learning according to the learner's characteristics (Sleeman & Brown 1982, Jr et al. 2013). In the training session, the user is introduced to a series of situations and, at the end of each scenario, the user is presented with the full history of their choices along with a full and detailed explanation of their negotiation profile, attitude and the goals they reached. The interface also allows the user to navigate through the past interactions with the artificial agent, in order to provide both an overview and a specific analysis of the actions taken by the two agents and give a constructive view of the negotiation process to help the user gain awareness of their own styles. The feedback is based on an Intelligent Tutoring System that has been developed and integrated in the training environment.

1.1.2.3 Software evaluation and pleasantness

All of the features of the Enact platform, while going through different cycles of development, have been extensively tested for their usability. Different experiments were carried out and the results were used to improve the design of the software and evolve

on the learners' need. The first released version was tested with a sample of participants in order to receive feedback to improve the interface. The second release of the platform was also tested for its usability and user-friendliness in a different but comparable group sample using the same self-report Likert-scale questionnaire. Other than that, the final version of the game was compared with Rahim's Inventory in terms of implicit pleasantness, operationalized as the perception of time spent to complete both, and explicit pleasantness, throughout the use of a questionnaire that was different and more detailed than the ones used solely for usability evaluation purposes. The pleasantness was tested between two groups of participants, those who had a background in psychology (either students or graduates) and those who were students or graduates in other disciplines.

1.1.2.4 Natural language analysis and categorization

The last stage of this work concerned the attempt to enhance the features of the platform and remove its limitations, leaving the user more free to explore the real life based scenarios and allowing a more natural approach with the virtual agents. In this sense, a relevant constraint was imposed by the presence of multiple choices to interact with the bots, therefore an artificial intelligence system that allows the recognition of natural sentences was developed. The main hypothesis that drove the experiments was that there were distinct features in the words usage between individuals categorised as each of the five styles theorised by Rahim and Bonoma's model. To investigate this aspect, an extensive dataset containing natural sentences occurring in dyadic conversations between the virtual agents and participants was created by leaving the participants free to write their own reply in the scenarios. An analysis of the usage of the pronouns and an attempt to map the words' frequencies and relations to Rahim Inventory scores of each participant was conducted on the dataset. Other than that, to allow an automatic classification, the content of the sentences was transformed into vectors in a multidimensional space created by all the words chosen during data collection, used to train deep neural networks and classified into a negotiation profile, i.e. the

unique combination of the five dimensions - styles - of the ROCI II obtained by each participant.

Further experiments included a comparison between different machine learning algorithms in order to find the one that maximised the fit between the sentences and the corresponding style of handling interpersonal conflict. Analyses show consistent correlations between styles and words usage, and light the path to a comprehensive interpretation of the behavioural and linguistic cues of different negotiation styles.

1.1.3 Main findings

This work encompasses the investigation of many different aspects of the development and testing of the Enact software. The first theoretical research was conducted in order to determine the best theoretical framework that could be a sustainable and reliable measurement throughout the future comparisons and assessment conducted on the platform. Considering that soft skills, intangible and opposed to the concept of "hard skills", are extremely difficult to uniquely define, in literature it is possible to find several definition of negotiation, whose borders overlap across different disciplines, such as economics, psychology, sociology or cognitive science. Rahim and Bonoma's theory, as will be shown in the next chapter, allows to define several negotiation styles providing a formal multidimensional space that can be integrated and implemented in a computational and mathematical model. The theory is also supported by a robust and widely used psychological test, the ROCI II, that could be used to provide a practical mean of comparison for the assessment provided by the platform. When the first version of the Enact software was released, it was recognised as the first serious game and simulation software for negotiation in literature to be grounded on a psychological theory.

The first steps toward the validation and evaluation of the reliability of the game included the recruitment of participants by all of the ENACT project partners in order to control the robustness of the architecture. The results obtained supported the following hypotheses:

- There is a current lack of psychologically grounded technology and methodologies to teach and assess soft skills, in particular negotiation skills viewed as daily conflict management;
- The ROCI II is proved again to constitute a reliable tool for assessing negotiation skills; in fact, a factorial analysis conducted on the test items confirmed the five styles identified by Rahim, and Cronbach's alpha is high in all the samples considered.;
- Enact's gameplay can substitute the ROCI test for the investigation of at least two of the identified negotiation styles, and can help provide a confident assessment of the other styles with the investigation of specific scenario-based features;
- Enact user interface is clear, intuitive and addresses the needs of the player; the usability, rated with a questionnaire on a Likert scale, has been measured as high;
- There is a gender difference in the distribution of negotiation styles and in the attitude towards the virtual agents' negotiation styles;
- The administration order of Enact and ROCI II does not affect the behaviour and results obtained to the tests;
- There are country-specific differences in the distribution of negotiation styles.

As soon as the validation experiments on the software demonstrated these results, further studies aimed at improving the structure of the game were designed. Keeping in mind that Enact was designed to fit the needs of adults, adolescents and different international cultures other than the English one, such as Italian, Turkish and Spanish, it was important to evaluate if the results and the usability of the software were applicable to all possible users of the platform. Other than that, in order to promote the integration of a definition of negotiation that is situational, teachable, subject to evolution, improvement and previous knowledge, Enact was organized in scenarios that depict real life

situations that can allow a more accurate detection of some negotiation style features that could appear to be stable across different contexts.

However, even if the the realistic factor has to keep its weight on the design of the scenarios to provide a comprehensive and inclusive experience, sometimes scenarios that are too long or hard for users to relate with could decrease the motivation to play as they emerge to be too different from the user playstyle and lifestyle. Since scenarios were standardized to be played by all users, and may include hypothetical situations as partner relationships or context-specific environment, sometimes users are not able to be correctly identified with the character playing their role, and if this fundamental gamification aspect is lost, the results lose their accuracy. In particular, a preliminary experiment investigated how British college students and adults judged the naturalness of the emotions showed by the agents, how realistic were the agents' behaviours, the usefulness of the interface, how intuitively structured were the controls, the graphics and video quality, how interesting they perceived the information provided, and if they would play the game again. The results were used to provide a direction of the improvement of the interface and the design of the scenarios. The same questions were asked after the release of the second version of the platform, and the results were compared, this time with a broader sample. After the final version of the software was considered stable and definitive, a study was conducted to compare the pleasantness of the interaction with the Enact software and with the ROCI II psychological test. The explicit pleasantness was judged using a questionnaire with several sections, the implicit pleasantness was measured using the concept of perceived time spent on both tasks. The results can be summarized as follows:

- About 95% of participants stated that they would play the game again in both conditions, regardless of age, gender and version of Enact software;
- Videogame graphics quality in the first release was judged as poorer than standard modern videogames, but this data improved in the second version;
- The animations and situations described within the scenarios are generally con-

sidered easy to understand and appropriate;

- The Enact game controls are intuitive and the instructions are clear;
- Information provided in Enact is useful, interesting, and the vast majority of the sample declared that the information contained in the tutoring environment would have been useful in future daily life situations;
- Enact was in general judged as shorter than its real length in terms of minutes spent to complete it, even if the amount is about double than the time needed to complete the ROCI-II;
- Enact is considered in general more engaging and more pleasant to complete than the ROCI-II.

Once the final version of Enact was released, and after the end of the European Project ENACT, the focus shifted toward creating new means of interactions with the bot, improving the scenarios, their explorability and accuracy using machine learning and natural language processing. Experiments aimed at collecting natural language sentences occurring between the participants and virtual agents were designed, using screenshots and images taken from the Enact game platform, in particular the introductions, the variables, the characters and the dialogues of the scenarios. The experiments were carried out extensively, including more than four hundred English speaking participants, for a dataset of over four thousand sentences. All participants, also, were asked to complete the ROCI II Inventory. The sentences were preprocessed following several steps, including grammar, syntactical and semantic corrections, punctuation and stopwords removal and stemming. The analysis of the word distribution, usage and frequencies revealed the following results:

- The similarity in the use of the personal pronouns is consistent with the similarity across different dimensions of the negotiation styles described by Rahim;

- Integrating and Compromising style display a frequency in the use of both the pronouns "You" and "I" and relational pronouns in a way that is similar but different from all other styles;
- The document obtained by merging all sentences associated to a participant who scored as most predominantly Obliging is the one that proves to be the most different and heterogeneous from the others;
- Dominating has been found to use more second person pronouns and significantly less relational pronouns, while Avoiding and Obliging were the styles that used the lowest amount of pronouns;
- in general it is possible to highlight specific differences between the pronouns usage - along with their combinations - and the score obtained by ROCI II, including single style dimensions and the predominant negotiation trait.

Finally, data obtained during these experiments were used to train machine learning algorithms for feature extraction and automatic classification of the sentences into the negotiation styles associated with the participants. This last stage of the present work aimed at creating an intelligent system to obtain an accurate classification of the negotiation styles using natural language sentences. Several models were tested, among which Support Vector Machine, Naive Bayes Classifier and Convolutional Neural Network. The words were encoded into the models both as one hot vectors and as dense vector representations obtained training Google Word2Vec network (Mikolov, Chen, Corrado & Dean 2013). The results obtained can be summarised as below:

- All models reached an accuracy measured by F score on 5 classes significantly higher than chance (set at 0.2), with the minimum accuracy reached being 0.32;
- All models' accuracy improved significantly by taking into account in the dataset exclusively participants who reached above a specific threshold in their predominant negotiation style score as measured by the ROCI II. This aspect, in particular, also provides more validity to the ROCI II model and to Enact assessment;

- All models considered reached a state-of-the-art accuracy and in particular the use of pre-trained word embeddings resulted in a higher accuracy than one-hot input vectors.

1.2 Contributions to knowledge

This research, inside and outside of the boundaries of the ENACT Project, aims at developing a simulated environment where users can learn, exercise and be assessed about their level of soft skills and in particular their knowledge and application of Rahim's model styles of handling conflict. The Enact game software, developed and enhanced during the course of this study, includes an assessment and a training environment and takes into account all the aspects to increase the user's participation and engagement (interacting with their intrinsic and extrinsic motivation), the latest researches' outcomes in literature about Intelligent tutoring systems, and a Natural Language Processing system to allow a realistic interaction with the computer-controlled avatar. It is thought to be as accessible and portable as possible. The development and testing of this platform is studied in order to increase its usability, effectiveness, reliability, distribution and comfort.

1.2.1 Theoretical contribution to knowledge

In details, this study aims to contribute to the current state of theories and speculative knowledge in the following ways:

1. Integrating in the Enact's tutoring system and user model the techniques and advances described in the recent years as part of the literature about ITSs. In particular, the gap that the present work aims to fill is to translate the techniques which were considered as exclusive features of well-defined domains to the field of soft skills, a multitude of concepts and definitions which can hardly be referred to in a unique way. In detail, what this work proposes to achieve is the definition of a common and agreed framework for the development of intelligent tutoring systems for ill-defined domains, including features about the characterization of

the virtual agents, facing unexpected situations, dealing with overlapped contents and definitions, and learning from the experience with users. In this direction it is important to adopt data analysis techniques from psychometrics to cloud computing and machine learning.

2. Providing validity to Rahim and Bonoma's theory of negotiation skills seen as interpersonal conflict - in particular between peers - by analysing wide datasets from culturally and demographically diverse regions and demonstrating consistent results. In particular, ROCI II test has shown its construct validity and its robustness in correlation with each of the negotiation styles specified by the theoretical framework, across all groups, nationalities, age and gender.

3. Showing the effectiveness of serious games as possible psychometric tools for assessment and training of soft skills, with a solid psychological theory as background. Many of the dimensions of the negotiation styles theorized by Rahim and Bonoma, in fact, overlap with the dimensions measured through the Enact game platform, and these indicators in most cases can uniquely refer to a specific style, guaranteeing an accurate assessment, whereas other tests which explored different constructs, such as self-efficacy, did not show any correlated dimensions. This once again represents a perspective change in the way psychological tests are conceived and an advance towards future investigation of either different soft skills, like leadership and problem-solving or more stable traits such as personality. Concerning soft skills which require social interactions, in fact, the use of simulations, role-playing and serious games could be extremely beneficial to improve the quality of assessment and to disambiguate between what the participants declare they would do - as in standard paper and pencil tests - and what their behaviour appears to be within the game.

4. Modelling a Natural Language Processing architecture to improve the means of communication and enhance the spontaneity of the interaction between the user and the virtual characters, along with allowing an automatic categorization and

feature detection of specific linguistic style during virtual-to-human negotiation, that could be potentially extended to human-to-human ones. In fact, this study wishes to address the problem of inferring the user's soft skills by analysing the factors contained in the sentences that compose their answer. In particular, the biggest aim is to include a natural language processing system in the user model, which has not been previously done nor reported in literature, especially concerning ill-defined domains. This means that, in order to interact with the virtual characters inside the Enact scenarios and receive, at the end of the game, a standardised psychological assessment based on the ROCI II test, it is possible for the user to write English sentences, with the only limitation of the number of characters, and a profile matching the linguistic characteristics of each of the negotiation styles as provided by Rahim and Bonoma's theory will be provided. First of all a dataset to use as training data for different classifiers was constructed, and this was a necessary theoretical and practical step, considering the lack of any negotiation dataset available in literature. Concerning the modelling of an automatic system to classify the sentences and the way it is included in the student model of the intelligent tutoring system and of the serious game platform, different architectures for NLP are compared, in order to give an insight about the efficiency of linear classifiers, such as Support Vector Machines and non-linear classifiers (like Naive Bayes Classifier) and artificial intelligence algorithms (specifically convolutional neural networks), along with descriptive and inferential statistic analysis. The development of such model, that has also proven to be quite effective in its results, represents an aim to reduce the gap in the reliability and validity of the final and improved Enact system and user model with respect to the ROCI II paper and pencil test, along with an attempt to provide a more immediate and tailored environment in which the participant can feel more free to explore and interact with less constraints.

1.2.2 Methodological contribution to knowledge

Along with the previously listed theoretical contributions, this work aims to present the results of the application of the models and theories explained. In order to provide an evidence of the efficacy of the methodology introduced, several experiments were conducted, which included hundreds of participants from different cultural and age groups. The methodological and practical contribution to knowledge achieved during this research project can be listed as below:

- The first advance is represented by the proposal of a computer science approach to the field of psychological assessment that includes models coming from the field of artificial intelligence, whereas the classical ways of profiling the level of soft skills of a user - be them a candidate, a student, a professional, or a group of employees - had to go through unnecessary time-consuming and expensive stages in which the presence of an expert was required to go through paper and pencil tests, which require time and expertise to be processed. The advantage of the automation of this assessment procedure is a key element, in that it allows immediate feedback, quick user adaptive responses, increased accessibility and affordability. This proposed methodological solution is also practical in terms of storing, analysing, comparing, shaping and maintaining the collected data. Thanks to the use of known cloud database services, user data loss is considered extremely improbable and there is a general higher accessibility and security compared to paper tests.

- Considering that the present work started within the framework of ENACT, it embraces since the beginning the project's general scope of giving role-play games a new assessment-related role in soft skill evaluation and training for their potential to increase motivation, engagement and help create a more immersive and realistic experience. For this reason Enact also includes elements strongly taken from the field of serious games and RPGs in general, with the aim of gamifying the user experience and collecting information about the players in a way that is

comparable to that obtained by the administration of the ROCI-II. For gamification it is meant the process of integrating game elements into non-game activities so as to increase user engagement and therefore, in this case, the effectiveness of any form of teaching or learning outcome. This setting allows the player to be immersed in a simulation and interact with virtual characters, which should also help relief the continuous feeling of being the subject of a standardized psychological test, although receiving the same evaluation. Other than that, while the ROCI-II items are based on the subject's ability to "recall the most recent conflict situation", as stated in the introductory description of the items of the inventory, and therefore rely on the subjective memory of someone's behaviour, within the Enact platform what it is measured is the actual behaviour in each specific situation, which cannot be biased by personal memory.

- Another interesting research aspect is represented by the integration in the Enact's training environment of intelligent training techniques in a standardised and reliable tool for the evaluation and practising of an ill-defined domain such as that of soft skills. In particular, the aim of the system is to provide a personalised, learner-centred and protected environment where the user can exercise different negotiation styles and gain awareness on the effectiveness of each among several types of situations. The possibility for the users to train themselves in the same context than the one where they were tested and the validated reliability of the platform allow a continuity between theoretical knowledge and application which can improve the user understanding of the process and effectively increase their negotiation abilities and interpersonal conflict management skills.

- Thanks to the large dataset created collecting the sentences which were produced within the interactions between participants and Enact's virtual agents, all data available about the linguistic components, features, indicators and frequency and its interpretation can be viewed as more robust and reliable than material collected via standard interviews. For this reason, the way it is included

in the user model within the platform can be seen as an novel methodology for the collection and analysis of user information and linguistic behaviour in serious games. Considering that the platform itself in all its releases and features has been extensively tested for its usability and localised in different countries, taking into account different age groups, participants' background, career, education, gender, and that the scenarios were developed and improved to fit de-localized real-life situations as much as possible, the experience and knowledge acquired throughout the Enact platform can be considered highly generalizable.

The direction of the present work aims at demonstrating that the release of the enhanced and complete Enact platform can add a fundamental dimension to the current use of technology and serious games in education, by grounding learning analytics and player's behaviours in shared frameworks that are accepted and recognised by psychologists and professional soft skills trainers.

1.3 Summary of the chapters

The structure of this thesis describes all of the stages, step by step, taken during the course of the project to design, develop, and release the Enact platform. The game represents a computational model and methodology for the investigation and assessment of virtual agents-to-human negotiation skills. The special focus of this platform is the promotion of social skills, soft skills competence and knowledge that can be applied to a broad range of contexts as well as human-to-human negotiations.

The focus of the work during the research was on the development of an adaptive system that integrates different means of interaction - verbal and non-verbal - as well as a descriptive and accurate user model. For this purpose, the work was defined by multiple research goals reached independently and in some cases in parallel, both inside and outside of the scope of the European Project ENACT, and with or without the cooperation of the project partners. Each of the chapters of the thesis will present the experiments made to find out the needs of different cultural and age groups in the field

of the current state of the art soft skills training and assessment tools, the literature researches to disambiguate, distinguish and provide a theoretical direction to the psychological framework of the Enact platform, assess the validity, reliability and efficacy of the platform compared to standard psychological tests, define the guidelines for the implementation of the adaptive intelligent tutoring system integrated in the platform, assess the interface, usability, realism and graphical appearance of the serious game compared to standard role-play games, assess the implicit and explicit pleasantness of the virtual agents, interactions, scenarios, and the whole experience with the platform, and finally create a natural language processing model of each negotiation style described in the theoretical framework, using the features evidenced by the results of a wide research to create an automated assessment and guarantee to the users an additional freedom in their action and response.

In particular, the following chapter will deal about the theoretical framework chosen for the implementation of the platform, starting from the concept of soft skills, their contrast with hard skills, their difference from other similar and overlapping constructs such as emotional intelligence and competence, the variety of psychological theories and definitions that arose around this concept and what are the state-of-the-art methodologies for effectively teaching and assessing soft skills. The chapter continues introducing the concept of serious game, seen as a videogame with educational purposes, and role-play games, useful tools especially for practising social abilities in safe simulated environments, and how the two methodologies are interconnected and mixed in online technology enhanced learning platforms, in particular those designed for ill-defined domains. Then, considering the special focus of Enact on negotiation skills, conflict management models are introduced along with the most recognized and up-to-date definitions. Finally, Rahim and Bonoma's 5-style theory is described, along with the motivations that made it the most reliable and currently used conflict handling model.

The third chapter deals about Intelligent Tutoring Systems, considered as an interdisciplinary research field that investigates on how to develop pedagogical technologies

that create educational content based on the need of the specific learner. The chapter goes on with a brief comprehensive literature review of the latest intelligent tutoring systems developed in the field of well and ill-defined domains and of what are the advantages of the different types of infrastructures and features that can be implemented in the expert and user modules in order for them to be more effective and specific to the teaching subject. Along with an analysis of the best arrangement of the modules within intelligent tutoring systems, an investigation of the variety of algorithms that have been used in literature as part of the expert module is described. Finally, the tutoring system structure of the Enact platform is explained.

The fourth chapter revolves around Enact software development and testing. In particular, the creation of the content of the scenarios, the interface, the virtual agents, the game parameters and online infrastructure is described, taking the results of the Training Need Analysis conducted by the ENACT project partners as principal research directions. The chapter also illustrates the experimental testing and results of the first version, on both the sides of the psychological validity in comparison with Rahim's standardized test and of the user experience and usability.

The fifth chapter deals about the main changes in the passage between the first and the second release of the platform and the differences and improvements in terms of graphics, contents, interface and user evaluation. It also illustrates the evaluation phases and a new validation phase conducted on the improved version of the platform. Finally, it explores one of the dimensions that can constitute a clear difference, from the user perspective, between a standard paper and pencil psychological test, as the ROCI-II, and Enact, a serious games based on role-play: the pleasantness. The chapter ends with the analysis of the advantages and disadvantages of the use of the two methodologies, in particular the experimental biases and the user experience. The hypothesis is that the interaction with Enact is explicitly and implicitly more pleasant for participants.

Finally, the sixth and the seventh chapters revolve around the design, implementation

and testing of a natural language processing model for negotiation sentences. In detail, the sixth chapter describes a research conducted to collect natural language sentences occurred between participants and Enact virtual agents, along with a statistical analysis of the collected dataset to allow linguistic features that distinguish each negotiation style to emerge. In particular, the aim is to highlight the most informative and relevant features of the sentences that can help recognize and classify negotiation skills abilities and behavioural styles. The seventh chapter, instead, focuses on the investigation of the most accurate machine learning model for the detection and classification of these descriptive linguistic cues.

Chapter 2

The importance of development and training of soft skills and negotiation

The definition of soft skills in literature is currently under debate. Soft skills are generally considered personal attributes that characterize interpersonal interactions and intrapersonal aspects in formal and informal environments. These abilities can also be assessed and enhanced throughout education and tutoring. Among these trainable competencies, negotiation skills, in their cooperative rather than competitive definition, are gaining importance across various disciplines such as psychology, law or economy as well as in organizational contexts. Negotiation skills training can significantly improve communication efficiency, social competence, job performance and conflict management. With the advent of serious games and game-based learning, that represent novel methodologies to approach traditional assessment and testing, the actual challenge is how to address the need to integrate these multidimensional psychological aspects in a pedagogical online environment that can establish a connection with the development of behaviour change in real situations. In this section an exhaustive summary of the most recognised definitions of soft skills is presented, with a particular focus on the concept of negotiation, its evolution in literature, and which acceptation has been chosen as theoretical framework for the present work. Other than that, an introductory analysis of educational technologies and serious games is shown, along with how methodologies taken from role-play techniques can help operationalize soft skills and create more engagement.

SOFT skills are involved in personal aspects and relationships in every stage of people's life transitions, from family to school, college or workplace, and are crucial for an effective emotional understanding and communication, for promoting personal performances, leadership and resource management, to accomplish goals in cooperation with peers, superiors or subordinates in both organizational and daily situational contexts. Soft skills are relevant for individuals and groups, and regulate teamwork, collective growth and help create a shared view in professional and educational set-

tings. The first approach to soft skills is born in contrast with the definition of "hard skills", that are quantifiable skills usually acquired through formal education, and can be evaluated, assessed and rigorously measured with recognised standardised methods. These skills include all the domain-specific knowledge and technical competencies required in a given organizational context to perform a job, such as programming, finance, technical writing or data analytics. Soft skills, due to their multiple intangible behavioural, interpersonal and emotional dimensions, are difficult to be uniquely defined, and sometimes the hard skills and soft skills borders overlap. Soft skills are meant to be a complement to hard skills, and in many jobs they represent a requirement that employees need to prove during interviews, and are evaluated through the use of specific standardized tools and psychological tests. One example of job that involves both soft and hard skills knowledge is project management, where leadership, communication, decision making, negotiation and active listening are essential requisites that allow to improve the performance of more technical skills like budgeting, planning or contract management. Apart from these characteristics, that allow a broad application of soft skills in several hard skills areas, the context specificity and cultural uniqueness of each soft skill is another crucial variable that changes from environment to environment across groups. What can be acceptable within a specific age and cultural group, for example, can be considered inappropriate in other contexts, and this must be taken into account as the ability to distinguish social differences is also determinant for the effectiveness of one's behaviour.

Among all the types of soft skills that have been identified, one that can be involved in all daily life situations, such as work or family-related issues, is negotiation. The act of negotiating allows two - or more - parties to communicate their needs, ideas, positions and to help achieve a solution that can represent joint interests. However, the concept of negotiation has been vastly discussed and it is still much debated in literature. The difficulty of providing a comprehensive definition of the concept of negotiation skills, considered as conflict management abilities used as a cooperative rather than competitive process, comes from the variety of contexts and disciplines to which it can

applied, from economics to psychology. Among all the negotiation theories, Rahim and Bonoma's model of conflict handling, based on five styles that are born from the combination of two factors, concern for self and concern for others, has been chosen as the most robust theoretical framework for the Enact software. Rahim's model, in fact, highlights the importance of the situational aspects of negotiations, in that there is not just a single style which is always appropriate, but all styles have their specificity and usefulness, and this view shifts the notion of negotiation toward the idea of a positive learning experience in any possible case, regardless of the final outcome.

However, assessing and practising soft skills, and in particular negotiation skills, is generally an expensive and time-consuming process which requires the presence of experts throughout the phases of testing and evaluation. All cases must be handled individually, and most small to medium organizations cannot afford this process during job interviews or other activities. Serious games and virtual reality can represent an optimal balance between the personal and organizational need of assessing soft skills and the cost in terms of time and resources that traditional assessment requires. The Enact software, based on role-play and daily life simulations, aims at filling this gap, introducing a new methodology for an interactive, reliable, cost-efficient and portable assessment tool that also provides adaptive training and personalized learning.

The following paragraphs will explore the most dominant approaches to soft skills both from a theoretical point of view, i.e. across how the term was defined in literature and a more applied perspective, introducing the most known and diffused methodologies to implement serious game and e-learning technologies to assess and train soft skills and the transition from a more traditional to more technological and intelligent tools. The last paragraphs will focus on the notion of negotiation across the different theories that helped define the concept and the multitude of styles identified. In particular, Rahim and Bonoma's model of conflict handling will be presented, along with the motivations that made their model one of the most known and their psychological test one of the most used in literature.

2.1 The concept of soft skills

2.1.1 Soft skills, competency and emotional intelligence

The term "soft skills" has gained different meanings throughout years, and its definition has been compared to that of competency, social intelligence, interpersonal skills according to the discipline of reference. Thorndike, in his pioneering work, described social intelligence as the ability to understand and manage relationships with young and adults (Thorndike 1920). Boyatzis et al. (2002) considers soft skills as personal behavioural competencies that guide the process of performing a task and affect the job performance rather than the final outcome, a definition that is similar to what the author defines as competency (Boyatzis et al. 2000). This view is in agreement with a general idea of competency as a predictor of job success (McClelland 1973), even though the concept of competency itself has undergone several changes (Woodruffe 1993, Dubois 1998, Snyder & Ebeling 1992, Pate et al. 2003, Sultana 2009). In particular Dubois (1998) identifies competency, as well as soft skills, as desirable behaviours needed to achieve an higher job performance.

Along with the meaning of competency, soft skills have been considered for long time as a different way to define emotional intelligence (Goleman 1995, Salovey & Mayer 1989, Sjoberg 2001), especially when associated with communication skills, social interaction and interpretation of social situations (Murata 2008, Zigler 1973, Bar-On 1997). Soft skills, according to this view, represent the core of personal development and management of interpersonal relationships, and are a more significant predictor than intelligence quotient or hard skills (Goleman 1995). For this reason, soft skills cannot be measured or based on standard IQ assessment as they represent complex combinations of how people manage emotions and interact with each other. Goleman (1998), whose definition of emotional intelligence overlaps significantly with the definition of soft skills provided by several recent authors (Kantrowitz 2005, Klaus et al. 2007, Dell'Aquila et al. 2017), considers it as the dimension of intrapersonal and interpersonal abilities, naming the evaluation of this combination of factors Emotional Quotient

(EQ). In this work, the author describes the findings observed during the close study of hundreds of organizations and states that self-efficacy, commitment, self-control and awareness are key personal aspects that can drive not only individual but also organizational success and growth. Dulewicz & Higgs (2000), in a later analysis, proved that a combination of EQ and IQ indices can predict successful outcomes more accurately than only of the measurements.

Neurocognitively, soft skills are placed in the same areas of emotions, since the ability to choose the correct strategy in each situation depends also on primary processes and individual responses directed towards the specific situation. Lately, much effort has been made in the field of neuroscience in order to clearly define the circuits involved in the genesis of complex affective states. While the boundaries of the "limbic system" - if the concept is not yet abandoned - have been trespassed turning it into a fading region which includes some parts of the mid-brain and neocortex (LeDoux 2003), there is good evidence about the role of the amygdala in decision-making abilities, and in general for the generation of the emotion of fear (LeDoux 2003, Seymour & Dolan 2008, Rudrauf et al. 2008). These behaviours, according to evolutionary theories, originated from simple reactions that were selected through evolution in order to promote the survivability of the species in their specific primitive environment. An example may be fear, an emotion that can arise in the situation of being alone at night and triggers the decision to try to perceive possible additional cues of the presence of a threat (Cosmides & Tooby 2000, Pacella, Ponticorvo, Gigliotta & Miglino 2017). Each ancestrally recurrent situation activates an associated "behavioural program" which in the past led to the best adaptation to repeated events and conditions posed by the environment. In order to distinguish the emotions which can be considered as elicited by a recurrent situation in the past from higher cognitive skills, Panksepp proposes the distinction between primary, secondary and tertiary processes, whereas the first and second types, which he recognizes as being basic, mostly unconscious, do not necessarily involve cognition in order to arise, while the third include cognitive executive functions like thoughts and planning, located in the frontal cortex, intrapersonal and in-

terpersonal emotion regulation, in the medial frontal regions and intentional behaviours in parietal regions and hippocampus (Panksepp 1998, 2010).

A subsequent attempt to characterize the main features of soft skills, in comparison with hard skills, has been proposed and vastly agreed in recent years (Klaus et al. 2007, Pant & Baroudi 2008, Hendarman 2012). In particular, hard skills are defined as:

- having technical or scientific procedures;
- involving the application of predefined and industry-specific tools;
- based on inflexible and replicable operations;
- defining a structured and predictable scope.

On the other hand, soft skills are broadly agreed as:

- irreducible to predefined procedures;
- context and situation specific;
- involving multiple solutions and outcomes;
- with a wide scope of application.

In addition to how soft skills can drive success in different areas of daily life, the lack of these abilities can determine failure in many aspects (Sternberg 1997). Sternberg (1997) in a pioneering work, described several soft skills areas of deficit that can lead to unsuccessful outcomes. Among them there is lack of motivation, which can affect significantly the presence of individual talent and, in organizational contexts, is mostly caused by job dissatisfaction. Impulsiveness, especially during team work or group work, can also represent a characteristic that can be extremely counter-productive, and can drive unjustified acts. Unbalanced perseverance, considered as the inability to become aware of when to give up or continue to pursue an objective, regardless of whether collective or individual, can lead to an increased waste of energy and efforts.

Extreme thoughtfulness, caused by an excessive attention to a specific detail of a task, can significantly slow individuals and therefore negatively affect performances, and is closely related to the inability to maintain a practical aim and concentrate on the result rather than the process. Fear of failure, and subsequent inability to initiate a task, is caused by lack of self-efficacy and self-motivation, and determines unsuccessful outcomes. The opposite, that is excessive perceived self-confidence, can also lead to failure due to constantly misattributed blame, i.e. accusing others of own mistakes, or procrastination, a consequence of a missing or ignored pressure. As lack of soft skills can contribute to unsuccessful results, also soft skills competence and awareness can be utilised in a maladaptive manner, in order to consciously introduce biases in interpersonal relationships, social interactions and emotion regulation. In particular Kantrowitz (2005) listed six possible uses of soft skills that can be counter-productive:

1. leadership skills to exercise power and influence on others;
2. communication skills as a way to comment inappropriately in a subtle manner;
3. team leading to micromanage group work;
4. social skills to propagate stereotypes or social differences;
5. self-awareness to underestimate or overestimate own capabilities;
6. self-efficacy to undermine the other's skills.

Even if the absence or the wrong application of soft skills can be easily noticed and recognized, it is still difficult to encompass this concept in an exhaustive definition or framework (Giloth 2000).

2.1.2 Definitions of soft skills

Concerning the definition of soft skills, this task has represented a big challenge for psychologists, trainers, counsellors, researchers and managers. Since it is difficult to dig deep into the conceptual framework that stands behind each soft skill theory, these

has more often resulted in loose and inaccurate definitions. Halfway through the 90's, a multitude of different studies tried to provide a general definition of this concept, in the attempt to discuss the weaknesses of the previous theories.

For Proctor & Dutta (1995), a soft skill is a goal-directed behaviour whose performance is improved with practice and that requires less and less effort every time. Moss & Tilly (1996), in their definition, underline how these behaviours concern personality, attitude and traits rather than formal and well-defined knowledge. Ming Chia (2005) also includes, in the definition, abilities that are involved in personal development and career, such as the ability to manage stress, intrapersonal and intepersonal communication, communication and thinking skills. James & James (2004), for example, also agree that soft skills are extremely important for their potential in workplaces, especially because they help a candidate to suit their specific job and help to achieve a higher productivity. As Jackson (2010) also points out, this characteristic is particulary evident in jobs such as management, customer service and other jobs that require to work in teams or groups. Soft skills are what makes an individual, a candidate, an employee different from the others and makes him or her stand out for the commitment, qualities, behaviours and personal attributes rathern than for the technical knowledge and hard skills competence (Deepa & Seth 2013). Some authors defined soft skills as a mandatory requirement for the correct application of hard skills in the workplace, in that technical knowledge is theoretical and cognitive, while soft skills are practical and behavioural aspects of the same process (Weber et al. 2009).A focus on soft skills viewed in particular as leadership skills is provided by Bancino & Zevalkink (2007), who believes these abilities are what help distinguish the roles inside an organization or team. For example, a true leader helps the group achieve the goals and aims set within each specific task.

While Pratt et al. (2009) also points out the importance of soft skills in the field of creating strong relations between coworkers and partners in the workplace, in that they enhance communication and team management, Duffy et al. (2004) restricts soft skills

self-mastery (intrapersonal)	self-awareness	emotional awareness accurate self-assessment self-confidence
	self-regulation	self-control trustworthiness conscientiousness adaptability innovation
	motivation	achievement drive commitment and initiative optimism

Table 2.1: List of intrapersonal (self-mastery) soft skills

to the range of interpersonal behaviours. This view, in fact, believes that soft skills are a verbal and non-verbal way of creating effective relationships and to identify the others' problems, stress, mood in order to decide a resolutive communication strategy. Boyce et al. (2001) broadens this definition by including all skills that are not domain specific and not exclusively relational, such as problem solving, critical skills or judgement. This intangible aspect of soft skills is also highlighted in Kajnc (2010) theory. These authors, in fact, show how these abilities are influenced by personal and cultural characteristics, experiences, and all of these aspects that do not depend exclusively on the individual's education. Sukhoo et al. (2005) also underlines that practical skills such as planning, team building, working with people and ensuring customer satisfaction, time management and meeting the group expectations are all individual skills that are learned by doing, in an ongoing process of practising and improving based on feedback and experience that is subjective and specific to each environment.

Finally, personal soft skills are the result of subjective learning based on acquired experience. They concern both self and people, in that they include intrapersonal and interpersonal aspects.

They enable the effective use of hard skills and technical knowledge in the workplace, creating the link between theoretical and practical abilities and competence, as listed in 2.2 and 2.1. They represent the bridge between job knowledge and job performance, in that they help meet personal, team or organizational goals, as well as helping to

people skills (interpersonal)	empathy	understanding others
		service orientation
		developing others
		leveraging diversity
		political awareness
	social skills	influence
		communication
		conflict management
		leadership
		change catalysis
	social co-ordination	building bonds
		collaboration and co-operation
		team capabilities

Table 2.2: List of interpersonal (people skill) soft skills

achieve the best requirements, opportunities and management on the workplace. Soft skills are also crucial in personal lives, as they are not context specialised but they can be applied in all environments and are particularly important to reach effective communication, to solve problems, manage stress and maintain trust in all kinds of relationships. For this reason, soft skills cannot be easily quantified, evaluated or communicated, nor taught in a unique way. Each individual will need to initiate a subjective internalisation process to improve and practice them.

2.1.3 Teaching and practising soft skills

(Goleman 1998, Boyatzis et al. 2000) and introduces the idea that these skills can be developed over time. Similarly, pioneering studies agree on the importance of leadership, management, organizational competences and shared values for the success of a team and, more in general, a company (Watson 1983). In recent years, a strong correlation between soft skills and superior performance in leaders was found (Cavallo & Brienza 2002). But can soft skills, competency and emotional intelligence be taught? A recent debate arose about whether soft skills have stable characteristics that are related, for instance, to personality traits (Robles 2012). From the answer to this question and from the fact that soft skills can permeate every stage of an individual daily life, it is possible to derive the need for a continuous, lifelong, training and assessment of these

abilities.

Another possible training method for soft skills - since these abilities can be considered as social dimensions - is role-playing. With the term role-play we define an experience that revolves around a creative and imaginative process, that becomes concrete throughout the interpretation of a real or imaginary role of a character inside a specific situation (Freedman 1969, Yardley-Matwiejczuk 1997). The interpretation of a role allows the participant to have a direct experience of a specific situation and also of the reactions that each act sets off in its interlocutor and/or in the other components of a group.

Role-play techniques derive from the classical psycho-drama, brought by Moreno in the psychotherapeutic framework, and he was the first to demonstrate the usefulness of the game and of the dramatic action in the expression of repressed feelings, creativity and spontaneity (Moreno 1946). Moreno felt the inextricable pedagogical value of the drama and invited researchers to adopt the psycho-drama beyond the therapeutic scope in order to improve and train social competencies. Role-playing techniques had a huge diffusion outside of the therapeutic framework and in particular in the context of pedagogy, education, and organization development - such as recruiting and assessment (Bell 2010). Role-play games share the basic mechanisms and the principles of the psycho-drama. In particular, a role needs a bipolarity, the presence of two distinct and opposite entities that, interacting, create a relationship. The term role-playing is often used to indicate different kind of experiences that range from the therapeutic to the pedagogical framework and can be addressed to groups of people (multiplayer) or an individual (single player).

Recently, researchers tried to experiment many transitions between classical and more modern methodologies in the field of role-playing. In education, EMORPGs (Education Massive Online Role Playing Games) allow many players - represented by in-game avatars - to simultaneously participate to multiple educational online game sessions; other role play games, instead, are chat based, and ask the users to interact with writ-

ten text or voice messages. Most of the times, anyway, the presence of a teacher that guides the interaction is required, and all the participants need to be online at the same time. This is where the importance of environments embedded with simulated virtual agents starts to be evidenced. Virtual agents can, in fact, provide a solid basis for practising soft skills in the presence or absence of a teacher, and can represent a safe and reliable simulating experience where the trainee can learn by doing and exercise soft skills in a broad variety of scenarios. The Enact software, creating a bridge between previous serious games and role-playing simulations literature, stands out as a psychologically-based role-play serious game specifically targeted at negotiation and conflict handling. The details and implementations of serious games for soft skills will be described in particular in the next chapter.

2.2 A review of negotiation skills theories

Since, as previously mentioned, the main difficulty lies in structuring the concepts inside ill-defined domains, it is important to stress that in the current literature there is no agreement about a formal definition of negotiation. The difficulty of establishing a universally accepted description of this concept is a direct consequence of the variety of disciplines in which the term has been defined and theorised, such as economics, law, international relations, psychology, sociology and conflict management. The negotiation concept adopted for the development of the Enact platform takes inspiration from the Integrative approach of Negotiation (Walton & McKersie 1965, Fisher et al. 2011, Zartman & Berman 1982) which is based on cooperative processes, rather than competitive-distributive processes.

The terms "distributive" and "integrative" were first presented and defined by Walton and McKersie in the field of collective bargaining and work negotiations in the 1970s (Walton & McKersie 1965). In modern literature, both terms are commonly used, especially when referring to the process of reaching mutual agreement and handling interpersonal conflict. Distributive negotiation process represents a method in which two - or more - parties strive to divide a fixed amount of resources, that can be money or

other possessions, and where each one has the need to maximise its distributed portion during the decision. The distributive approach is also defined as a "fixed-sum" or "fixed-pie" game, as the negotiation revolves around the division of a fixed amount of objects or possessions between parties and it is not possible to add external resources. This process is also named "zero-sum" because, once the negotiation has settled, one of the parties will lose the same amount gained by the other. The integrative approach, instead, also involves "making concessions" to the other party in order to reach an agreement, although what defines it regards the search for a reciprocal and mutually satisfying solution. This process is also referred to as the "expanded-pie" approach - in contrast with the distributive fixed-pie - because parties can engage in an active and creative search for more agreeable and unexplored solutions than the two standard possibilities that meet just one party's personal interests. In other words, parties are willing to cooperate in order to reach the highest amount of mutual interests by integrating their needs in a unique agreement that possibly satisfies all parties. The integrative approach is comprised and well defined in the theoretical framework of the well-known Win-Win paradigm (Henderson 1996). However the term integrative distribution may be preferred in the psychological context as the concept of "win-win" has been transformed and used sometimes to indicate a situation that negates and escapes conflict, rather than being just defined as the integrative style of conflict handling.

Other than that, the concept of negotiation can also be defined as the methodology used to effectively communicate between the two parties - the so-called "constructive feedback process" -, and reach together a joint view or solution that concerns all the needs or ideas expressed during the process. Within this definition, another important ability required is the one that allows to adopt the others' positions, and therefore transforming the opponent into a partner, and the bargaining into a relational exchange. With a joint work, both parties can have a greater opportunity to reach a mutually satisfying solution that can represent a benefit for both sides. The main difference between this approach and the distributive approach to negotiations is that in the latter parties are always in competition, as it is not possible to add external resources or solutions to the

negotiation, and therefore they also try to obtain the greatest gain and leave the other party with the smallest advantage. This constant and continuous competition makes it impossible in the majority of distributive negotiations to regard the other side as a partner rather than an opponent. In fact, one party will tend towards the most extreme solution - his or her own maximum benefit - and may engage in other behaviours such as lying, manipulating or bluffing to achieve it. Typical daily life examples of interpersonal conflict are the discussions about two employees contending a senior role, or about two partners moving to a new house.

Regarding negotiation processes seen as interpersonal conflicts, these can be handled with various styles of behaviour. In literature it is possible to distinguish different conflict resolution styles taking into account a two, three, four or five factor model. Deutsch (1949) first suggested the two factors cooperative-competitive model in his research about social conflict. Similarly to the perspective introduced by the game theory, this model uses a cooperative-competitive continuum to simplify the categorization of conflicts. Deutsch and colleagues suggested that the cooperative style compared to the competitive style is more effective when managing conflict and leads to more functional outcomes. However, these studies did not show evidence of a correlation between cooperative style and job performance and productivity. Apart from theoretical speculations, dealing with practical situations using a purely cooperative or a purely competitive conflict is implausible, so game theorist recognised that conflict situations can be characterised by both cooperative and competitive aspects. Example of models who propose three dimensions of handling interpersonal conflict were those by Putnam & Wilson (1982), who defined them as non-confrontation, solution-orientation, and control, while Lawrence & Lorsch (2009) listed forcing, smoothing, and confrontation. Other two models of the three styles of handling conflict belong to research in the area of marital conflict, respectively developed by Billingham & Sack (1987): reasoning, verbal aggression, and violence; and Rands et al. (1981): attack, avoid, and compromise. However there is no evidence of the relationships between the three conflict styles and organisational behaviour, and individual, group, organizational outcomes. According to

Rahim (1983), the main limits of these models is that the theoretical basis for the three-category conflict styles is not clear and the statistical analyses and methods used for investigating the factors are not sufficiently robust.

Pruitt (1983) suggested a four style model of handling conflict based on the two continuous dimensions of concern for self (high or low) and for others (high or low), whose combination results in the following styles: yielding, problem solving, inaction, and contending. Empirical evidence from laboratory studies (Pruitt 1983, Pruitt & Carnevale 1993) has shown that problem solving is the most effective style for managing conflicts. Another four styles model of conflict management useful for the conceptualization and operationalization of marital conflict was proposed by Kurdek (1994), and comprises the following four dimensions: problem solving, conflict engagement, withdrawal, and compliance.

The first five factor model of handling interpersonal conflict in organizations, instead, was conceptualised by Follett (Schilling 2000). The author first found that conflict was managed using three primary behaviours: domination, compromise, and integration, and two secondary approaches, avoidance and suppression. The first conceptual scheme for classifying the styles into five types was proposed by Blake and Mouton (Blake & Mouton 1970): forcing, withdrawing, smoothing, compromising, and problem solving. The model was based on two main dimensions: concern for production and concern for people. These dimensions describe the attitude of the manager of being a task or relation oriented leader, and the different combinations result in five different leadership styles.

Blake and Mouton's scheme was reinterpreted and extended by numerous researchers. For example Thomas (1992) extended this model including the intentions of the parties involved, and classified the conflict handling styles using two dimensions: assertiveness (attempting to satisfy one's own concerns) and cooperativeness (attempting to satisfy the other party's concerns). A combination of the level of these dimensions determines the following five conflict handling modes employed by the parties: competing

(assertive and uncooperative), collaborating (assertive and cooperative), compromising (moderate in both assertiveness and cooperativeness), avoiding (unassertive and uncooperative) and accommodating (unassertive and cooperative).

Thomas has suggested that an individual's conflict style is a behavioural orientation and general expectation about one's approach to conflict (Thomas 1992). This idea of conflict style does not preclude the possibility that the individual changes style or enacts behaviours not typically associated with a particular style, but asserts that they choose - though often not consciously - a pattern of principles that guides them through the episode of conflict. Conflict management style is a "general and consistent orientation toward the other party and the conflict issues, manifest in observable behaviours that form a pattern and share common characteristics over time" (Kuhn & Poole 2000). Within this framework, the author, along with Killman, developed the TKI (Thomas-Killman Inventory) a standardised test for measuring the disposition to each of the 5 styles which provides the predominant style, along with an assessment of the less frequently used conflict modes. Thomas, in his structural model sustains that peoples responses' styles are hierarchically ordered, meaning that each individual has a dominant style, back-up style and a least preferred style, depending on their personality (Thomas 1992).

The TKI, along with the ROCI II developed by Rahim, are two of the most well-known self-report instruments for the evaluation of conflict management styles. However studies on the TKI revealed weak to moderate and concurrent variability, little evidence on content validity, and failed to find external and predictive validity questionable predictive validity (Van de Vliert & Kabanoff 1990). Rahim criticised directly the TKI in its form, and therefore developed the ROCI II that has achieved an extremely high validity and reliability up to date.

2.3 Rahim and Bonoma model of interpersonal conflict management

Based on the conceptualizations of Follett (Schilling 2000), Blake & Mouton (1970), and Thomas (1992), Rahim and Bonoma differentiated the styles of handling interpersonal

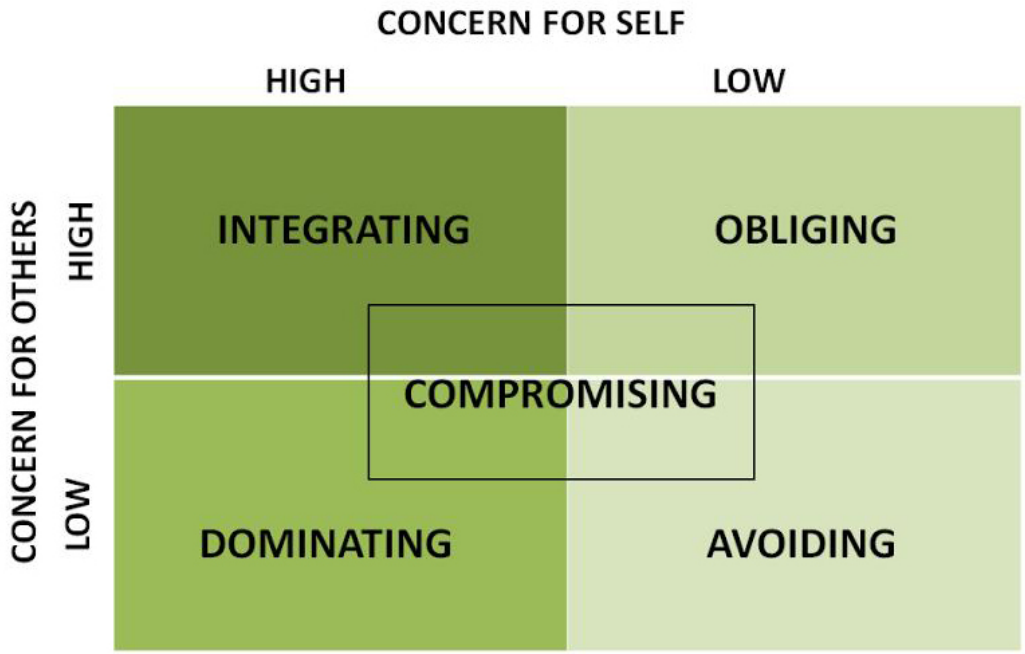

Figure 2.1: Rahim and Bonoma's theory of conflict handling styles

conflict on two basic dimensions: concern for self and concern for others (Rahim & Bonoma 1979).

The first dimension explains the degree (high or low) to which a person attempts to satisfy his or her own concern. The second dimension explains the degree (high or low) to which a person attempts to satisfy the concern of others. The combination of the two dimensions results in five specific styles of handling conflict: integrating, obliging, dominating, avoiding and compromising. In particular, as shown in 2.1 :

1. Integrating (high concern for self and others) style involves collaboration between the parties that are willing to reach a mutual and acceptable solution through openness, exchange of information, examination and exploration of differences for arriving to a constructive solution that goes far beyond personal and limited visions of the problem.

2. Obliging (low concern for self and high concern for others) also known as accommodating style is associated with attempting to play down the differences and emphasizing commonalities to satisfy the concern of the other party. An obliging person can be defined as a "conflict absorber" terms describing a reaction of low hostility or even friendliness to a perceived hostile act.

3. Dominating (high concern for self and low concern for others) style has been identified with win-lose orientation or with forcing behaviour to win one's position.

4. Avoiding (low concern for self and others) style has been associated with withdrawal, buck-passing, or sidestepping situations. This style may take the form of postponing an issue until a better time, or simply withdrawing from a threatening situation. This style often reflect little concern toward the issues or parties involved in conflict, and the attitude to refuse or denying to acknowledge the existence of a conflict in public.

5. Compromising (intermediate in concern for self and others) style involves give-and-take whereby both parties give up something to make a mutually acceptable decision.

Rahim insists on the value of the possibility of training individuals on how to choose and make use of all the different styles of handling conflict in order to deal effectively with various situations. For example, the integrating style can be useful for dealing with conflicts that involve complex problems or strategic issues, and when decisions cannot be taken by a single individual since the value of the differences (skills, information, experience) of both parties can be useful to define the problem, and identify effective alternative solutions. This style can be inappropriate when an immediate action is required, but there is no time for problem solving strategies and the task is not simple. Moreover the style may not be efficient when parties are unworried about the outcome.

The obliging style can be useful when the individual is not familiar with the issues involved in a conflict. It may be useful when the party is unconcerned about outcomes,

and for preserving a relationship that might be more important that the immediate outcomes, yet as a strategy when a party is willing to give up something with the hope of getting some benefits in the future. This style is not appropriate if the issue involved in a conflict is important to the party, and when there is a belief that the other party is wrong or acting unethically.

The dominating style should be the first choice when an immediate action is needed, or when an unfavourable decision taken by one of the parties involved in a conflict can be harmful to this party itself. It should be used , for example, by supervisors dealing with subordinates who lack technical expertise to make decisions or when the implementation of unpopular courses of action is needed. This style is inappropriate if the issues involved in conflict are complex and there is enough time to discuss a mutually agreed decision, by using problem solving strategies and when the issues are not important to the party. If used by equally powerful parties it may lead to a deadlock. The avoiding style is considered appropriate when a confrontation with the other party may be risky and the negative effect on the relationships between the parties would exceed the benefits of the resolution of the conflict. It may be also useful when the task or problem to solve is trivial or simple. The style may not be appropriate to use when the issue involved in a conflict is important to a party, when party it is given responsibility to make decisions, or when prompt action is required.

The compromising style is advantageous when the goals of the conflicting parties are mutually exclusive, and when an impasse occurs between equally powerful parties. It can be used when it is difficult to reach a consensus, parties need a temporary solution to a complex problem, conflicts are protracted for long time, or other styles have been used and found not to be effective in resolving the issue. Often the use of the compromising style for dealing with complex issues fails to reach durable long-term solutions. This style may not be appropriate when the conflict involves dealing with important or personal values.

In the first operationalization of the Enact training scenarios compromising is not iden-

tified as a specific style of interaction, rather as "a lack of a style", as it emerges from the combinations of the choices of the other four different styles. Also in literature has been debated whether considering the compromise as a distinct conflict handling style, or as a dimension similar to the integrating (Van De Vliert & Hordijk 1989). In a second stage, the compromising style has been included as a different style from the others, in order to provide the same evaluation process for all the possible styles.

Rahim's model (Rahim & Bonoma 1979) was chosen for the following main reasons:

- The model stresses the learning process and the idea of enabling organizational members to learn the appropriate use of the handling conflict styles depending on situations (Afzalur Rahim et al. 2001). In training contexts, this implies that the appropriateness of a style should be determined by the context and by the learning objectives. People, when facing conflict experiences, might employ different styles, and not only one. Learners through the interactions within the Enact scenarios can have the opportunity to reflect on the different styles employed during a conflict episode, and being aware of the most appropriate effective style for that specific given situation;

- In this model, the idea of conflict is not seen as a negative concept itself. When conflict is handled appropriately, positive learning experiences may result in the development of individuals and teams. This variable directly depends on how a conflict is managed and handled. Key aspects of conflict management start from self-awareness, understanding group dynamics, and finally taking action.

- Rahim and Bonoma developed the Rahim Organizational Conflict Inventory-II (ROCI II), for assessing the 5 styles of handling interpersonal conflict (Rahim 1983, Afzalur Rahim et al. 2001). The ROCI II, as previously mentioned, is the most widely used conflict management instrument and it has been rigorously tested by many researchers through the years. It is available in 3 forms (A, B, C) to assess respectively the individual's styles of handling interpersonal conflict with superiors, subordinates, and peers. The ROCI II has been used for the

evaluation and feedback to the group targets, and for validating the model itself after the trialling phase of the software.

The whole negotiation framework used inside the Enact platform is based on the description of the 5 styles provided by Rahim, from the design of the scenarios to the variables, the generated profile, the tutoring system and the adaptive feedback. In particular, the whole process of assessment and training is supported and validated by the use of this instrument and compared with other scales such as the Big Five Inventory (short version), the Assertive efficacy and Self Efficacy scales and a short version of the psychological test "Ways of Coping", originally developed by Folkman and Lazarus (Folkman & Lazarus 1988).

2.4 Conclusion

An attempt to categorize soft skills has been provided, encompassing the conceptual view that considers them as task-related, self-directed or people-directed behaviours complementary to hard, more technological knowledge in both private and public settings. Soft skills, similarly to emotional and social intelligence allow individuals to effectively perform and achieve personal, collective or organizational goals. These competencies, applied on specific tasks, can facilitate the application of hard skills as well as support decision-making and the resolution of problems, in order to effectively obtain the desired outcome. In their intrapersonal aspect, these abilities enable the development of self-awareness, self-efficacy, problem solving, managing appraisal and judgements made and received and help relieve everyday challenges, as pressure and stress in organizational settings. These abilities can help finding a balance between self and other interests and finding creative solutions to problems. They can also empower the individual increasing leadership behaviours and allow an effective teamleading and management. In their relational aspect, in fact, soft skills are beneficial for all types of social relationships, like those with subordinates, peers and superiors.

Increasing the performance in dyadic relations as well as group collaborations, these

skills have a significant effect on interpersonal aspects like trustworthiness, verbal and non-verbal communication, conflict management and negotiation. The last one, in particular, appears involved in the majority of daily life interpersonal interactions, and therefore represents one of the skills that requires continuous training and practising. Since soft skills and negotiation can only be learnt-by-doing, there are few possibilities of effective training and testing, and fewer if we consider that traditional psychological methodologies are time-consuming and expensive and might be conducted only in specific settings and big organizations. An attempt to cover this gap in the educational tools available for soft skills is represented by serious games and technology-enhanced learning software, whereas portability, availability and self-assessment is possible, reliable and quick.

Based on Rahim's theory of negotiation, that involves the definition of five different negotiation styles from the combination of two dimensions, concern for self and concern for others, Enact tries to convey this need by proposing a new methodology for 3D serious games and simulations in order to provide free online training and assessment of negotiation skills that is as close as possible to traditional negotiation tests.

Chapter 3

A review of Serious games and Intelligent Tutoring Systems

In recent years, the need of technology-based platforms to assist traditional learning methodologies arose, and several experiments showed the importance of one-on-one tutoring compared to one-to-many. Many of the first developed learning platforms based on curriculum sequencing evaluate the user's knowledge through direct questions only. In the last years, new platforms that automatically collect data during the user interaction, named Intelligent Tutoring Systems, were developed as a new standard inside serious games to provide the learner a more personalized curriculum. However several obstacles still prevent the wide use of these tools, among which the difficulty of finding the right way to deploy the information to the user and to collect data from the learner, or the lack of evaluation standards to assess the user's knowledge. Also most recently developed platforms only focus on hard skills. This review will explore the actual state of art of serious games and Intelligent Tutoring Systems, analysing strengths and weaknesses of the actual platforms and the barriers which prevent the widespread of these learning methods; it will offer a view on the different implementation methods, including situated tutors and game-based learning; it will also describe the development of an Intelligent Tutor able to provide a personalized training on the user's negotiation skills and the possible solution to the actual urgent need to develop a methodology to reliably assess the user's soft skills.

THE interest in applying technology to the field of education starts to rise in the 50's, when behaviourists, involved in the study of the possible applications of the conditioning paradigm, designed the first automated teaching systems. These simple systems were not able to build and sequence material or to provide a tailored feedback, and the way they were designed did not provide any specification on the type of instruction they were deploying but just on how to deploy it. This means that there was no difference in the architecture according to the kind of taught subject. These systems were not registering neither responding to the user answer to their instruction, since

the path was predefined and there was no possibility to switch it or go on a previous step. Only in the late 60's, when the behaviourist approach led its way to the "cognitive revolution" more complex systems started to emerge, which included non-predefined paths and a slightly more tailored material, resulting, for instance, in evolved finite state machines with branch architecture.

However, the implementation of these systems, despite their simplicity, had two main aims: a better understanding of the processes that lie behind education and a demonstration of the effectiveness of one-on-one instructions, even if deployed by a non-human system. In fact, also in recent years many studies demonstrated how effective one-on-one tutoring is, proving that students who receive this kind of instruction perform two standard deviations better than students in traditional classrooms (Ong & Ramachandran 2005). But what kept this type of systems far from perfection was the lack of consideration of cognitive aspects, such as emotions, motivation and specific user temper. Beyond the idea of "one size fits all" there is a variety of shades into the user behaviour that cannot be ignored as the performance of the learner will strongly rely on these variables. The thought that not every learner could approach a subject in the same manner as the others led to the idea of structuring the teaching material considering the single student's needs and characteristics. Sleeman and Brown, in 1982 defined Intelligent Tutoring Systems (ITS) as "adaptive systems which use intelligent technologies to personalize learning according to individual characteristics such as knowledge of the subject, mood and emotion" (Sleeman & Brown 1982, Jr et al. 2013).

A recent definition of Conati (2009) states that "ITS is the interdisciplinary field that investigates how to devise educational systems that provide instruction tailored to the needs of individual learners, as many good teachers do." Intelligent Tutoring Systems learn about the user while the user learns about the subject and their evolution is synchronous with his progress and the interaction with him. The field of ITSs is wider than it can seem and embraces a big range of intelligent systems, like conversational

tutors, simulation-based tutors and modern serious games. As Conati underlines, "There are, however, other educational activities that can benefit from individualized computer-based support, such as studying examples, exploring interactive simulations and playing educational games. Providing individualized support for these activities poses unique challenges, skills and mental states often not as structured and well-defined as those involved in traditional problem solving." (Conati 2009, Jaques et al. 2014). Many studies underline how the use of ITSs provides better results (Shute & Psotka 1995, Paviotti et al. 2012). However, the higher the complexity of these systems is, the higher amount of time is required to develop them; it has been shown that in order to implement 1 hour of information deployed via an intelligent tutor, it is required an average of 200 hours of labour (Woolf & Cunningham 1987, Steenbergen-Hu & Cooper 2014). The cost of developing an ITS is high in terms of time and resources, especially considering the variety of disciplines which can be taught, the different purposes and user target (Dell'Aquila et al. 2017).

One of the actual challenges of the field is, therefore, to build an Intelligent Tutoring System which can be used through different domains, from well-defined to ill-defined ones, that can cover both strictly hard disciplines and soft skills, and that can be easily adapted to a wider range of users considering their gender, age and other personal characteristics; also, its features should include accessibility, mobility and easiness to structure even for non-professionals.

3.1 Classical methodology for developing technology-enhanced tutors

3.1.1 Computer-Assisted Instruction

When speaking about tutoring and assessment, one must not forget how integrated these two processes are. The concept of tutoring itself includes the ability to guide the learner and fill the "gaps" in the learner's knowledge, and, in order to perform this, the tutor should be able to correctly diagnose the current state of the knowledge of the student, assessing in real time and modelling his own belief of the student. So an ITS is an attempt to reproduce in a computer what in a human would be called "good teach-

ing" (Elsom-Cook 1993, Heift & Schulze 2007). The ideal tutoring systems should be able to assess the student for his absolute knowledge of the subject, for his knowledge related to other learners (i.e. previous interacting agents) and for his progress during the session with the platform. To reach this point, the research on tutoring systems has gone far since its beginning. The first tutoring systems did not contain any representation of the learner and of the knowledge inside of it. The bases of Intelligent Tutoring Systems lie into Computer-Assisted Instruction (CAI), the first way to implement a tutoring system on a PC. These systems were structured to present the material to the learner on the screen in a programmed path. The first implementations had a simple and linear structure, did not behave according to the user characteristics, and just "ignored" the user response, deploying the programmed information in the same order and path for every learner. They, in fact, did not possess a student module to build a representation of the student they are teaching to. This type of CAIs was called "ad-hoc frame oriented" (AFO) in order to specify its exclusive dependency on the author's implementation and material structure. Latter implementations of these systems were based on more classical and operant conditioning principles, i.e. they favour an immediate positive feedback after a correct response that acts as a reinforcement of the learned knowledge strengthening the stimulus-response association (Skinner 1963); for this reason they provided an immediate feedback about the correctness of the answers (Pacella & López-Pérez 2018).

The act of the tutor of switching his behaviour during the curriculum path and try to avoid another learner's mistake is called "remediation". In these simple systems, if the response provided by the student was incorrect, the CAI interrupted his curriculum path and presented the previous instruction again, followed by an item - theoretically similar to the connessionist stimulus - identical to the former by category and type, in order to "condition" the student to the correct answer. The theory behind it was that, in disregard to the nature of the error, the student could have been conditioned to learn the right answer by the repetitive presentation of similar kind of stimuli, as stated by the behaviourist's "law of frequency" (Watson 2017). These programmed instructions

systems dominated the 70's until the need of a more cognitive approach emerged.

3.1.2 Generative systems

Generative systems represent an evolution of CAIs, in a way that allows a more complex structuring of the teaching material from the system. This implies that the system is required to sequence material and to solve meaningful problems posed by the student. These systems, mostly used to deploy arithmetical and geometrical knowledge, represent the real precursors of ITSs: although they do not have a structured knowledge module, they are able to respond to the student's questions and to provide exercises tailored to the student's needs. The problem analysis and solution was generated by comparing parameters in the student's answer to a predefined list of possible outcomes, and then, according to the relevance, categorized the error providing the most useful exercise. It is considered the first type of system to provide a sort of scaffolding - conceptualised as "guided prompting that pushes the student a little further along the same line of thinking [...] giving direct feedback on a student's response" (Yamauchi 2001, p. 490) - to the user through his learning process. The privileges of using this approach were the reduction of memory usage and a more dynamic management of the knowledge, even though the system was not "aware" of the subject matter. The instructions, in fact, were still stored as group of texts, without any categorization or modelling. Another name for this kind of systems was "adaptive systems", since they could adapt - in a simple but functional way - to the student's mistakes in the particular domain. The behaviour of these tutors cannot be really considered fully adaptive, since they lack of a student model, and also they are not able to provide dynamic solutions to problems. The next step in this direction would have been to add machine learning techniques into Tutoring as will be explained in section 3.4.

3.1.3 Curriculum sequencing

The main aim of Curriculum Sequencing lies behind the need of deploying information to the learner in an order that suits their need and takes into account their existing knowledge, their personality, learning style, individual traits and context of work

(Brusilovsky 1992, 2012). This allows the student to receive a personalised flow of instructions and type of material in a way that is best tailored to him. This important ITS module was the first one to include a preliminary form of student model, which used the stored learner's responses as well as the user's personal information to provide material in an order that would most suit the student. . Some of the early ITSs totally relied on a finite state machine structure, while in recent works curriculum sequencing is still integrated into more intelligent, complex and advanced systems, as in the case of the Enact platform. This technique, for instance, is still widely used especially in well-defined domains, since a reliable algorithm based on expert knowledge is enough to provide an adaptive curriculum to the learner.

3.2 The structure of an Intelligent Tutoring System

3.2.1 Requirements for ITS development

A total redesign of Tutoring Systems was made with the integration of Expert Systems and artificial intelligence into non-human tutors. A definition by Polson & Richardson (2013), list three core features an Expert Tutoring System must have in order to be defined an Intelligent Tutoring System:

1. It must have an expertise on the subject, that is, to have a dynamical representation of the subject domain that allows autonomous problem solving;

2. It must be able to store information about the student expertise on the subject which includes the knowledge prior to the interaction, his progress and a meaningful assessment per each step;

3. It must adapt on the learner, using curriculum sequencing to structure the material in order to reduce the difference between the expert's and the learner's knowledge.

More recent works investigated the pros and cons of automated tutoring systems compared to human tutoring. In particular, the main advantages of using ITSs are (Latham

et al. 2012):

- Curriculum sequencing: as discussed above, the ability to decide an algorithm to sequence the information presented in a way that suits the learner's needs is an extremely powerful tool that allows customised experience and reliable student assessment;

- Intelligent solution analysis: this technique allows dynamical problem decomposition and a better analysis of the mistakes of the learner, providing a detailed and customised feedback (Kenny & Pahl 2009);

- Problem solving support: according to the constructivist approach, ITSs allow a deeper understanding of the subject by helping the learner to autonomously reach the solution to the problem (Redondo et al. 2007).

On the other hand, the disadvantages of ITSs are (Juárez-Ramírez et al. 2013):

- Expensive research: the development of every ITS requires a team which includes psychologists, computer scientists, educators and experts of the subject, plus an amount of time which may vary but still bigger in average than the preparation of a class lecture;

- Lack of intrinsic motivation and human contact: most of the first ITSs did not give any real context in which the deployed knowledge could be used and did not provide any source of reward during the interaction. The contact with a human tutor, also, can be motivating itself for the learner thanks to the positive feedback given by words, facial expression and social context.

- Lack of focus in learning content: rather than focusing in the learning content, many of the first ITS developed focused on the structure of the algorithm, thus the quality of the deployed content was poor, incomplete or limited and difficult to use in broader contexts.

- Lack of evaluation standards: since the specificity of the deployed content and since the lack of availability of other non-human platforms it is still very difficult to compare the results obtained by automated instructions and those by traditional courses.

Many of these problems have been faced and approached with different proposed solutions; one of the most important is the intrinsic motivation, which allows the student to be tied to the PC until the interaction with the tutor becomes effective enough to provide a progress of his knowledge. This has been achieved in several ways, as, for instance, in the field of serious games, where contexts have been introduced in the form of game scenarios and where the presence of an extrinsic motivation often results in the creation of challenges which involve personal aspects of the learner. Also simulation-based tutors aim at recreating a virtual environment as close as possible to real life situations, thus arousing the same feelings and emotions.

3.2.2 Structures and modules

A pioneering project by Carbonell, SCHOLAR, represents one of the very first attempts to create an ITS, and helped to delineate the actual structure of a general Intelligent Tutoring System (Carbonell 1970). According to this model, an ITS is composed of three core modules:

- the Instruction Module
- the Expert Module
- the Learner's module

These three components are interconnected and each can communicate directly with the other two. Another more recent model of ITSs was structured by Dede (1986) and included a fourth module, the User Interface, giving the Instructional module a different name, "Pedagogical module", which had the aim of building a customised path of knowledge for the student using the most efficient pedagogical strategies. The User

Interface, instead, being the most external module of all, allows the collection of information about the user in order to modify the existent relations among the components and is crucial in modern ITSs (Ma et al. 2014).

3.2.2.1 The Expert Module

The Expert Module is the component which contains the full knowledge of the domain. This knowledge is generally provided by a team of human experts of the subject. Once the content is defined, the way it is organized, coded and fragmented during the development of the ITS can be chosen among several options:

1. Sorting the instructions using a mathematical order
2. Sorting the content using human-like way of reasoning
3. Collecting data from the target group.

The way the content is deployed is chosen by the Instruction Module which will access to the Expert Module database in order to provide the best material suited to the student.

3.2.2.2 The Learner Module

The Learner or User Module is in charge to interpret all the data obtained by the interaction of the user with the interface in order to assess their performance. The evaluation provided is called Learner's diagnosis, and represents the basis on which the Knowledge Module builds an adaptive instruction/curriculum. According to Polson & Richardson (2013), the diagnosis is based on three dimensions:

- The bandwidth. The bandwidth represents the width of the information about the student that the system has access to. It may be at three different degrees: mental state, when all possible information are available; intermediate state, when almost all the information through the interactions is available; final state, when only the last student answer is available to be processed by the system.

- Type of knowledge. The knowledge type of the subject to be taught. It can be distinguished into three different types: flat procedural, when the knowledge is only procedural and concerns learning how to execute a specific task; hierarchical procedural, when it is procedural but necessitates the completion of intermediate subgoals that are required in a predefined sequence; declarative, when it does not concern task execution but rather learning a concept, a principle or a fact.

- Student-Expert difference. This dimension concerns the gap between the student model and the expert module representation. The degree of complexity the system can have is tridimensional:

  1. The system is able to recognize the learner's missing knowledge;
  2. The system is able to categorize and infer the learner's incorrect information;
  3. The system is able to map the student's errors in a dynamical way.

3.2.2.3 The Instruction Module

The Instruction module, or the Pedagogical module in Dede's notation, ensures that the information is deployed to the student in the more efficient way. The structured material, that will be the outcome of this module, must have this three attributes:

- Flexibility: every instruction can be adapted to multiple purposes and logical order;

- Structural transparency: every instruction must be self-explanatory and structurally identical to the others;

- Individualization: every instruction must be implemented in different ways that can be deployed to different kind of users.

The instructions should be implemented in a modular logic, dividing the material into units with small instructional goals and ensuring that in any case each unit achieves the goals.

3.2.2.4 The Environment

The Environment is the part of the system specifying or supporting the activities that the user does and the methods available to the user to do those activities. Introduced in the fourth modules model (Dede 1986, Ma et al. 2014), it includes both the User Interface and the more general setting in which the user is supposed to access the platform.

3.3 The application domains of Serious games and Intelligent Tutoring Systems

3.3.1 Well-defined and ill-defined domains

While nearly all of the former literature on serious games and Intelligent Tutoring Systems regarded building systems for well-defined domains, where there was a black-and-white distinction between what was wrong and what was right, one of the main challenges of designing tutoring systems regarded the need to structure platforms able to face the problem of the multiplicity of possible student behaviour that have shapes of correctness which may vary according to the concept, the contest and other involved agents. E-learning tools for ill-defined domains differ on the former systems mainly on the following points (Thomas et al. 2013):

- It is needed a certain amount of information regarding the student in order to retrieve the correct diagnosis and assessment;
- The system's feedback on the correctness of the answer is delayed until a partially complete student model has been made.

Most of the tasks posed to the student do not have a defined answer and this ambiguity is reflected in the complexity of the interaction between the tutor and the student, the need of shaping the student model through time and the constant endeavour through the interaction to catch the granularity of the aspects which determine the behaviour of the user.

3.3.2 Diagnosis and adaptation

As mentioned earlier, the diagnosis of the student knowledge must take into account several dimensions and information which must be gathered throughout the interaction of the user with the interface. In order to process and categorise these inputs, there are some commonly used heuristics which allow an assessment and a prediction of the student knowledge.

3.3.2.1 Rules-based models

Rule-based ITSs are based on the idea of establishing a set of rules that a student should follow while finding a solution to the problem presented. Properties of rule-based systems, according to Hayes-Roth are (Hayes-Roth 1985, Kadhim et al. 2014, Xavier et al. 2015):

1. The ability to incorporate practical human knowledge in conditional IF-THEN rules;

2. Their skill increases at a rate proportional to the enlargement of their knowledge bases;

3. They can solve a wide range of possibly complex problems by selecting relevant rules and then combining the results in appropriate ways;

4. They adaptively determine the best sequence of rules to execute;

5. They explain their conclusions by retracing their actual lines of reasoning and translating the logic of each rule employed into natural language.

Every learner error should be predicted according to this model, and the system can provide a feedback which helps the user identify the right steps to follow (Campuzano et al. 2014).

3.3.2.2 Case-based models

In case-based models, the student's errors are not coded into IF-THEN rules but in more meaningful cases, which are more specific, related and reveal a strategy. The cases are built gaining data from a large amount of learners with the help of machine learning techniques and each case is built according to the sample of students collected.

3.3.2.3 Constraints-based models

The concept of constraint-based systems lies in Ohlsson (1987) theory of the difference between declarative and procedural knowledge, in which the former indicates the pure knowledge of the subject while procedural knowledge allows the generation of actions. The declarative knowledge is important to evaluate the consequence of the actions, and therefore to expect a certain outcome. According to this model, the only way for the user to learn from mistakes is to create constraints on the possible correct answers, generating set of rules that give a feedback in order to restrict the range of the declarative knowledge of the user. The main difference between rule-based systems and constraint-based systems lie in the fact that the latter concentrates on the obtained result in each of the steps to reach the solution to the problem and does not require a global Expert Module to process the sum of the information required. Assuming that the constraints are respected, every solution that solves the problem correctly can be considered satisfying, and, for this reason, each step can be auto-conclusive, denying the scope of a wider goal. This is the main reason why Constraint-based models are usually applied in ill-defined domains.

3.4 Machine learning approaches to technology-enhanced learning

3.4.1 Bayesian Networks

Regarding the integration of Expert systems into serious games, there are many examples in literature of researches which used AI techniques as Bayesian networks, Neural networks and Fuzzy sets to allow the systems to be autonomous, adaptive and

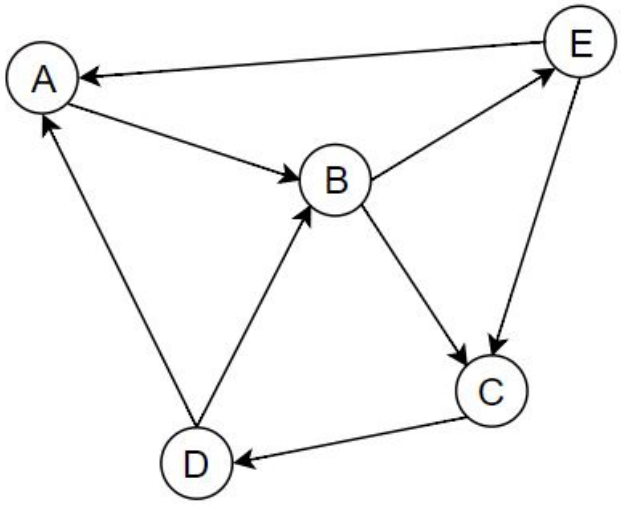

Figure 3.1: Structure of a basic bayesian network.

"smart" in the concept as we know it. Costa and colleagues used Bayesian Networks to design a system which possesses an Open Learner Model, in the sense that it allows the student to be involved in the creation of his own Module (Costa et al. 2012). This approach leads to a more dynamical interaction between the Instruction Module and the user, in order to improve the efficiency and the time needed for a reliable assessment. A bayesian network is a directed acyclic graph where the connections between the nodes are modelled by estimated conditional probabilities; a basic bayesian network is displayed in Fig. 3.1. The system in Costa et al. (2012) uses two Bayesian networks which are structurally identical, one to compute the learner's belief in his knowledge (Ms) and one to compute the tutor's belief in the learner's knowledge (Mt). Each of these has two parts, the domain-general circuit, which computes the general belief throughout the whole interaction, and the task-specific circuit, which is created at run-time and concerns the problem faced by the student at the moment. When the student has finished solving a task, the probability and data obtained are directed as input to the domain-general circuit, modifying the structures of the nodes. Every node of the networks have two values: mastered or not-mastered; the weights are established within 5 possible degrees of knowledge. At the end of each task, instead, the learner is asked the belief regarding his or her degree of knowledge. If the beliefs contained in

Ms and Mt differ, the tutor starts to negotiate with the student in order to find an agreement. The negotiation process can include tasks like proofing, by solving or analysing problems posed by the tutor, that the tutor's belief is wrong. If the learner believes he or she has less knowledge than estimated by the system, a support strategy will be engaged in order to convince the learner that is more proficient than believed.

3.4.2 Neural Networks

The concept of artificial neural network stems from an attempt to recreate with an algorithm the connections among neurons that happen within the human brain, with the purpose to create an architecture that is able to learn and be trained as similarly as possible to its biological reference. An artificial neural network consists of a set of input units x , a set of outputs y and a set of hidden layers l containing hidden units h . The connections among layers are defined by weights w that can change upon training. For a fully connected feedforward neural network, a hidden layer unit value is computed by:

$$h_j = f_1(\sum_{i}^{l_h}(W_{ij}x_i + b) \tag{3.1}$$

where f is the activation function of the single unit and b is a bias unit. The most common activation functions in a standard feedforward neural network are sigmoid and tanh. Output value instead is computed as follows:

$$o_k = f_2(\sum_{j}^{l_o}(W_{jk}h_j + b) \tag{3.2}$$

A schematic representation of a feedforward neural network is shown in Fig. 3.2. Only few authors in literature used neural networks to predict student models based on previously assessed learners on ill-defined domains. This mainly happens because, even if neural networks usually perform well at both categorising unpredictable situations and learning from data, it is not easy at all to retrieve free available datasets in order to train them. This situation is slowly changing especially since the growing interest in big data science, and famous MOOC (Massive Open Online Courses) websites are collecting

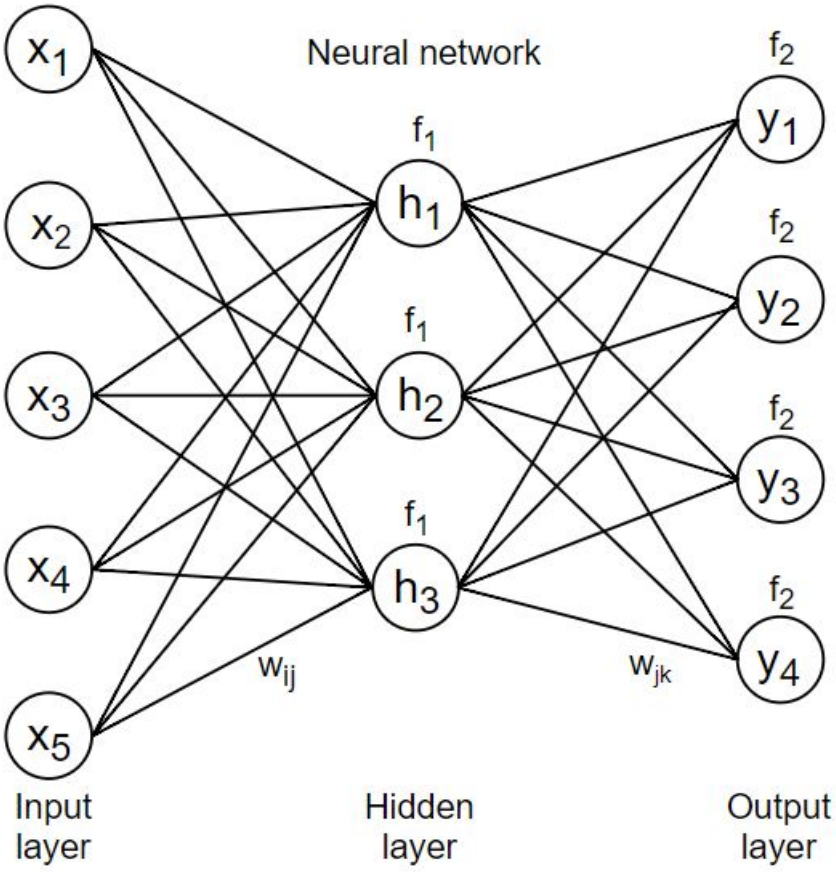

Figure 3.2: Feedforward fully connected artificial neural network.

learner's data to implement adaptive tutoring systems. However, it is still needed a huge amount of work in order to give the right inputs and obtain generalizable data out of these kinds of complex network architectures. One of the first approaches used a C programmed feedforward neural network with three layers to compute subtraction problems (Mengel & Lively 1992). The neural network was trained using an already available set of records. A more recent approach by Wang and Mitrovic describes the use of data collected by SQL-Tutor, a constraint-based model tutoring system, to train a neural network in order to map student errors (Wang & Mitrovic 2002). The neural network was feedforward and its input layer took the following parameters:

- the time needed by the student to reach the solution;
- the level of help requested by the student;
- the level of complexity of the posed problem;
- the current level estimated of the student.

After the training session, the network was used to predict the behaviour of the students. It was found that six submissions of the student answers were requested in order for the network to assess his behaviour properly and to select an appropriate problem afterwards. Lo and colleagues recently used a multilayer feedforward neural network to identify the users' cognitive styles in a web-based learning environment (Lo et al. 2012). This methodology represents another trial to gather information about the learners' psychological traits in an unobtrusive way, without the use of explicit psychometric data and thus without separating the assessment from the learning process itself. The designed system was able to analyse the user browsing behaviour, to infer a cognitive style out of it and to organize the educational path according to it. Since the importance for the authors to design a system that collects information about the users in a totally implicit way, there was no need for a login and only the material for each session was stored, this meant that when the user disconnected and then tried to connect again, it was not possible to recover the previous session. The model of

cognitive styles adopted in this work is the one by Myers-Briggs, which is based on Jung's theory of cognitive types and has four possible profiles: Interpersonal, Mastery, Understanding and Self-expressive. In the first part of the project, the authors collected data to investigate the relationship between the browsing behaviour and the cognitive styles and then used these data to train the neural network. The neural network was three-layered and its input neurons received data from:

- Ratio and frequency of the selection of contents and interactive components;
- Average staying time on content and interactive components;
- Selection ratio of content link types;
-

The output units were four and the activation of each of them returned a different profile. The approach used by the authors is extremely interesting, though their results cannot be generalised for several reasons: the students used for the experiments were not in the same classes and had different teachers; there is no demonstration of an increased attention paid to the contents when the system adapted to the identified style, but only of a very low increase of the time spent on the webpage. Also, time spent and attention are a measure of motivation, self-efficacy and other psychological concepts that could act as intervening variables on the outcome and add a bias on the cognitive style inference, greatly reducing the reliability of the system (Grondin 2010).

3.4.3 Fuzzy sets

Fuzzy sets are a powerful tool for overcoming the limitations of classical sets and allow a higher degree of freedom in classification tasks (Jeremić et al. 2012). Among the main conceptualizations of the fuzzy logic, Zadeh lists some of the features and properties this logic has (Yager & Zadeh 1992):

- Exact reasoning is viewed as a limiting case of approximate reasoning;

- Everything is a matter of degree;
- Any logical system can be fuzzified;
- Knowledge is interpreted as a collection of elastic or, equivalently, fuzzy constraint on a collection of variables;
- Inference is viewed as a process of propagation of elastic constraints.

The use in literature of fuzzy inference systems is wide, especially concerning ITSs.

A recent work by Hsieh and colleagues aimed at building a system to support foreign languages learning for Taiwanese people (Hsieh et al. 2012). The system aimed at suggesting tailored English article by using the data gathered through multiple interactions with the user. Fuzzy sets and memory cycle updates are used in order to build the learner's preference. A tutoring system was implemented through an analytic hierarchy process to provide a motivating environment for English learners. The analytic hierarchy process is a technique that allows to structure a decision problem into subgoals that constitute a hierarchy, and uses mathematics to formulate judgements on the relationships among such goals (Saaty 1980). Both the ability of the user and their vocabulary cognition were evaluated. Right after reading a new article, the user was asked to answer to a questionnaire about motivation, satisfaction and feedback. The learner feedback about the understanding of the suggested article was collected using a five point Likert scale. Additionally, a test asked about the newly learned words in order to help the memory to retain the information. The next suggested article took into account both new words and previously learned words, to efficiently improve the volume of the vocabulary of the learner. To calculate the volume of the vocabulary of a new user, they were asked to answer a set of 10 questions using the Computerized Adaptive Testing (CAT) approach. Tests based on CAT are able to estimate the ability of the learner using the rate of correct and incorrect answers against the difficulty of the items, and to use an item selection algorithm in order to provide an adaptive testing experience (Wainer et al. 2000).

Most of the actual platforms use stereotype-based approaches in order to initialise the tutor in his first approach with a new user. This means that the first interactions with the student will always follow a predefined pattern and, once the system has collected enough data to build a student model out of the user, then the process of modifying the existing stereotype starts. This may lead to problems as lack of flexibility and lack of creating new stereotype by learning from the interaction with users. Another modern approach to intelligent teaching in software pattern design is DEPTHS (Jeremić et al. 2012), a platform to build an educational path for each student taking into account their ability, background and characteristics. The approach used by the authors is the overlay-based model. According to this approach, the student model is a subset of the general learner model and can be just a part of it (in the strict overlay approaches) or a misconception of it (in the extended overlay approach). In DEPTHS, each domain concept is associated with the student's knowledge status in relation to that concept. The computation of this state is based on fuzzy sets and certainty factor theories. Curriculum planning is computed by Jess, the inference engine, which provides the right concept to follow next according to four rules:

1. student's current knowledge state should be the same as or lower (if possible) than the minimal knowledge level predefined;

2. for learning that concept;

3. the concept should not be learned earlier;

4. all prerequisite concepts should be already learned;

5. the concept should not be already in the concept plan.

A Diagnostic engine uses a set of pedagogical rules and domain knowledge in order to diagnose test results and infer the student's knowledge level based on these results.

3.5 Conversational Intelligent Tutoring Systems

Conversational Intelligent Tutoring Systems (CITSs) are platform which aim at replicating the interaction of students with real teachers, as in classrooms, with the introduction of natural language. The development of such platforms is extremely time-consuming and the speech recognition must be at a high level in order to prevent the student to abandon the platform for lack of motivation. For these reasons, there are only few platforms which use conversational agents to communicate with the user and/or to provide an assessment. An example of this kind of platforms is OSCAR (Latham et al. 2012). Oscar Conversational Intelligent Tutoring System is an ITS developed in order to implicitly predict and adapt to the user's learning styles using natural language recognition. The learning style characteristics are taken from the Index of Learning Styles (ILS) model by Felder and Silverman, according to which the learning style of a subject can be described by the score obtained in a dedicated 44-items questionnaire along four dimensions: perception (sensory vs. intuitive), input (visual vs. auditory), processing (active vs. reflective) and understanding (global vs. sequential) (Felder & Silverman 1988). The model also describes the typical behaviour expressed by each prevailing style. The aim of OSCAR CITS was not to just predict and assess the user learning style, as already done in Cha et al. (2006), but to imitate the behaviour of a human tutor during a conversation and presenting the material tailored to the learner, giving adaptive hints and analysing the problems posed. The CITS architecture detected and categorised behavioural cues using a set of 33 logic rules to apply to the information gathered. The first implementation was used to teach an online course of SQL and .NET framework.

A more complex and complete system is INES (Mikic Fonte et al. 2012). The INtelligent Educational System (INES) is a prototype of an e-learning platform which has the aim to deploy information, manage multiple users, assess the learners, guide the learner and grasp natural language speech using multiple virtual agents. In order to do that, the platform uses the following set of tools:

- BDI technology agents, according to the BDI model. The BDI (Beliefs, Desires, and Intentions) model is based on the idea of logical agents which possess the knowledge of the actual state (the Beliefs) and has a set of goals to achieve (Desires). If the knowledge of the current situation differs from the one expected by the set goals, the agent has a list of plans to execute in order to try to reach the desired state. Multiple and different agents compose the ITS of the INES platform:

- Jadex platform, a tool which allows easy programming and initialization of virtual agents following specific features and variables set by external coders. The agents created by Jadex are known as Goal Oriented Agents, and the possible actions they can execute are written in Java and XML;

- CHARLIE, a chatbot which allows natural speech recognition.

In particular, the structure of INES provides that, for every new user connected to the platform a new instance of a BDI agent inside the Expert Module of ITS is created, which dynamically starts adapting to the user and stores the information gained. The user, throughout the interface, does not communicate directly with the system. First of all the natural language processing goes through a first agent, a chatbot, which analyses the words and the concepts used by the user. This information is then delivered to a particular BDI agent called EMMA, which elaborates the information and computes whether to send those to the Expert Module for a decision-making process or send a response back directly to the user without a higher level processing. If a higher order message (as "the user has completed the session") must be sent to the user, it overrides the messages sent by the other two agents. If the information gained by EMMA needs a different level of processing they are sent to the BDI agent correspondent to the user inside the Expert Module, which is called ISMAEL (Intelligent System Manager Agent for E-Learning). The instance of ISMAEL for each user contains their educational progress, their credentials and personal information. The whole platform organization and information exchange is able to plan the educational path

for each user and suggest them personalized tasks to achieve specific learning goals. What lacks in these kinds of systems, as already evidenced by the authors, is that the system seems not to take into account emotional aspects of the user interacting with these agents. Another problem is that, although advanced, implementing a chatbot mostly results in a limited interaction that prevents users to fully express themselves, since even the most complex chatbots are currently very limited in their performances. What really is the turning point is using natural language processing not just to allow a more natural interaction with the system, but to improve the user model, analysing the utterances to infer more about the learner and his or her style. This has been the main drive of the development of the natural language architecture in the Enact platform, as will be shown in the last chapter of this work.

3.6 Additional embedded tutors

There are other examples of intelligent tutoring systems that are embedded in more complex video games, simulation games, role-playing games and MOOC platforms. The Interactive Strategy Training for Active Reading and Thinking-Motivationally Enhanced (iSTART-ME) tutor, for example, is a recently developed game-based platform to enhance self-explanation of science text in young adolescents (Jackson & McNamara 2013). This serious game is built on top of an existing ITS which had the same aim, but that required too long interactions in order to have long-lasting results, thus reducing the motivation to stay. There are several modules in this platform:

- the Introduction Module, during which three virtual agents introduce the concept of self-explanation by making examples and providing strategies for an efficient reading;

- the Demonstration Module, where two characters demonstrate self-explanation by reading science texts and the student must guess the strategies used in these examples. A tutoring character gives a feedback about the self-explanation given by the other virtual agent;

- the Practice Module, where the student is asked to use the acquired knowledge to use self-explanation to understand the proposed science texts.

Another example of the gamification of an ITS for children is described in the work by Sandberg and colleagues (Sandberg et al. 2014). In the work they presented, it is described the development of a motivationally enhanced version of the already existing MEL platform, an application to improve the native English speaker's vocabulary of the animal and continent words. There were 25 stimuli of animals and 5 continents, along with embedded videos and short stories about animals. Several kinds of tests were presented to the children, among which multiple choice quizzes, puzzles and spelling questions. The enhanced version of the application contained two different game modes:

- Zoo Animals: there was a storyline in which the animals escape and need to be found back.
- Neighbourhood: items that were needed to be found in a collage.

The authors tested three hypotheses:

1. Children who used MEL-enhanced outperformed those who used the simple version of MEL;
2. Children who used MEL-enhanced learned in a more efficient way than children who use MEL-original;
3. Children who use the MEL-enhanced spent significantly more learning time on the platform.

While the hypotheses 1 and 2 have been verified, the third revealed to be false.

Another recent platform which integrates multiple tools previously mentioned is the architecture described in Giuffra & Silveria (2013). This model wishes to integrate a Virtual Learning Environment and ITSs techniques to create responsive agents which

can guide the learner throughout their interaction with the teaching material. The tool used for the Expert Module form is Moodle, a well-known and common platform to build MOOCs, interactive teaching environments and it is fully customisable, allowing researchers and teachers to save log files of the users. This platform is massively used by institutional websites. The Pedagogical Module is held by a virtual agent called "Bedel", which accesses the database in order to retrieve relevant information to get the material sorted. The Instruction (tutoring) module is set into the form of an agent called "Tutor", which takes the information both from the database and from Bedel, allowing it to provide suited and efficient educational path to the learner. Both the agents are coded using the already mentioned BDI model and the Jason platform, a program in Java.

In the actual context, where most of the platforms are extremely well designed, implemented, and rich in their potentials, like in the case of customisable online scenario editors, game environments that allow to practice interpersonal skills and communication styles, what really emerges is the lack of standardised technologic tools to assess (neuro)psychological characteristics. Paper and pencil tests are still widely used even by human resources managers, though these kinds of tests require time to be completed, time to be interpreted, money, double processing, the opinion of experts and they result in a slow candidate selection and assessment. The present work aims at describing a 3D environment that overcomes these problems and where users can receive scientific, standardised and reliable profiles of their negotiation skills in a quick and handy way. The advantage of using three-dimensional models and environment is twofold: on one hand, 3D models allow a higher spatial resolutions of the gestures, facial expressions and gaze of the avatars, considering how crucial non-verbal aspects are in the context of soft skills and negotiation (see section 4.2.2 of this work); on the other hand, 3D environments increase realism and immersion in that they resemble familiar daily places and activities. The ENACT platform, as will be described in the next sections, is composed of two main features:

- A virtual agent which can interact effectively with the user, answer according to their responses and have a dynamic behaviour;

- An intelligent tutoring system, that evaluates the interaction and the performance of the user within the scenario and gives a customized profile; other than that, it allows the user to navigate through the history of his or her interaction, in order to provide a molar and a molecular view of the choices taken by the two agents.

The virtual 3D agent is able to express a range of basic emotions in an effective way using verbal cues, as vocal tone and structure of the sentence, and non-verbal indicators, as facial expression, eye contact, body posture and gestures. The agent possesses a series of internal states which vary during the progression of the interaction and influence its behaviour. The tutoring system is implemented to intervene at the end of the game session and provide useful information to the user about his behaviour related to the bot plus more static traits. The user will be given a profile about his negotiation skills and tips about how to improve the efficacy of communication styles. The potential of the platform developed by ENACT is dependent on four primary objectives:

- The development of a 3D game-based platform to assess the user's negotiation and communication skills in realistic scenarios during the interaction with an Intelligent bot;

- The development of an Intelligent Tutor able to provide a personalized training on the user's negotiation skills;

- The implementation of a system to provide a reliable and scientific-based assessment with the use of modern psychological models;

- The development of a free, cross-platform software, handy to access for wide and heterogeneous groups of users.

3.7 Serious games and technology-enhanced learning for soft skills

As previously mentioned, there is no formal or unique definition of soft skills. A possible attempt of a comprehensive definition that embraces what is intended for soft skills in the extensive literature on the concept could be that soft skills are not domain-specific and are both self and people oriented and represent an inseparable addition to hard technical knowledge that allows the completion of activities and accomplishment of results both on the job and in private life (Kantrowitz 2005). The continuous innovations in advanced learning technologies have provided opportunities for enhancing traditional methods of soft skill training, adding the possibility to gain quick, reliable and low-cost personalised assessment using portable devices and globally accessible software. In fact, the new web technologies have allowed to rapidly and effectively reach geographically sparse learners and group them into social communities, creating new means of communication and space for peer-to-peer comparison and evaluation. For this purpose, many different computer based learning environments have been developed to promote soft skill development, in the form of standalone tools or as features of larger systems. As an example, Small business Game is an online interactive simulation that can be integrated into ICT, Business and Finance courses, that helps players develop interpersonal and intrapersonal skills related to the management and setup of a small business (http://www.thesmallbusinessgame.co.uk/).

Unigame, instead, is a web based game that can be used by teachers within lifelong learning frameworks to support various areas of domain, as part of face-to-face teaching activities (He et al. 2010). The final aim is to share and provide knowledge within a specified topic area by interacting with other users and different team groups using private or public discussion forums, allowing audio, video conferences or written chats. The nature of the interactive experience helps give confidence and improve problem solving, communication skills and teamwork.

The BuT Courseware (Rogmann 2008) is integrated into university programmes as a preparatory module to be completed prior to the commencement of the study curricu-

lum. The course material contains a series of video lectures that show real life cases in which problem solving and conflict resolution techniques are used to effectively resolve situations. In this case, the presence of an expert trainer to guide and support learners is crucial to guarantee a successful training.

Among the game-based e-learning technologies for soft skills development, instead, they can be distinguished between team and single player platforms. At-risk and E-circus ask users to interact with virtual characters controlled by a computer programme (Aylett et al. 2006). In At-risk, users, by playing an online role simulation game, can build interpersonal skills and learn to effectively communicate about behavioural and health issues. It is designed to help university employees to recognise the signs of psychological burnout as depression or anxiety and identify students at risk. Role play simulations introduce players into 45 minutes conversation with an a 3D avatar that possesses its own story and personality. E-circus is a single virtual role-play animated by "synthetic characters" that proposes to effectively establish empathic relationships with the learners. It has been designed with a particular focus on children and young adolescents in schools and its purpose is to be used in combination with other educational methods. Another e-learning tool in the form of a game, Eduteams, developed by TPLD, is an online avatar based software that allows to design personalised training scenarios in the form of single or multiplayer games and 2D and 3D environments. The aim of Eduteams is to provide students with an environment where to experience different situations to practice their social skills.

Eutopia (Miglino et al. 2010) constitutes an example of a multiplayer platform that provides role-play simulations focused on the development of soft skills. The aim of this tool is to represent a training system able to promote the improvement of mediation skills. In particular, the goal of the training is to allow users to get in touch synchronously in verbal and non-verbal modes within a predefined simulated scenario and test their communication strategies. In the actual context, where most of the platforms are extremely well designed, implemented, and rich in their potential, like in the case

of DOKEOS Game (http://www.dokeos.com), a game environment which allows to build fully customised scenarios where to practice interactive skills and communication styles, what really emerges is the lack of standardised technologies that allow to precisely assess psychological characteristics and abilities.

All the previously mentioned e-learning tools are able to allow the learner to practice in safe and customizable scenarios, either individually or within peer groups, but they do not base their training methods validated on scientific theories and standardised models, therefore this leads to two major issues:

1. it is possible to provide an effective training but not to provide a former unique and precise assessment of soft skills, since it is not possible to refer to a specific psychometric framework;

2. it is not possible to evaluate the real results of the training using standardised psychological tests or compare the outcome with those of other tools.

The present work, along with the development and enhancement of the Enact platform, aims at overcoming these difficulties by integrating both the customization and the validation aspects and, after a series of experiments to evaluate its reliability, being able to provide a training and assessment fully comparable to that of standardised psychological tests.

3.8 Conclusions

In this review the actual state of art in literature of serious games and Intelligent Tutoring Systems has been explored, underlining which of the early systems' features have evolved through time and which techniques are the most efficient and used in literature, including machine learning and artificial intelligence systems. The actual gaps between technology and theory have been highlighted in order to identify the possible directions of future research in the area; huge and fast progresses have been made since the first Intelligent Systems, but there is still plenty of dark corners in need to

be explored, some of which will be shortly summarised below, in particular the lack of platforms which investigate in a reliable and scientific way the same category of variables currently assessed by paper and pencil psychological tests as well as the lack of a standardised and agreed framework for the assessment of soft skills, which are typically distinguished by their undefined boundaries. Other open challenges involve the ability to find models in which to code and frame ill-defined domains, which include the need of creative solutions in problem solving, facing unexpected situations, deal with overlapped contents and definitions, and learn from the experience with users. In this direction many developers are extending their user data analysis techniques from psychometrics to cloud computing and machine learning, allowing a deeper understanding of the user behaviour and a more meaningful use and access of many related variables whose real correlation is yet unknown.

Chapter 4

Enact Game: a serious game for practising and assessing negotiation skills

The implementation of the Enact software, a free and portable tool for the assessment and training of the user's negotiation skills according to Rahim and Bonoma's theory, is described in the present chapter. Prior to the design of the platform, need analysis is conducted on the target groups the game is mainly addressed to, with the scope of investigating the concept of negotiation, the difference in the perception and use of technology and the actual need of a technology-enhanced learning tool for practising soft skills in several context and geographical areas and age groups. Results underlined the need of a higher awareness about the importance of soft skills and negotiation abilities, and a need of teaching and assessing these abilities for both young and adult, in different contexts. The development of the various features of the software is then described, in particular the assessment and training environment, the integration of the tutoring aspect and the design of the scenarios, the virtual agents and their overall preliminary evaluation. The testing of the platform in its first release is then presented, including the details of all components and how they were evaluated in the trialling phase. The platform usability, consistency and reliability is also assessed in comparison with both traditional video games and traditional psychological tests. The results presented in this chapter, that showed the importance of integrating a model of the Compromising profile, the need of improving the scenarios according to the reliability obtained in the validation and the need to remodel the aspect and graphics of the environment and of the virtual agents, constitute the guidelines for the release of the second version of the platform.

ENACT is a 3D serious game focused on soft skills that allows an adaptive user assessment and training of negotiation skills and conflict management as defined by psychological theories with the use of virtual agents and simulated environments. The game is divided into scenarios, each dealing with a real-life situation, in which the user is asked to play the role of one of the two characters in the scene. The game targets dyadic negotiations because it aims at training and assessing the outcome of

conflicts that involve convergence, i.e. discussing a decision on a problem that involves many solutions, and divergence, i.e. the distribution of a single resource between two contenders, and also aims at investigating the direct effect on the user of the interaction with each of the possible negotiation styles that can be adopted in the process. This chapter describes the steps taken to design and implement the platform and how Rahim's theory of negotiation skills was modelled and introduced within the game variables. A training need analysis has been conducted by the Project Partners in order to establish the needs of the target groups Enact was mainly addressed to. The interface of the assessment and training environment are described, along with the methodological choices that led to the structure of the scenarios and of the interactions between the agents and the user within the game. Then, the experiments conducted to test the platform, its efficiency and usability are presented. These include the release of the first version of the software, its evaluation in terms of usability, and its validation with the help of the partners of the ENACT Project. Among the other results, the platform shows satisfying consistency with the ROCI-II standardised test on the dimensions of Dominating and Integrating style. The results of the analysis of the data obtained by the usability evaluation and validation phase of the first release is then used to improve the platform.

4.1 A Training Need Analysis

The Training Need Analysis (TNA), conducted by all ENACT Project partners in Italy, Spain and Turkey, aimed at identifying the views and opinions of Enact's target groups concerning the concept of negotiation and process of handling conflict. The findings are informative about the process followed during the design and implementation of the Enact platform and its interactive scenarios. In particular, the scope of the TNA is:

- To investigate the level of technology in schools and institutional environments;
- To investigate the level of technological and general ICT (Information Communication Technology) knowledge and savviness of the teachers and trainers in

various contexts;

- To evaluate and identify how the concept of negotiation skills is considered in different contexts and what initiatives are proposed and developed for practising them;
- To evaluate how would be considered the idea of a role-play game and simulation tools as new methodologies for practising and assessing soft skills in schools and sport contexts;
- To investigate the possible benefits of the use of role-play games and simulations in small and medium enterprises or in other business-driven environments.

4.1.1 ICT and negotiation practices for educational purposes

The importance of ICT as a tool to reach educational aims is well-known in primary schools (Moseley et al. 1999). In Italy, since the first year of primary schools it is possible to find indications about the use of technology-enhanced educational tools, along with the importance of acquiring early digital competencies, while in secondary school it is a cross-curricular subject. Secondary and high schools are usually provided with computer laboratories where learners can access and familiarize with different types of technological devices like laptops, desktops or tablets and it is also possible to find interactive whiteboards. Unfortunately it is not always possible to guarantee an immersive experience of ICT technologies in all schools because teachers and principals may be unsupportive or not qualified (Polizzi 2011). Most soft skills and negotiation skills training and assessment tools are developed by training organizations specifically employed by private companies or by universities. These include Massive Open Online Courses (MOOCs), role-play simulations, serious games, but also one-to-many traditional courses.

A similar situation sees Spain promoting ICT technologies for children and teachers since early schooling years, where whiteboards and PCs increase from year to year (Buabeng-Andoh Charles 2012). Traditional learning tools are replaced and comple-

mented by electronic devices. The importance of ICT as an educational methodology is a framework applied not only to school contexts but also in professional ones, such as in small and medium enterprises (SME). Soft skills and negotiation skills, in particular, are especially valuable in business contexts but it is not always possible to find appropriate training opportunities for employers and employees.

In Turkey, there is no report of specific training practices for soft skills or negotiation competencies in schools or other educational contexts, even if a general interest in the adoption of e-learning and educational technologies has been expressed by the Turkish Ministry of Education and by the Ministry of Youth and Sport, among the project partners, which underlines the importance of enhancing soft skills as a mean to improve inclusion and fair play among young adults in sport environments (Akkoyunlu 2002).

4.1.2 TNA Interviews and Results

The TNA was conducted to explore the different points of view and negotiation experiences, as well as the training design, development and delivery of Enact's learning scenarios for the identified target groups. Specifically, the TNA was focused on the involvement of the following target groups and for the following investigation purposes:

- Middle school students (Italy), in order to evaluate the perceived importance of negotiation at early stages of life and before the beginning of a professional career, to identify educational gaps and assess how technology can help fill them;
- Researchers/teachers (Italy), to investigate the use of technological tools in any experience of teachers training realized and its impact on soft skills development, in particular on negotiation skills development;
- Managers and employees of small and medium-sized companies (Spain), since negotiation skill plays a vital role for building relationship and communicating effectively, and the ability to negotiate can strongly impact success and failure in today's business world;

- Young and Athletes (Turkey), since Turkish Ministry of Youth and Sport has underlined the importance of sports to improve negotiation skills at both physical and psychological level, and generally sport can help enhance the ability to assess when to compromise and supports the need of feeling part of a team.

The data collection was conducted with the following tools:

1. the administration of an exploratory questionnaire to better define the context of the research;
2. 10 semi-structured interviews in depth (questions are listed in Appendix A);
3. the administration of an online questionnaire completed by 55 users (questions are listed in Appendix B);

All the recruited participants were volunteers, and in all cases the research advertisement, first contact and approach was done by email. The exploratory questionnaires made it possible to gather information on existing procedures and practices for education, development and professional training of soft skills and negotiation skills, as well as the level of dissemination and use of ICT in the different target contexts. In all countries, the research has highlighted the existence of a gap between the educational practices recommended in the national guidelines and the actual educational practices, use of technologies and teacher training. The presence of training courses on soft skills for teachers and/or students was found in none of the three partner countries. Concerning the presence of projects to encourage the development of negotiating skills as well as other soft skills, only Italy, in literature, showed clear plans for the development of innovative training tools, such as role-playing games and simulation tools for the development of soft skills (Miglino et al. 2010).

The results of the 10 semi-structured interviews highlighted the aspects that the various target groups consider crucial in their definition of negotiation. Each partner involved into the pilot experimentation was asked to interview at least three participants and

no more than five, as representative of each specific target group. The 10 interviewees were distributed as follows: 3 Spanish managers; 3 Italian teachers and 2 Italian students; 2 Turkish young athletes and 1 Turkish coach. The concept of negotiation showed to be very context specific. Concerning the economic context, the Spanish managers stated that there are two needed abilities for a successful negotiation: problem solving and time management. In Italy, instead, teachers define negotiation as the ability to manage conflict effectively and positively, while students state the quality of the interpersonal relationship between the parties is the key of an effective negotiation. For the three Turkish respondents, negotiation comprises persuasive ability, awareness of one's own resources and personal skills.

The results of the 55 online surveys revealed the importance of a successful training and in particular of active, rather than passive, learning methods, where the learner does not only listen to information and theories but has the chance to apply and practice them, such as in role-play games and simulations, for the development of communication and negotiation skills. Each partner involved into the pilot experimentation was asked to invite at least ten participants to complete the questionnaire. The participants who completed the survey were distributed as follows: 24 Italians (20 students, 3 teachers, 1 trainer); 21 Spanish (10 managers and 11 employees); 10 Turkish (5 students, 4 managers and 1 employee). More specifically, for Italian subjects, the good use of critical thinking (64%) and active listening (64%) are the skills that allow to choose the correct negotiating attitude. Furthermore, negotiation skills are considered fundamental for managing situations that involve both school and professional life, as well as the private sphere (33%). For the Turkish participants belonging to the sports field it is communication (80%), together with problem solving (50%), empathic ability (50%) and creative thinking (50%) that play a key role in the negotiation process. Moreover, for 60%, the negotiation seems to be more important for the professional life than for the private one (30%). For the Spanish respondents the negotiation process is associated with three fundamental social skills: effective communication (90%), empathy (70%) and decision making (65%). The ability to negotiate seems to be instead perceived

Nationality	Most considered as key negotiation factors
Italian	Critical thinking, active listening
Spanish	Communication, problem solving, empathic ability, creative thinking
Turkish	Effective communication, empathy, decision making

Table 4.1: Summary of the most important negotiation factors considered by the participants of each nationality.

Nationality	Situations that negotiation is considered fundamental to manage
Italian	School/professional context, private context
Turkish	Professional context only
Spanish	Professional context, social relationships

Table 4.2: Summary of the situations where negotiation is considered fundamental by the participants of each nationality.

as a dimension more linked to performance on the job rather than to social relations (81%). The results are summarised in Tables 4.1 and 4.2. Considering the common belief that the act of communicating effectively (sharing information and states of mind, having clear objectives to be reached, actively listening, giving and receiving feedback) is the tool through which a successful negotiation takes place, these results, complementary with Rahim's theory, drove the first design and implementation of the Enact platform, which will now be presented.

4.2 The Enact Game interface

The Enact platform has been designed to bring Rahim's model into a 3D virtual environment. The game is organised in scenarios, each independent from the others, where the user plays a different character and negotiates with various virtual agents in reality-based everyday life situations. The main game components are:

- the User Interface, that connects the user with the underlying algorithms and logic of the Assessment and the Training environments and that allows the player's data to be stored and retrieved;

- the different 3D virtual characters inside each game scenario, who represent the interacting agents of the software, are anthropomorphic and play a specific role;

- the Assessment environment (developed in collaboration with the ENACT Project partners), that provides an evaluation of the player's negotiation skills by detecting the features that characterize each style;
- the Intelligent Tutoring System, core of the Training environment, that uses the participant's data to create the User profile and define the most useful path for training and practising negotiation styles in different scenarios, by manipulating the scenario variables like time, importance and relationship with the other agent.

The details of each component will be described below.

4.2.1 The interface of the game

The User Interface is designed to be as intuitive as possible, and its usability and intuitiveness was repeatedly evaluated during the trialling (Marocco et al. 2015). The game has been accessed both from browsers, through the Unity3D player plugin and from the executable compiled with the graphic engine. The game is cross-platform and is currently accessible on the Enact game website: `http://www.enactgame.eu`. As soon as the game is loaded, the user is asked for registration. The user is asked to submit the following information:

- nickname (mandatory);
- password (mandatory);
- age (mandatory);
- gender (mandatory);
- preferred language (mandatory);
- institution who redirected the participant to the website (optional).

Once registered, the user will be redirected to a selection screen where he or she will decide if to proceed to the Assessment environment, where a predefined and standardized path to a negotiation profile will be provided, to the Training environment, where

Figure 4.1: Screenshot of the selection screen presented to the user after registration in the Enact platform.

the user is presented with scenarios adaptive to their negotiation profile, or to the Free play mode, where the user can simply select and play one of the available scenarios in the preferred order, without receiving feedback nor a guided path. The selection screen is shown in Figure 4.1. If the user has not previously completed the Assessment, it is required to do so in order to proceed to the Training environment. All the information about the user, the games that are played, the choices made during the game sessions and the data about the scenarios are contained in a cloud database structured on Parse backend, an opensource server infrastructure (`http://parseplatform.org/`) developed by Facebook that can be hosted on a variety of cloud services like AWS, App42 and also on Parse web service itself, before its shutdown - that happened at the beginning of 2016. Scenarios consist of xml files and are uploaded to the Parse backend, were they can be accessed by the browser game. The scenarios are originally developed in English and then translated into Italian, Turkish and Spanish with the collaboration of professional translators hired within the project and include different situations and different roles for the user. The registration step, which is needed to

initialise the game according to the parameters of gender and nationality, is required unless the player wants to solely run a trial demo of one scenario, but the virtual character played will not reflect the user characteristics and features.

4.2.2 The design of the virtual agents

As mentioned, the game is composed by a sequence of scenarios presented to the participant in a different order according to the environment. Inside each scenario, after a brief introduction which deals about a specific daily life situation, there are two 3D characters. The 3D virtual agents, which are present throughout all scenes, one for the player and one for the artificial bot, are implemented as human-like animated characters inside the game scenarios. Their appearance is modelled to represent different genders, culture and ethnicity, and their combinations create a total of 24 different characters. Concerning their dialogues, they are displayed as text inside a speech bubble floating above their heads. The interface, animations and environment are all developed with the use of the 3D graphics software Blender and the Unity graphic engine framework.

The agents interact dynamically and are able to perform a range of basic expressions using verbal cues (vocal tone, shape of the speech bubble and structure of the sentence) and non-verbal indicators (facial expression, eye contact, body posture and gestures). The character that represents the user is shown to the player, so that he or she can identify with it and decide the gesture and facial expression that better suits the hypothetical reaction in that situation. The gender of the user character is the same as the gender specified by the user at the registration, mandatory for accessing the platform. The virtual agent-user interaction inside each scenario is divided into states, which include one turn of speech for each party. In every state, the user can choose one among a set number of possible sentences (four in the beta and in the first release of the platform, five in the second release), each of which correlated to a gesture and facial expression that shows the way the sentence will be told to the agent. Screenshots of the scenarios within the first game release and the second game release are

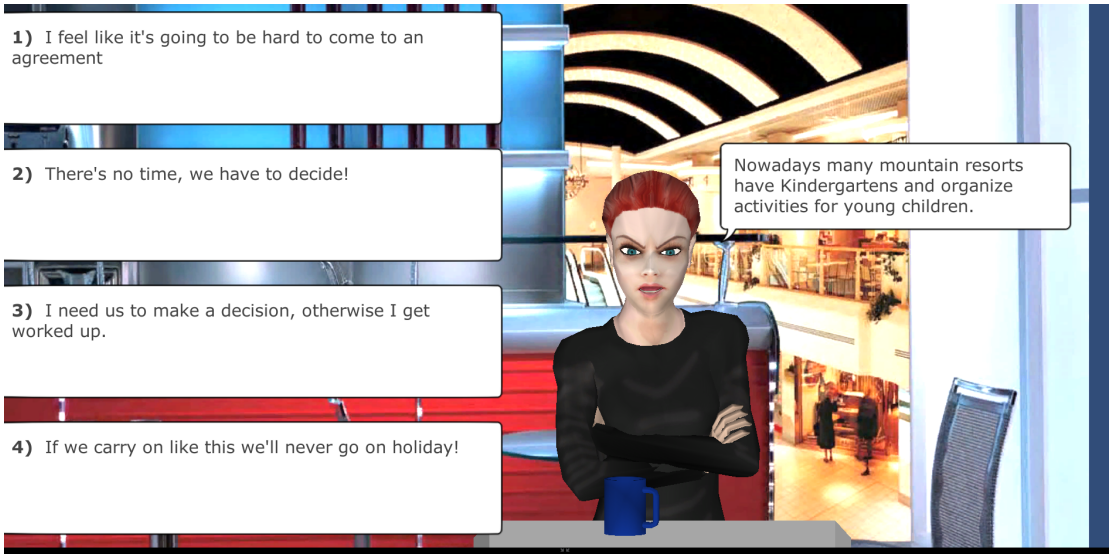

Figure 4.2: Screenshot of the scenario of the beta release of Enact. On the left, the user choices, on the right the virtual agent.

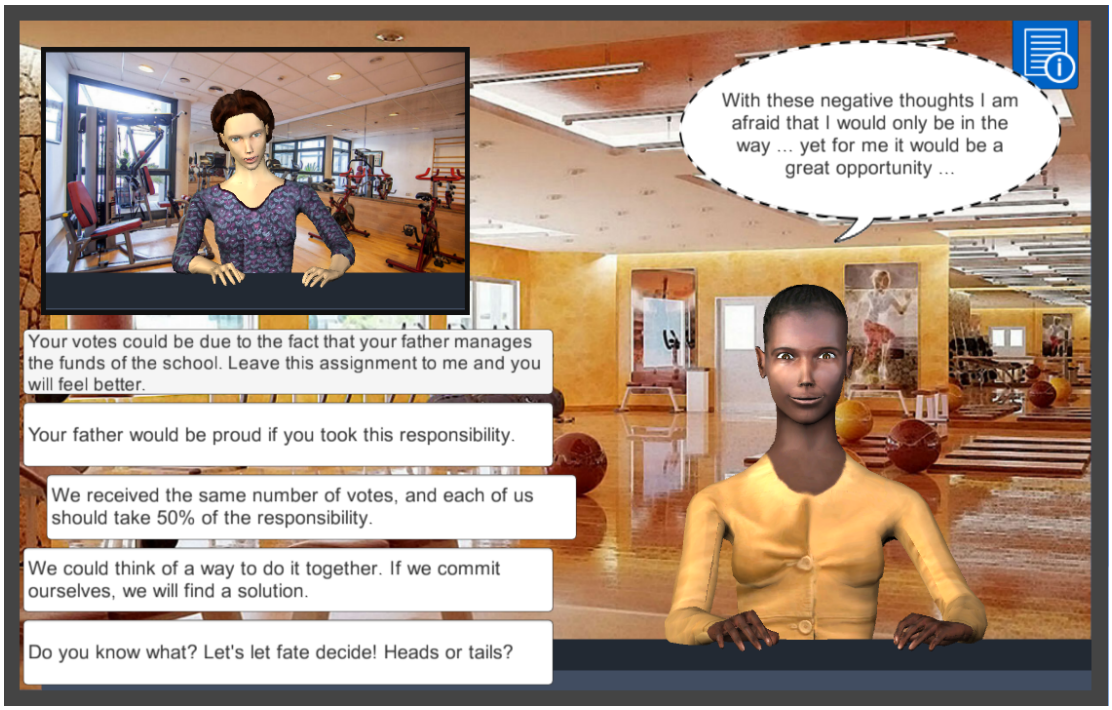

Figure 4.3: Screenshot of the scenario of the definitive release of Enact. On the top left, the user character, on the right the virtual agent.

	Integrating	**Dominating**
I statement	specific	prominent - You statement
fact-opinions	distinction - objective	no distinction
enquiring	enquiring others' opinions, preferences	provoking fear
criticism	constructive - feedback	destructive -judging-manipulative
vocal tone	steady and warm	rigid-cold-patronizing
vocal volume	to fit surrounding (not too quiet-not too loud)	harsh loud
facial expression	congruent	taut
eye contact	maintaining	intense
intent	to communicate	to dominate

Table 4.3: List of verbal and non-verbal indicators obtained by operationalizing Rahim and Bonoma's Integrating and Dominating negotiation styles.

shown respectively in Figure 4.2 and Figure 4.3. After the player's answer, the artificial agent computes its answer and waits for the next user answer. The behaviour of the artificial bot always reflects one of Rahim's styles, which is set at the beginning of each scenario and does not change - all scenarios are as long as a brief dialogue. The sentences of each scenarios for both the virtual agent and the user are constructed using the indicators suggested by operationalizing Rahim's styles, with the help of two negotiation experts that were recruited within the ENACT Project to check the consistency of the dialogues and of the behaviour of the bots. The list of the indicators for each of the styles can be found in Table 4.3 and Table 4.4.

Examples of utterances and locutions used by the virtual agents in the sentences of each of their styles is in the list below:

- Integrating
  - I would like to...; In my opinion...; What do you think about...;
- Dominating
  - You must...; Why didn't you...; We will do it my way...;
- Accomodating

	Accomodating	**Avoiding**
I statement	rambling-approval seeking	Evasive - Psychological withdraw
fact-opinions	putting down preferences	denial of reciprocal preferences
enquiring	dismiss preferences	Not bringing up preferences - topic shifting
criticism	rigid-self-criticism; self-devaluation	subtle-destructive criticism
vocal tone	soft quiet	rigid but quiet
vocal volume	tapering off	loud at the beginning
facial expression	uncomfortable	uncomfortable
eye contact	evasive	poor
intent	to please	to escape

Table 4.4: List of verbal and non-verbal indicators obtained by operationalizing Rahim and Bonoma's Accomodating and Avoiding negotiation styles.

  - I could be wrong but...; I'm sorry to bother...; I shouldn't have done...;

- Avoiding

  - We can discuss this later...; We had a difficult day...; Instead, let's talk about...;

The non-verbal expressions of the agent are coded with movement scripts modelled on real people acting; their efficacy and unambiguousness is assessed during the demo sessions using data collected through paper questionnaires administered at the end of game sessions and whose results will be presented in the next paragraphs.

4.2.3 The Assessment environment

The Assessment session is a game mode that allows the user to be assessed in a standardised way according to Rahim's model. In this environment, the system collects data about the user's behaviour and choices and creates a model of the player that will then be used for generating tailored information in the training session. The score and the profile of the player's negotiation skills are actually calculated by summing the independent concern for self and concern for other values of the variables accumulated during the interactions, which are represented within every sentence that the user can choose. The assessment environment is composed by a specific series of different

Title	Agent	Conflict	Gender
Who is taking the motorbike?	Integrating	Convergence	Different
Pizza or chinese?	Integrating	Divergence	Same
The disputed CD	Avoiding	Convergence	Same
A weekend in London	Avoiding	Divergence	Different
The contended parking space	Dominating	Convergence	Same
Question Time or Masterchef?	Dominating	Divergence	Different
Logo Design	Obliging	Convergence	Same
Which present to buy?	Obliging	Divergence	Different

Table 4.5: List of scenarios and variables in Enact's first release

scenarios that concerns a negotiation situation between two peers. The virtual agent's behaviour is static and not adaptive in order to provide a standardised comparisons between users and shows a specific negotiation style for each of the scenarios. The user does not get any feedback about his behaviour in-between the scenarios, and the user is told that there is no right or wrong answer. There are three parameters according to which the scenarios were designed:

1. The negotiation style adopted by the bot-controlled virtual character, respectively Integrating, Dominating, Obliging or Avoiding - Compromising was added only in the second release;
2. If the player and the agent have the same or opposite gender, so the interactions can be male-male (or female-female) and male-female (or female-male);
3. If the negotiation concerns a decision about two different possibilities (divergence) or concerns a single object which must be exclusively assigned to one of the two characters (convergence).

The eight scenarios of the first release of the platform and their topic is listed in Table 4.5.

It is worth noting here that the Compromising style is not present among the negotiation styles shown by the virtual agents in the beta and the first release. This has been a careful methodological choice. In all releases, anyway, the assessment of the user results in 5 style, where, in the first release, the compromising corresponds to a moderate

interaction of the two variables, concern for self and concern for other. Therefore, while in the beta and first version of the platform the user interacts with 4 distinctive styles and the user's style results by summing the independent variables and by identifying the position of the users profile within the graph depicted in Chapter 2 (Figure 2.1), in the second release the Compromising style has been introduced as independently evaluated from the others. The features that characterize the Assessment environment are:

- Each scenario is played only once;
- The scenarios are relatively short (4-6 exchanges) and equal for all users;
- The story and the context change from one scenario to another;
- There is not a specific goal for the user;
- The interlocutor (the virtual agent) has a specific negotiation style and does not change accordingly to user choices;
- The negotiation result is not showed at the end of each scenario;
- There is no feedback at the end of a single scenario;
- The assessment feedback is given only at the end of the entire session.

4.2.4 The Training Environment and Intelligent Tutoring System

Enact's ITS tries to take advantage from the literature results and proven structures that were described in Chapter 3, such as the structure of the ITS along with its modules Polson & Richardson (2013), rules-based models, curriculum sequencing (Brusilovsky 2012) and neural networks (Xavier et al. 2015). In this sense, the Enact platform fits perfectly in the culture of one on one tutoring for the user assessment of soft skill, where, as mentioned, the aim is not to teach well-defined subjects, but ill-defined ones, where there is no completely inclusive definition or established "correct response", but a whole pattern of user characteristics and behaviour needs to be taken into account

to correctly evaluate the user performance, context-specific and difficult - if not impossible - to generalise in many cases. In a domain such as conflict management and negotiation, many profiles can cross their boundaries and overlap; for this reason it is extremely important to consider as many user variables as possible, in order to be able to correctly model the learner and match any possible user profile which may occur. In the Enact Training environment the user modelling and feedback generation that the final user experiences is dynamic and context-specific, with an underlying adaptive tree-structured model. The core of the Enact training is based on curriculum sequencing, which conveys to the user tailored information personalised on their needs. The user path is built around the learner throughout the training scenario and each feedback is constructed dynamically including a variety of details which specify, for example, how the performance of the user matched the expected appropriate style in the specific situation and the user tendency in the dimensional map of the negotiation styles. The User Model in Enact is able to build a customised profile which takes into account all the information gathered until that point and the kind of choices made by the user, following four dimensions:

1. the appropriateness of the style to the situation;
2. the consistency of the style throughout the interaction;
3. the consistency of the style to reach the final aim of the scenario;
4. the number of choices belonging to each style in the scenario.

4.4 shows the whole sequence path and the variables which determine the final feedback.

Concerning the Expert and the Instruction Module, two are the possible types of information which can be included into the feedback structures and changed according to the path followed by the user: the first concerns the description of each negotiation style; the second specifies the type of situation each style is the most appropriate in. This knowledge is injected dynamically into the final feedback sentences provided

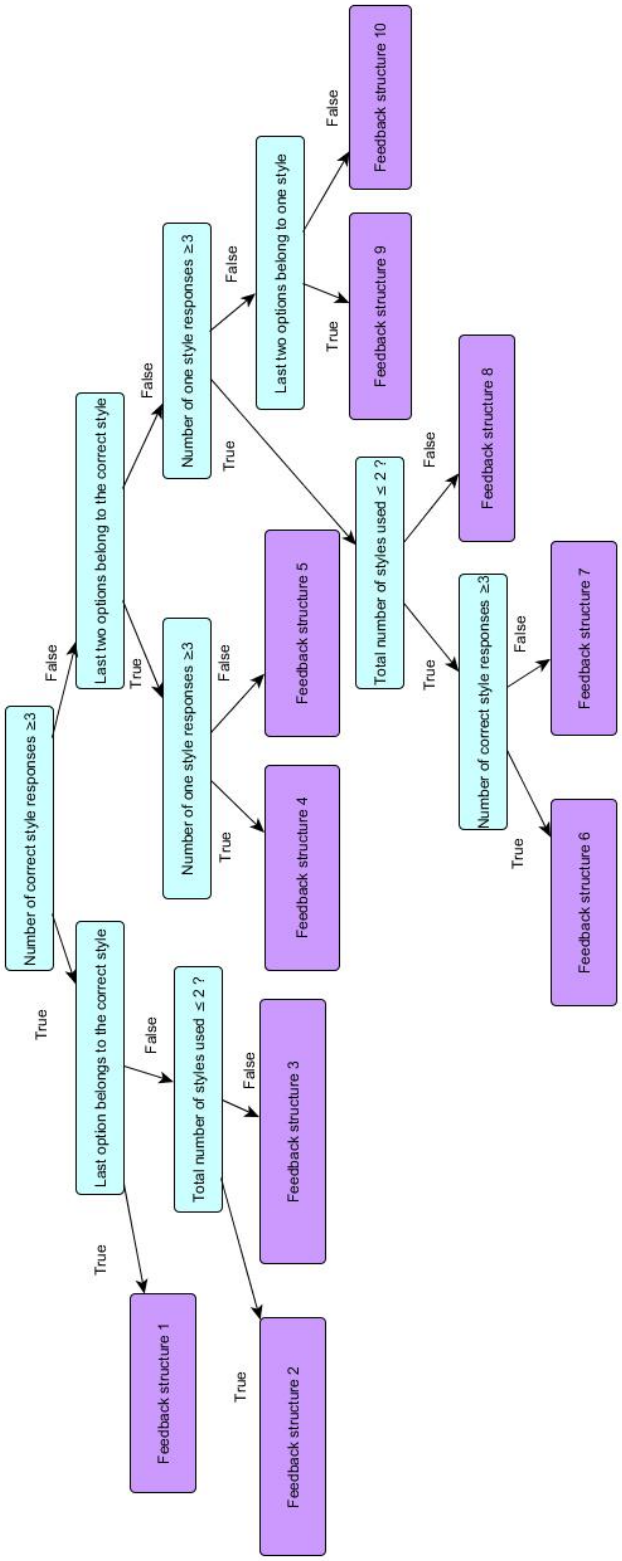

Figure 4.4: Possible assessment paths computed by the Enact tutoring system according to the order, number and appropriateness of the style chosen during the interaction with the agents in the scenarios.

at the end, creating a rich and complete feedback to help the user understand the complexity and situatedness of Rahim's negotiation styles. This tree-based curriculum sequencing structure, along with a character-based scenario and dynamic user profiling aims to integrate the evaluation of the learner provided by the Assessment environment, guaranteeing a tailored profile, detailed information and a motivating user framework to experience real-life negotiations in a safe and stimulating environment. The specific features that characterize the Training environment are listed below:

- Each scenario can be played more than one time;
- The user is given a specific goal to achieve at the beginning of each scenario;
- Similar scenarios can have different contexts, variables and initializations;
- The negotiation result is showed at the end of each scenario;
- For each scenario there is a suggested and more adequate style that the user should adopt that would result in a successful negotiation;
- A detailed feedback is given at the end of each scenario, where the user can see his predominant negotiation style, the other styles adopted in the interaction, and can also have a look at the history of the choices made during the dialogue with the virtual agent.
- Tips and advices are given on how to improve the user's negotiation style, as an example playing again the same scenario or suggesting another one.

The Training environment has been under development during the course of the first release of the platform, which included only the Assessment environment. For this reason, more details and the full implementation of the Training environment, scenarios and variables will be described in the next chapter.

4.3 First release evaluation: Materials and Methods

Concerning the deploying and release of the game throughout different channels, the game has been made available for free online and a user manual has been created for the software. The manual has been written in English and translated in the languages of the three partners involved in the trials (Italian, Spanish, Turkish). It is important to note that the versions used for the different phases of testing had some special features. In particular:

- they required to fill additional questionnaires in order to compare the Enact methodology with existing tests (e.g. the ROCI II or other personality tests);
- they required to specify the ENACT project partner who was responsible for that participant;
- they didn't provide any feedback to the users, after the completion of the scenarios or the assessment procedure;
- the links to these versions were not public, nor linked in any other part of the website, to avoid any possible influence on the subjects before playing the scenarios.

The first step in the direction of the software evaluation and investigation of the effects and reactions of the users to it has been done after the development of the beta version of the Enact platform. At the Sci-Tech Showcase 2014, held at Plymouth University in the days 16-17 September 2014, the ENACT Game first demo was presented to more than a hundred people and feedback was collected in the form of short questionnaires administered right after the game session. The demo was represented by a single scenario, and it was not yet possible to view the user's virtual agent in the interface. In the scenario of the demo, people were requested to deal with the artificial intelligence in order to find a compromise on the holiday location. In particular, the user dressed the part of Paul, the husband of Carla, the artificial agent, who prefers the mountains over the seaside, and tries to convince the user to renounce to a holiday on the seaside. The virtual agent keeps a Dominating style toward all the dialogue. At the end

of the interaction, the user receives a profile on their communication and negotiation skills based on their amount of answers given for each of the styles and a short tip about which aspects to take into account in order to improve their ability. There were 4 possible profiles based on Rahim and Bonoma's model of negotiation and conflict handling (Accommodating, Avoiding, Dominating and Integrating) and 3 possible short non-standardised tips, even though this feedback was provided more as a debriefing than as a real user assessment.

4.3.1 Experimental Procedure

During the Showcase, there were two PCs showing a "Play" button on their screen. If interest was shown toward them, the person was asked if he or she was willing to try the demo version of a new videogame to assess negotiation skills. He or she was warned that the final assessment would have not been accurate, but that his or her honest feedback on the platform and the experience with it was welcome. Then, the participant was invited to sit alone in front of one of the two PCs and to read the instructions on how to interact with the interface. People who tested the game demo were in an age range between 6 and 60, but the questionnaires were administered after the game only to people who reached at least the age of 11 as it is considered an appropriate age for performing cognitive tests and also reading abilities, comprehension and literacy are satisfactory in the UK (Deary et al. 2004, 2007). Only participants who spent at least 5 minutes interacting with the game were asked to complete the questionnaire. It took an average of 10 minutes to both interact with the game and fill the questionnaire.

The questionnaires collected were completely anonymous and the only demographic information required were age and gender. They contained 8 questions which concerned various aspects of the game and ranged from the user interface to the usefulness of the game content. The first 7 items had their possible answers based on a 5 points Likert scale, where 1 stood for "Extremely bad" and 5 for "Extremely Good"; the last question's possible answers, instead, were "Yes" or "No". A final empty box for comments was left as an option at the end. 152 subjects tested the game, of which 72

Questions	Average points / 5
How appropriate and natural were the emotions expressed by the artificial bot?	3.7
How much realistic did you find the conversation with our artificial bot?	3.4
How would you rate the user interface (the controls, buttons, settings, etc.) of the ENACT Game?	3.75
How interesting were the information and the profile given at the end of your game session?	3.77
How would you rate the graphics and video quality of the ENACT Game?	3.29
How would you rate the usefulness of such game in dealing with real life situations?	3.70
Can you rate your overall experience with the ENACT Game?	3.91
Would you play this game again with different scenarios and characters?	95.3 % Yes, 4.7% No

Table 4.6: Means of the participant score obtained by each question in the first demo evaluation

in the age 6-10. A total of 79 valid subjects between the age of 11 and 60 completed the questionnaire, with a mean age $\approx$ 20.6, of which 41 males and 38 females.

4.4 First release evaluation: Results and Discussion

Even though it was not the aim of the investigation, it was observed that the most common profile obtained after the game sessions was the Avoiding, and in general the majority of the answers belonged to the Avoiding style. This is a natural consequence of a particularly evident and strong Dominating style, as that shown by the agent in the demo scenario.

4.4.1 Analysing group differences

Data showed that the overall feedback was positive, the average rating for each of the question never scored below 3.5 points except for Question number 2 (regarding the realism of the conversations) and 5 (about the game graphics) which scored a mean of 3.4 and 3.3. 95.3% of the subjects answered "Yes" to Question number 8, showing an overall interest in the future development of the game (as shown in Table 4.6). It is

important to highlight that the game graphics was still immature, and that the first version released already included a greatly improved version of the interface. Considering the gender differences, as shown in FIgure 4.5, data revealed a significant difference in the answer to question number 5, about the game graphics, where male participants generally rated this aspect much lower than females (measured with Mann-Whitney U Test, $p < 0.05$). The sample was also divided by age range, in particular between aged under 18 (age 11-17, 42 subjects) and over 18 (age 18-60, 37 subjects). Scores are shown in Figure 4.6, and the only significant difference in the response was on question number 3, about the game interface, where younger participants gave a much lower score than adults (Mann-Whitney U Test, $p < 0.05$). In the additional comment box one participant (Male, 18 years old) mentioned the importance of adding more sound and improving the graphic quality of the game. Another participant reported that the interactions were too long and therefore not realistic (Male, 33 years old). Both the scenarios and interface were modified and improved in the second release.

4.5 First release validation: Materials and Methods

4.5.1 Validity and Tools

Concurrent validity refers to how a test measures what it purports to measure. The objective of the validation process is to demonstrate a reliable link between the Enact tool and what is meant to be measured, that is the style of handling conflict of users. Since Enact assessment tool has been designed to measure conflict management style, the simpler and effective way, to verify whether it is able to measure such a construct or not, consists in calculating the correlation between Enact and Rahim's test scores. In order to prove the correlation between the Enact game platform and the questionnaire ROCI-II (Rahim Organizational Conflict Inventory-II), the procedure that was used is the administration of both to the participant in a randomised order. This procedure is also similar to the one followed to validate the second release, described in the next chapter. In the validation of the first release, the Enact platform scores were additionally measured in respect to their ability to evaluate and/or predict other constructs such

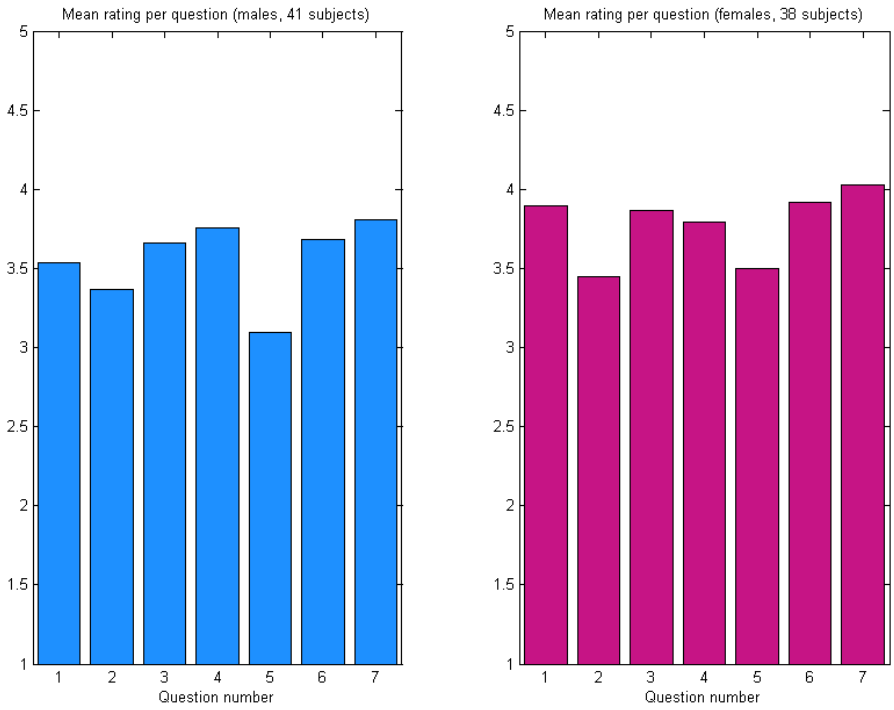

Figure 4.5: Score differences in the answers of the questionnaire between male and female.

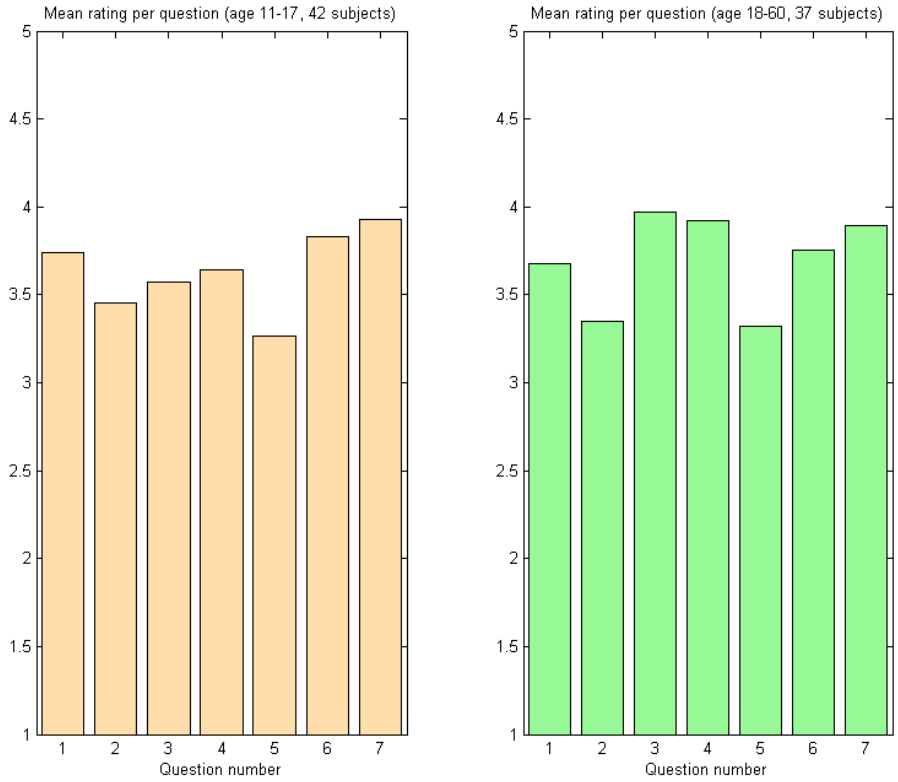

Figure 4.6: Score differences in the answers of the questionnaire between under 18 years-old and adults.

as personality factors, assertiveness, self-efficacy and Coping strategies. The other scales chosen for this purpose were in particular:

1. Adjectives of BIG Five personality inventory (20 items);
2. Assertive Self-efficacy Test (6 items);
3. Negotiation Self-efficacy Test (8 items);
4. Ways of Coping Test (26 items).

Although fundamental, in fact, concurrent validity is not the sole desirable psychometric property of a test. Divergent or discriminant validity is another important face of the validity coin. Discriminant validity is used to check whether a construct, in this case negotiation skill, that is supposed to be unrelated is actually unrelated. This extra step was therefore taken in order to ensure that negotiation can be confirmed as being a separate concept, independent from other stable personality traits and that the software was effectively aimed at assessing exclusively this soft skill.

The main tool for evaluating negotiation skills, the ROCI-II questionnaire, consists of 28 statements on a 5-point Likert scale measuring five independent dimensions of the styles of handling interpersonal conflict: 7 statements for Integrating (IN), 6 statements for Obliging (OB), 5 statements for Dominating (DO), 6 statements for Avoiding (AV), and 4 statements for Compromising (CO). The instrument contains Forms A, B, and C to measure how an organizational member handles conflict with supervisor, subordinates, and peers, respectively. The subscales have adequate reliability and validity. A higher score represents greater use of a conflict style. Sample items of the instrument for each of the styles are:

- (Integrating): I try to investigate an issue with my supervisor/subordinates/peers to find a solution acceptable to us.
- (Obliging): I generally try to satisfy the needs of my supervisor/subordinates/peers.

- (Dominating): I use my influence to get my ideas accepted.
- (Avoiding): I attempt to avoid being "put on the spot" and try to keep my conflict with my supervisor/subordinates/peers to myself.
- (Compromising): I try to find a middle course to resolve an impasse.

The ROCI II it is not an evaluative tests of intelligence or behavioural skills. It is brief and can be usually administered in a range between 5 and 10 minutes. It is self-administering (directions for completion are written on the questionnaire), and must be taken individually. The full form used in the research and its instruction can be found in Appendix C, while more psychometric information about the test can be found in Rahim & Magner (1995) and Weider-Hatfield (1988). Concerning the game, the Enact tool asks to the users to negotiate with an avatar in 8 different scenarios. On average this activity takes 5 minutes for each scenario. The total amount of time needed for each participant is about 65 minutes.

4.5.2 Experimental Procedure and Details

For each round of validation every partner of the ENACT project was expected to test the software reaching out to as many participants as possible following the common procedure and then report the results to the other partners. All psychological tests were available to the participant on online Google Forms, automatically and anonymously connected to their Enact game account and randomly administered before or after the gaming experience. Once a participant was involved, the step forward was to be sure that the whole battery was complete (Enact along with the 5 psychological tests). It was possible to participate to the validation both in person or remotely, since participants in person were recruited in the different partner countries and advertisements for the study were also sent out by email. It was not needed that the participant was alone in order to proceed with the test. In particular, the validation procedure provided two possibilities:

- testing people under the experimenter control. The advantage of this case is

obviously the ability to help the participant on the spot for any technical enquiry, clarification and control for any possible disturbance that may occur during the testing process. However, in order to make the process identical in both cases, the participant does not receive further oral instructions;

- testing people remotely by sending them an email with a guide and instructions for the installation of the Unity 3D Webplayer plugin. This procedure allows to control for the selection bias, and allows to reach out to many more participants, which are needed since validating a new tool for psychological assessment requires a broad sample.

Before to begin the process, the user is provided with a consent form (which is attached in Appendix D) and, after he or she signs it, an Information sheet (Appendix E). Then, a brief introduction to greet the participant is shown on the screen:

"Thank you for participating to the Enact Survey. We gently ask you to play with the Enact game, a game in which you are required to start a discussion and negotiate with an artificial player, and answer to a set of questions related to 5 psychological tests regarding conflict management, personality, assertive efficacy, self-efficacy and coping. The testing procedure is completely anonymous, however by leaving your email address you will receive a feedback when we will finish to process the data."

Then, the user is asked to register with an arbitrary ID and the information requested by the platform, and is either redirected to the tests webpage or sent to the game. The online tests have a brief instruction on top of each, while before the game, the user is welcomed with the following message on-screen:

"Welcome to the Enact Assessment tool, you will be asked to play a character (avatar) and negotiate in 8 different scenarios with an artificial player. Each scenario will be introduced by a brief description. Please, start a conversation by choosing the desired sentence among a set of four possible sentences displayed on the left of the screen. Your avatar will be shown on top left of the screen. Press Start to start playing."

Style	Cronbach's alpha
Dominating	0.862
Integrating	0.83
Obliging	0.636
Avoiding	**0.326**

Table 4.7: Psychometric reliability of Rahim styles in the first release. The styles which have not resulted reliable are in bold.

The only requirement for the participant was its ability to correctly understand and read in one of the languages provided in the Enact software (English, Italian, French, Turkish) and have normal or corrected-to-normal vision.

The total sample for the validation of the first release of the Enact game platform - assessment environment - consisted of 169 individuals, 113 males and 56 females, aged between 10 and 63 (mean = 27.34, st.d.= 8.93). Of the total, the participants recruited by the Plymouth University project partner were drawn from a pool of native English speakers throughout online and offline advertisement and were composed of a total of 41 subjects, of which 24 males and 17 females, in the age range 10-63 (mean = 29.73, st.d.= 10.25). This sample, which was collected in an attempt to maintain the balance age/gender, was added to other 125 subjects recruited by the other project partners. No significant statistical difference has been found between these distributions, therefore the full sample was analysed as a whole.

4.6 First release validation: Results and discussion

The reliability of the Enact platform was measured through Cronbach's alpha - also called tau-equivalent reliability - whose theoretical value varies from 0, no reliability at all, to 1, maximum reliability. Table 4.7 presents values of alpha for the four conflict management styles considered. Values considered unacceptable in regards to their internal consistency have been marked with asterisk. The Dominating and Integrating values present a very good reliability; Obliging measure is acceptable whereas the Avoiding style results completely unreliable. Tables 4.8, 4.9, 4.10, 4.11 present detailed analyses for each scale within the 8 scenarios.

	Scale average score	Scale variance	Corrected scenario total correlation	Alpha
Dom_S1	7.43	49.654	0.618	0.846
Dom_S2	6.90	48.253	0.584	0.848
Dom_S3	7.02	47.124	0.628	0.843
Dom_S4	7.29	48.402	0.576	0.849
Dom_S5	7.22	47.772	0.616	0.845
Dom_S6	6.83	46.212	0.556	0.853
Dom_S7	6.61	44.184	0.648	0.841
Dom_S8	7.03	45.323	0.683	0.837

Table 4.8: Reliability of the Dominating style for each of the 8 scenarios.

	Scale average score	Scale variance	Corrected scenario total correlation	Alpha
Int_S1	22.28	62.810	0.564	0.809
Int_S2	22.05	64.846	0.501	0.817
Int_S3	21.83	66.226	0.495	0.818
Int_S4	21.59	61.651	0.550	0.810
Int_S5	20.99	59.179	0.577	0.807
Int_S6	21.68	63.597	0.523	0.814
Int_S7	21.88	59.579	0.573	0.807
Int_S8	21.30	57.178	0.656	0.795

Table 4.9: Reliability of the Integrating style for each of the 8 scenarios.

	Scale average score	Scale variance	Corrected scenario total correlation	Alpha
Obl_S1	6.97	16.579	0.389	0.588
Obl_S2	7.51	19.383	0.246	0.626
Obl_S3	7.49	17.341	0.473	0.565
Obl_S4	7.59	19.347	0.309	0.610
Obl_S5	7.99	19.489	0.264	0.621
Obl_S6	7.60	17.427	0.389	0.587
Obl_S7	8.11*	20.733*	0.154*	0.645*
Obl_S8	7.77	17.652	0.409	0.582

Table 4.10: Reliability of the Obliging style for each of the 8 scenarios. Scenarios with asterisk are those that, if removed, improve the reliability of the measure.

	Scale average score	Scale variance	Corrected scenario total correlation	Alpha
Avo_S1	5.32*	7.859*	-0.026*	0.379*
Avo_S2	5.54*	7.450*	0.070*	0.328*
Avo_S3	5.66*	7.703*	0.044*	0.338*
Avo_S4	5.53	6.348	0.261	0.216
Avo_S5	5.80	6.671	0.270	0.222
Avo_S6	5.88*	7.138*	0.132*	0.296*
Avo_S7	5.40*	6.696*	0.155*	0.282*
Avo_S8	5.90*	7.182*	0.148*	0.288*

Table 4.11: Reliability of the Avoiding style for each of the 8 scenarios. Scenarios with asterisk are those that, if removed, improve the reliability of the measure.

Scenarios with asterisk are those that, if removed, improve the reliability of the measure. The most unreliable measures correspond to the Obliging an Avoiding scales, as shown in Tables 4.10 and 4.11, being the latter the worst. The avoiding scale, in fact, lacks consistency across items (intended as single interactions). Regarding the correlations between the conflict styles and the other psychological tests involved in the study, results show that Rahim's test (ROCI-II) and Enact present a high statistically significant correlations on the dominant and integrating scales, as shown in table 4.12. Lack of correlations between personality traits and Enact is an evidence of divergent validity, since the assessment tool was not meant to measure such stable traits (Table 4.13). Considering Assertive Self-Efficacy and Negotiation Self-Efficacy, a further proof of the itnernal validity of Enact is shown by the correlation between the negotiation efficacy scale and the Integrating score in Enact, as in Table 4.14. Correlations between Enact and the 6 coping strategies provide further evidences of convergent validity, as evident in 4.15. In fact the Coping test alike the ROCI is meant to address a particular conflicting event. In particular significant positive correlation between confrontive scale (that relates with hostility and risk taking) and Enact dominating scale was found. Interestingly, enact dominating scale correlates positively with escape-avoiding coping strategy, suggesting that mediating with an opponent can be a stressful activity. Significant opposite correlations between confrontive, escaping-avoiding coping strategies and enact integrating style support this last point. Moreover, the significant positive cor-

		Enact-Dom	Enact-Int	Enact-Obl	Enact-Avo
ROCI_Dom	Correlation	,261**	-0.14	-0.173*	0.016
	Sig. (2-tails)	0.001	0.092	0.037	0.851
	N	145	145	145	145
ROCI_Int	Correlation	-0.206	0.320**	-0.125	-0.222**
	Sig. (2-tails)	0.013	0.000	0.134	0.007
	N	145	145	145	145
ROCI_Obl	Correlation	-0.035	0.041	0.013	-0.054
	Sig. (2-tails)	0.678	0.627	0.875	0.522
	N	145	145	145	145
ROCI_Avo	Correlation	-0.053	-0.004	0.118	-0.042
	Sig. (2-tails)	0.528	0.965	0.158	0.618
	N	145	145	145	145

Table 4.12: Pearson's correlation between Enact and ROCI-II. Significant values with p<0.05 are marked with one asterisk, highly significant values with p<0.01 are marked with double asterisk.

relations between seeking social supports, accepting responsibility coping strategies and Enact integrating conflict management style show another point in favour of the convergent validity of the Enact assessment tool. Accepting responsibility and seeking social support, in fact at face value play an important role in integrating conflicting parties' interests.

4.6.1 Additional variables analysed

The administration order of the tests and the Enact game was randomly set for each test taker. In order to evaluate this effect, the data were analysed with the use of ANOVA. Results indicate no significant difference between the group that received the test battery before playing Enact and the other group ($p > 0.05$), therefore this variable was excluded in the validation of the second release of the platform.

The effect of the gender of the virtual agent compared to that of the user (same sex or different sex) was analysed. Figure 4.7 shows the frequencies of Dominating, Integrating, Obliging and Avoiding responses in both cases. The differences are not statistically significant. The effect of the style displayed by the virtual agent, instead, shows significant differences among the groups according to the scenarios (Measured with Chi square, p=0.006702). In particular, as shown in Figure 4.8, when the virtual agent was

		Enact-Dom	Enact-Int	Enact-Obl	Enact-Avo
Extraversion	Correlation	0.111	-0.044	-0.139	0.066
	Sig. (2-tails)	0.182	0.595	0.094	0.415
	N	146	146	146	146
Agreeableness	Correlation	0.143	-0.128	0.015	-0.017
	Sig. (2-tails)	0.084	0.124	0.061	0.839
	N	146	146	146	146
Conscientiousness	Correlation	0.112	0.032	-0.189*	-0.086
	Sig. (2-tails)	0.179	0.697	0.022	0.301
	N	146	146	146	146
Neuroticism	Correlation	0.085	-0.074	-0.043	-0.070
	Sig. (2-tails)	0.310	0.374	0.607	0.400
	N	146	146	146	146
Openness	Correlation	0.034	0.014	-0.89	0.014
	Sig. (2-tails)	0.685	0.867	0.283	0.868
	N	146	146	146	146

Table 4.13: Pearson's correlation between Enact and BIG Five factors. Significant values ($p<0.05$) are marked with asterisk.

		Enact-Dom	Enact-Int	Enact-Obl	Enact-Avo
Assertive Self-Efficacy	Correlation	0.111	0.011	-0.124	-0.126
	Sig. (2-tails)	0.182	0.893	0.136	0.128
	N	146	146	146	146
Negotiation Self-Efficacy	Correlation	-0.153	0.246**	-0.137	-0.117
	Sig. (2-tails)	0.065	0.003	0.099	0.160
	N	146	146	146	146

Table 4.14: Pearson's correlation between Enact and Assertive/Negotiation Self-Efficacy. Significant values with $p<0.05$ are marked with one asterisk, highly significant values with $p<0.01$ are marked with double asterisk.

		Enact-Dom	Enact-Int	Enact-Obl	Enact-Avo
Confrontive	Pearson Correlation	0.3*	-0.2*	-0.1	-0.01
	Sig. (2-tails)	0.0002	0.0137	0.2129	0.9512
	N	146	146	146	146
Distancing	Pearson Correlation	0.08	-0.07	0	0
	Sig. (2-tails)	0.3243	0.3912	0.9836	0.9626
	N	146	146	146	146
Self-Controlling	Pearson Correlation	0.09	-0.13	0.07	0.04
	Sig. (2-tails)	0.2885	0.1301	0.4225	0.6497
	N	146	146	146	146
Seeking social support	Pearson Correlation	-0.07	0.16*	-0.1	-0,14
	Sig. (2-tails)	0.3835	0.0473	0.2291	0.0852
	N	146	146	146	146
Accepting Responsibility	Pearson Correlation	-0.11	0.16*	-0.09	-0.06
	Sig. (2-tails)	0.1869	0.0472	0.2639	0.5044
	N	146	146	146	146
Escape-avoidance	Pearson Correlation	0.23	-0.27*	0.12	0.02
	Sig. (2-tails)	0.0048	0.0009	0.1480	0.8523
	N	146	146	146	146

Table 4.15: Pearson's correlation between Enact and Coping strategies. Significant values ($p<0.05$) are marked with asterisk.

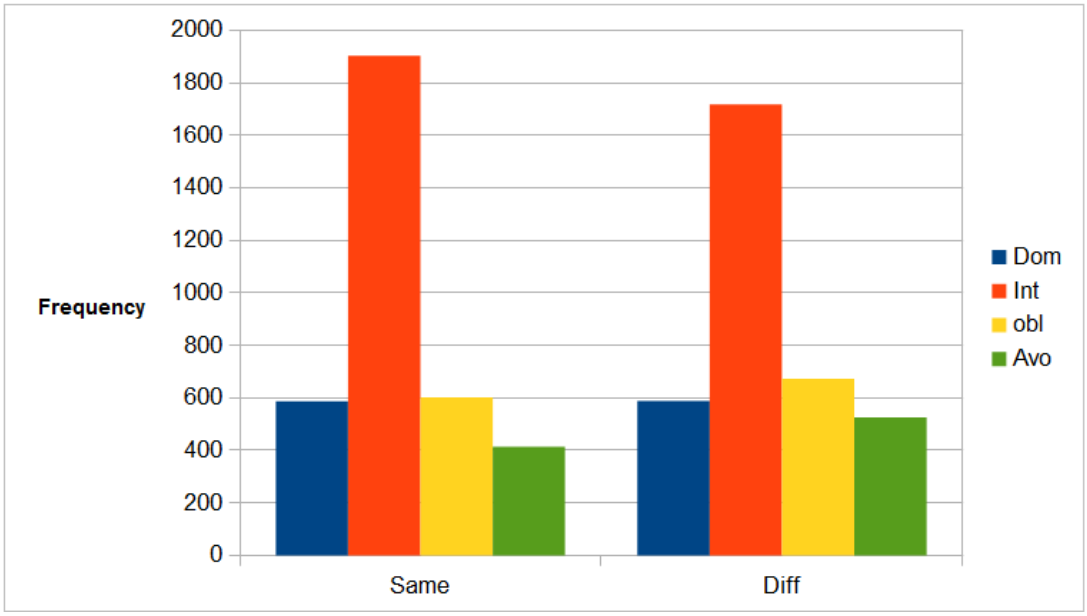

Figure 4.7: Graph showing the frequency of the answers for each negotiation style when the gender of the bot and of the user were equal (Same) or opposite (Diff).

Dominating there were significantly less Obliging and Avoiding responses and significantly more Dominating responses, and it is possible to conclude that a Dominating attitude elicits more concern for self. With an Avoiding agent, instead, more Obliging and Avoiding responses were elicited and less Integrating stragies, in fact the Integrating style requires to take into account the other's interests, which are not clear when someone is adopting an Avoiding strategy.

Regarding the Compromising style, the initial hypothesis that was tested in this first study was that this particular style could emerge as an interaction of the other four styles, as shown in Chapter 2. However, no significant correlation was found, therefore it was not possible in any way to map the answers of the subjects who scored as prevailingly Compromising in Rahim's scale with their behaviour and answers recorded in the Enact platform. This has brought to the acceptance of a new conceptualization of this style as independent from the other four, and this operationalization is also in accordance with how it is measured inside the ROCI-II. In fact, in this scale there are distinct items that aim at assessing the level of Compromising style. Other than that,

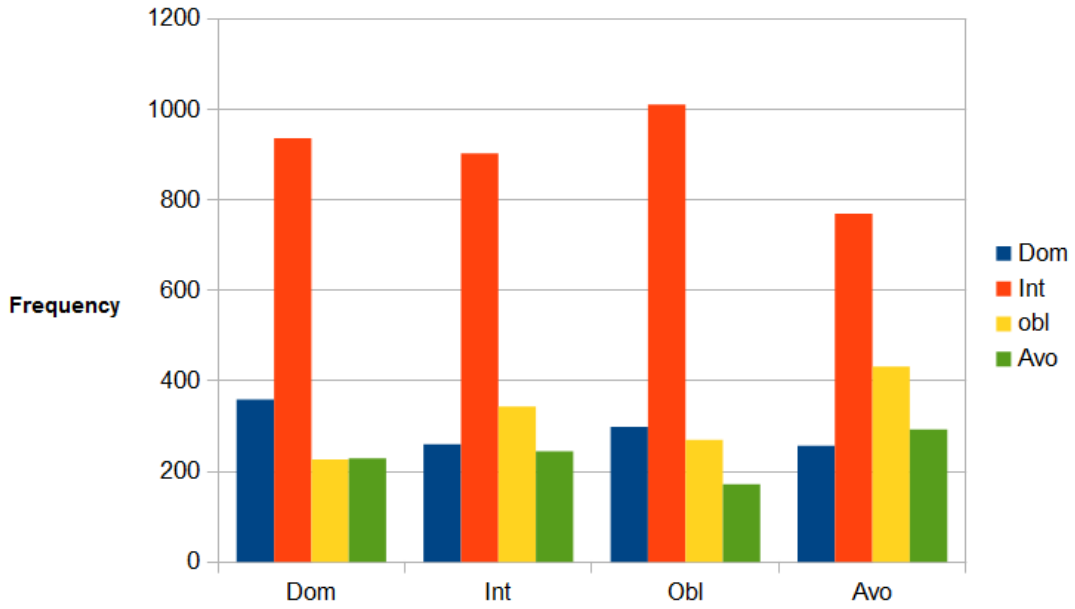

Figure 4.8: Graph showing the frequency of the answers for each negotiation style according to the virtual agent's style.

one implication of the previous conceptualization was that the Compromising was a lack of style, rather than a specific style itself, which led to the impossibility of measuring its prevalence when compared directly with the other styles. The Compromising style, instead, theoretically represents a completely different proposed solution than that of the Integrating and of the other styles, even if this solution often comprises characteristics which are present in more than one style.

4.7 Conclusion

This chapter described the design, implementation, first evaluation and validation phase of the Enact serious game. In order to design the components and features of Enact, a Training Need Analysis was conducted on the target groups in order to investigate the concept of negotiation, the attitude towards ICT, the need for new e-learning technologies and the technical knowledge and understanding of children and adults in the countries of the ENACT Project Partners. In particular, negotiation resulted as commonly believed to be associated with effective communication and social relationships. This structured concept led to the adoption of Rahim and Bonoma's model of negotia-

tion skills as conflict management rather than economic bargain. The Enact software, in its first release, is composed by the Interface, the Assessment environment and the Training environment, which reflect the need for training and evaluation of soft skills that emerged during the conducted interviews. After the first release of the platform, what has been evaluated and validated is the Assessment environment, composed by 8 scenarios - 4 for each negotiation style considered, respectively Integrating, Dominating, Avoiding, Obliging and concerning divergent or convergent conflict - depicting real life situations such as decisions about sharing a motorbike or a CD, and the interface. The evaluation consisted in the collection of assessment questionnaires to investigate the pleasantness of the experience with the platform, the usefulness of the information provided, the naturalness of the interaction with the agents and the interface and the likeliness of playing again with the game. The questionnaires indicated that the experience was generally indicated as pleasant, the information useful and the interaction fluid, where more than 90% of the sample expressed the will to play the game again.

After this evaluation, the platform underwent a validation process aimed at comparing how well it could measure negotiation skills - as modelled by Rahim - in comparison with the ROCI-II test, and if this construct also correlated with other scales measuring possibly associated factors like personality traits, self-efficacy, assertiveness and coping strategies. The sample was administered the Enact platform before or after the psychological tests and the results were analysed and compared. Enact scores correlated on the scales of Integrating and Dominating, but it was not possible to derive the Compromising style from the other scores. No specific correlation was found between Enact's items and the other psychological tests, confirming negotiation as a separate construct. The administration of the tests before or after the game did not affect the results while the virtual agents' negotiation styles in each of the scenarios significantly affected the strategies chosen by the participants. Following the results, feedback and comments obtained on the first release of the Enact platform, many changes were applied to this version in order to improve the user experience and its reliability. As will be described in the next chapter, the scenarios were renewed and shortened, the

possibility to choose and interact with the Compromising style was added, the Training environment was improved and detailed information were included at the introduction; in regard to the validation procedure, only the ROCI-II test (Form C) was administered and, since no effect of this temporal variable was found, always before to play the Enact game.

Chapter 5

Enact Game: second validation and a comparison of perceived pleasantness with traditional tests

For the second release of the platform, several changes were made to both reflect the results obtained after the first validation and improve the general user experience with the software. Concerning the evaluation of the negotiation styles, after the first validation of the software it was not possible to derive the information required about the Compromising style in Enact so to compare it with the ROCI-II dimension, so, in the second release, the Compromising style was introduced as a separated style, added as a behavioural aspect of the bot in a new scenario, as possible negotiation profile and as a response option to the list of sentences shown to the user at each interaction. Regarding the user interaction with the software, both the number of scenarios and the number of interactions per scenario was reduced. Detailed game-like information about the scenarios were added at the beginning of the scenarios, that helped provide a more reliable values to the users about which negotiation style to adopt in each situation. The latest platform release was evaluated and validated again: results show that adding a fifth style did not negatively impact the potential of Enact to recognize the player's style, while gaining likeability and user-friendliness. As a final test, the platform was compared to the ROCI-II in terms of implicit and explicit pleasantness, respectively measured with time perception and a questionnaire. The sample explicitly indicated Enact as much more pleasant, and while psychology students were able to correctly estimate the amount of time spent on the platform, student or professionals from other faculties even presented a time estimation bias, evaluating Enact as shorter than the ROCI-II.

ENACT, after the evaluation of the first release, was subjected to many changes that involved several aspects in order to improve its accuracy and user-friendliness. These improvements aimed also at allowing a more precise comparison with the ROCI-II test. Changes to the scenario structure, the virtual agents, both the environments

and the interface were made. Other than that, the server as well as the database were moved by project partners from a locally hosted to a Parse hosted backend (www.parse.com) and finally again to a local hosted one with a Parse backend structure. The migration process allowed a simplification of the connection to the database and was accompanied by a total renovation of the website, whose code was also ported from PHP to Asp.NET, making it more secure.

The details of the changes of each component of the Enact platform will now be described, along with the results of the new evaluation and validation carried out on the second and final release of the game.

5.1 Enact Game second release: Gameplay and content improvement

5.1.1 Changes to the Assessment Environment

5.1.1.1 The interface

Concerning the interface, a fresh new look was given to the platform. It was made easier to navigate between a "Free play mode", where it is possible to just select a scenario and play it without any data being recorded and any feedback provided, and the other two environments, Training and Assessment. In the scenarios, a "info" button on the top right side was added to quickly show again the introductory scenes containing the scenario description and variables like time and importance. The graphics and consistency of the animations of the user character in the top left side was improved, as well as the responsiveness of the cursor when navigating in the scenario. The interface of the scenario introductions was made more similar to modern role-play games, depicting the two characters and visualizing the variables (time and importance) and situation information in a clear manner. Figure 5.1 shows the first introductory scene, where the two agents are presented, while figure 5.2 shows the second introductory scene, showing the aim and the variables. Also, sound effects and a background music was added to the game.

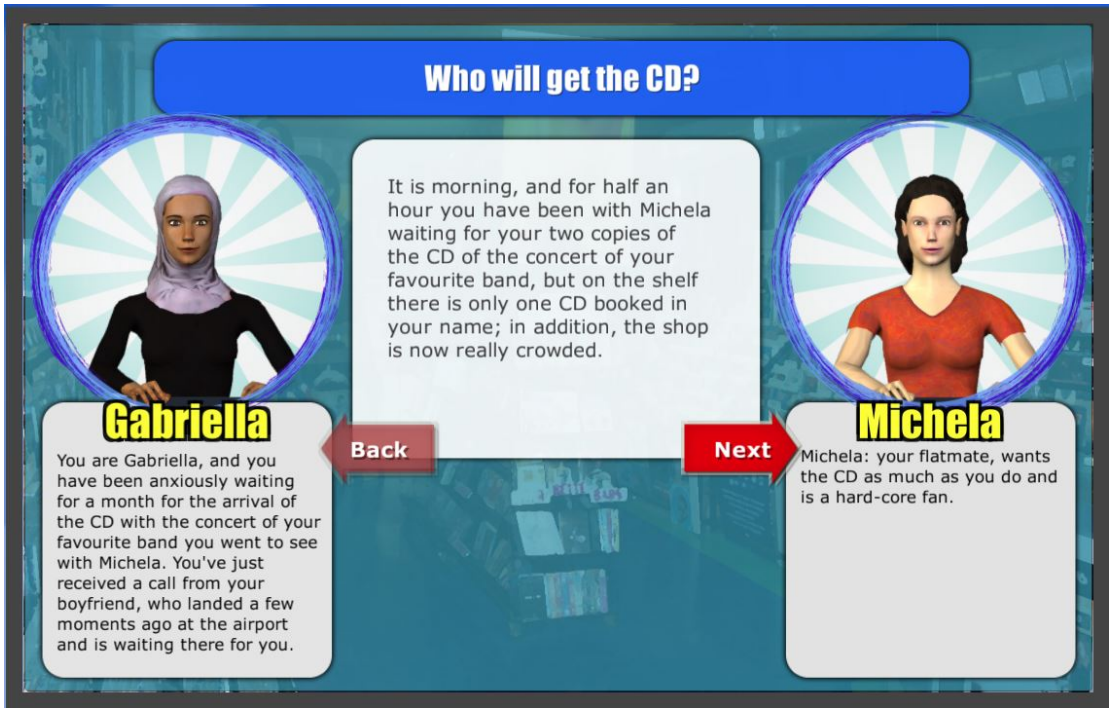

Figure 5.1: Introductory scene of the second release of Enact presenting the situation and the virtual agents.

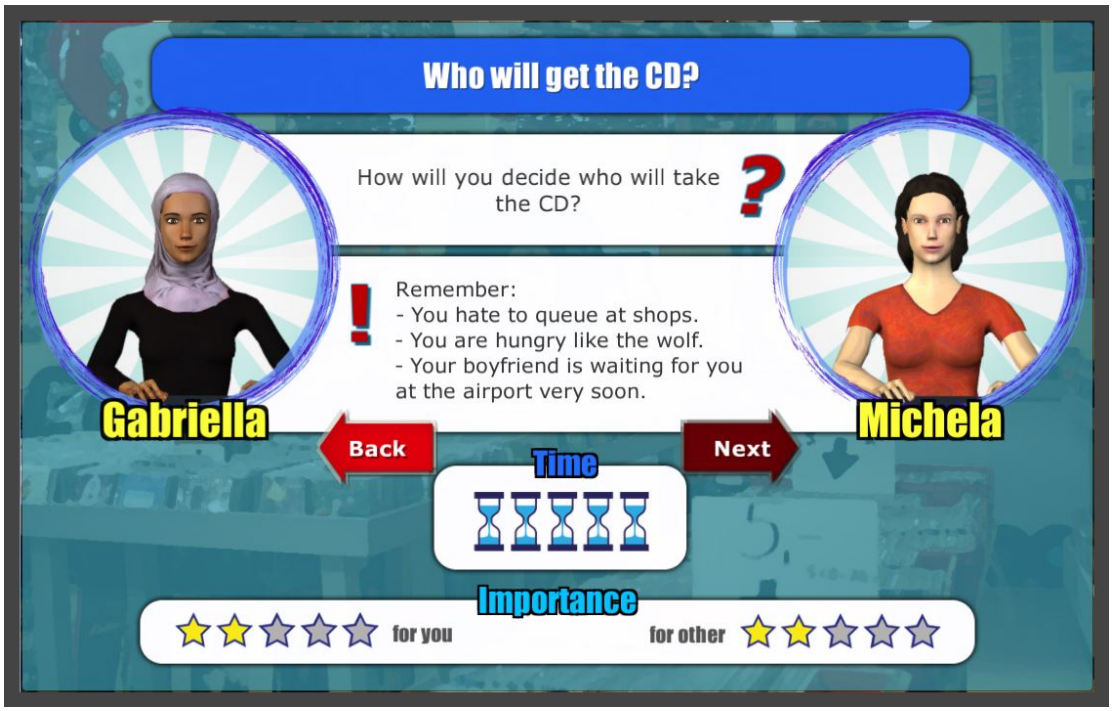

Figure 5.2: Introductory scene of the second release of Enact presenting the aim of the player and the variables of the scenario.

Figure 5.3: List of female characters in the second release of Enact.

5.1.1.2 The virtual agents

After the comments received in the first evaluation of the software indicating that the quality of the graphics was not modern enough and comparable to the standard of desktop and mobile apps and games, the virtual agents' gender, ethnicity and graphic quality was highly improved, as can be seen in figures 5.3 and 5.4. Gestures and facial expressions were made more realistic. Considering that the tone of the sentence can be cultural specific, related to the experience of the single individual and subject to misunderstanding, speech bubbles were visually designed to indicate the tone with which the sentence was uttered, as explained in 5.5. During the virtual agent speech turn, the appearance of the sentence in the speech bubble was accompanied by a gibberish sound, i.e. meaningless, unintelligible talk made of sounds, that was consistent with the tone.

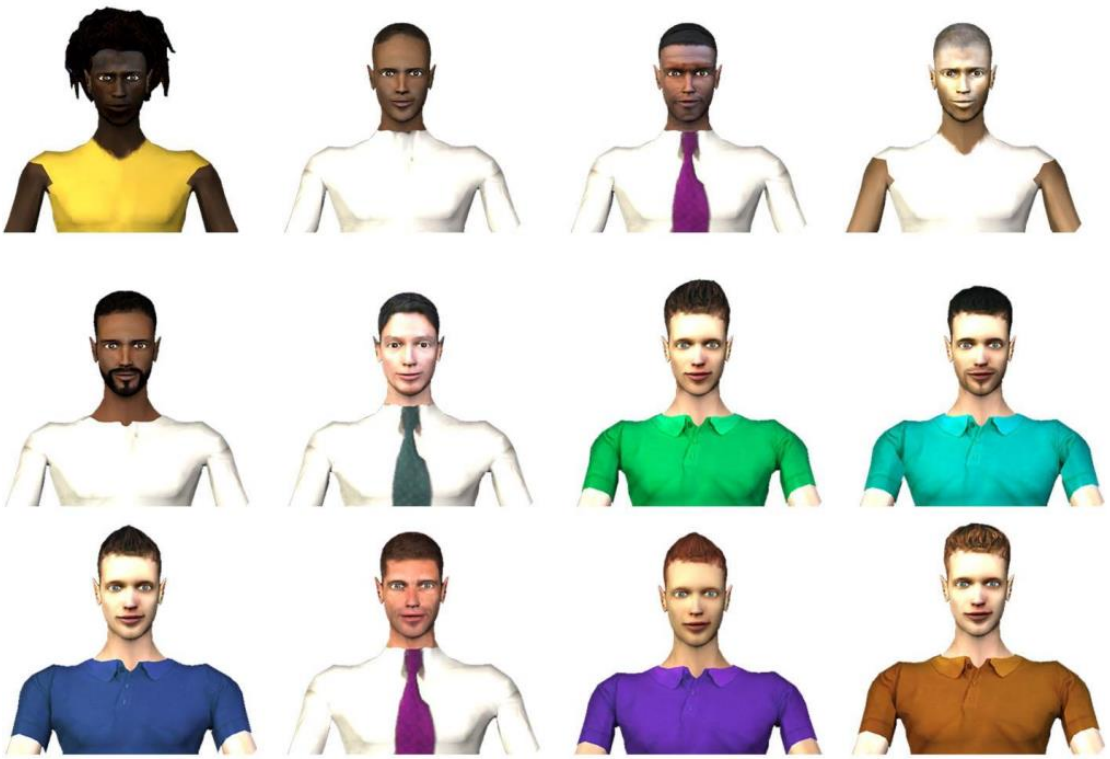

Figure 5.4: List of male characters in the second release of Enact.

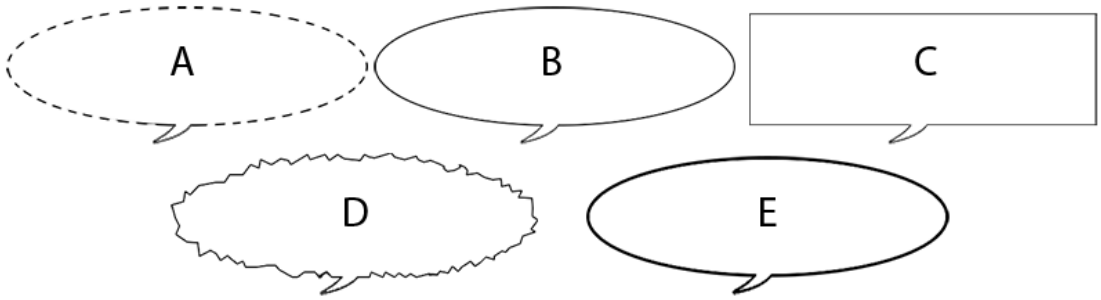

Figure 5.5: Shape of bubbles designed to help the user understand the tone of the sentence uttered by the virtual agents. A: passive tone; B: neutral tone; C: assertive tone; D: aggressive tone; E: very assertive tone.

5.1.1.3 Negotiation styles

Concerning Rahim and Bonoma's negotiation style model, since it was not possible to extrapolate a compromising dimension from the other four factors in the previous validation of the platform, the Compromising negotiation style was added as a separate style to the other four. A separate operationalization for the Compromising style was made, and the style was introduced as a possible profile in the outcome of the assessment, as possible fifth answer in the list of user choices in each scenario, as virtual agent behaviour in one new scenario, in the Training environment tutoring system, as separate gesture/tone/utterance configuration of the virtual agent. This major change drove most of the subsequent modifications to the scenarios.

5.1.1.4 The scenarios

The scenarios were subjected to consistent and massive changes driven mainly by the need to reduce the number of the interactions with the virtual agent, to smoothly introduce a fifth negotiation style and to make the situation and the player's aims clearer and more direct. Considering the introduction of a fifth choice, the time spent by the user reading and choosing the next sentence had an inevitable increase. For this reason the number of interactions in which the user was asked to make a choice was reduced to 5 per scenario. Other than that, to reduce inconsistencies with the agent's responses and to prevent the agent having the final word, the last interaction was not between the user and the virtual character, but asked the user to summarize his behaviour during the specific scenario and provide a final choice about the situation. This allowed the player to review the scenario from an outer perspective, be aware of other and self concerns and rethink of his or her choices in order to give a final decision about his or her attitude and adopted negotiation style.

The number of scenarios was reduced from 8 to 5, one per negotiation style, and this number includes a new scenario for the assessment of the interaction with a Compromising. Scenarios that less contributed to the evaluation of the user profile were removed. Since the interaction with an agent of each style is now unique and ex-

Title	Agent	Conflict	Gender
Who is going to take the motorbike?	Compromising	Convergence	Different
Pizza or Chinese?	Integrating	Divergence	Same
Logo design	Obliging	Convergence	Same
The disputed CD	Avoiding	Convergence	Same
Report or Masterchef?	Dominating	Divergence	Different

Table 5.1: List of scenarios and variables in Enact's second release

clusive, this change allowed to exclude the possibility of a test-retest variability and carryover effect. The list of the new and updated scenarios can be found in Table 5.1, while the full text of the interactions and choices within the scenarios can be found in Appendices D, E, F, G, H.

5.1.2 Changes to the Training Environment

The training environment was finalized, and a new user training system was developed to increase the width and variety of the scenarios and provide a more appropriate tutoring path. In the new training environment, in order to elicit a specific negotiation style that is independent from the style shown by the Virtual Agent, the user is presented the same scenarios played in the Assessment environment but with a distinctive difference in the introduction. In particular, the variables and the descriptions of the scenarios and of the characters inside them are modified in order to make them consistent with a situation that requires the adoption of one of the five styles to achieve a more successful outcomes, rather than the other four. These introductions were written following Rahim's guidelines and theory that explained in which situations each style is considered appropriate, as specified in Section 4, Chapter 2. A summary of the variables and the appropriateness of the styles can be found in Table 5.2.

5.2 Second release evaluation: Materials and Methods

The evaluation of the second release of the Enact platform was conducted in two different countries and events. The first part of the evaluation phase was conducted in the UK: questionnaires about the second demo interface and the full assessment system

Style	Appropriateness	Variables
Integrating	Issues are complex.	High importance for both parties
	Commitment is needed from other parties for successful implementation.	High engagement for both parties
	Time is available for problem solving.	Much time
Obliging	You believe that you may be wrong.	Low confidence
	Issue is more important to the other party.	High engagement for other, low engagement for self
	Preserving relationship is important.	High importance of relationship
Dominating	Issue is trivial.	Low importance for other
	Speedy decision is needed.	Short time
	Issue is important to you.	High engagement for self
Avoiding	Issue is trivial	Low importance
	Potential dysfunctional effect of confronting the other party outweighs benefits of resolution.	The other party is too engaged
	Cooling off period is needed.	Low engagement for self
Compromising	Goals of parties are mutually exclusive.	High importance for both parties
	Parties are equally powerful.	High engagement for both parties
	Temporary solution to complex problem is needed.	Short time

Table 5.2: Variables that influence the appropriateness of negotiation styles in each situation

were collected at the British Science Week hosted at Plymouth University in March 2015, where subjects were asked to play 4 different scenarios and to provide a feedback about the platform. The pool was composed of 39 people, in the age 11-28, of which 18 males and 21 females. The questionnaire was composed of 13 questions, of which 12 of them had a 5 points Likert scale as possible responses that ranged from "Totally Disagree" (represented as 1 point) to "Totally Agree" (represented as 5).

The second phase of the evaluation was conducted in Italy during the scientific fair "Futuro Remoto 2016", held in Naples, October 2016. The sample of participants was drawn from the people who attended the fair and were invited to play the game. When they completed the demo version of the Enact game, which was composed of one scenario randomly selected from the 5 available and a feedback about the negotiation style of the participant, they were asked to fill an 8 items questionnaire. The scenario lasted from 4 to 6 minutes on average, and the participant was asked to complete the questionnaire only if the time spent interacting with the game was at least 4 minutes. The questionnaire was composed of 7 items with scores on a Likert scale where 1 stood for "Extremely bad" and 5 for "Extremely good" and a final item which required a "Yes" or "No" answer. The items are the same as those used in the first evaluation, described in Chapter 4 Section 4, since the sample size was comparable. In fact, a total of 151 people participated to the data collection, of which 146 constituted filled the questionnaire with valid data and were included in the analysis. The age of the sample ranged between 10 and 64 (mean = 27.67, st.d. = 13.47), of which 78 females (mean = 27.71, st.d. = 13.84) and 68 males (mean = 27.62, st.d. = 13.13). The sample is considered homogeneous in regards to the variance and the two groups can be compared.

5.3 Second release evaluation: Results and discussion

Concerning the evaluation conducted in the UK, the sample was not as wide, however all the questions which were also asked in the previous event showed at least a slight improvement in their average score, and in this case only Question number 5, about

Questions	Average pts / 5
The conversation with the agents is realistic	3.59
The emotions expressed by the agents are appropriate and natural	3.9
The agents' movements and gestures are appropriate and natural	3.69
The emotions and gestures of the agents are useful for the comprehension of the speech	3.77
The graphics and video quality is modern and appropriate	3.41
The user interface (controls, buttons, settings) is intuitive and good-looking	3.87
The information and the profile given at the end are useful and clear	3.69
The scenarios deal about real life situations	4.1
The agents are behaving differently in each scenario	3.72
I found it easier to negotiate with some agents than others	3.97
I am motivated to negotiate even with the toughest agent	3.97
I find the overall experience with the ENACT game positive	4.33
Would you play this game again with different scenarios and characters?	94.7% Yes, 5.3% No

Table 5.3: Means of the participant score obtained by each question in the second demo evaluation - UK sample

how modern the graphics look, scored less than 3.5. Two questions even scored above 4, and, in particular, the highest score was reached by the 12th question, which was about if they felt the experience with the Enact game positive, while that same question had scored 3.9 on the previous demo. 94.7% of the subjects answered that they would play the game again with different scenarios. Table 5.3 shows a summary of the average scores obtained by the platform.

Concerning the Italian sample, the summary of the average scores obtained by each question is shown in Table 5.4. The highest frequency of answers is observed in all questions around the scores "Normal" and "Good", with the frequency of choice "Good" being higher than "Normal" in Q2, Q4, Q6 and Q7, which are all the questions which concern the content, while score "Normal" was chosen more frequently in case of questions that dealt with the graphical aspect of the videogame. Q6 is the only question which obtained a higher frequency of scores on point "Very good" than on point

"Normal", and this data shows that subjects, as players, were able to observe the usefulness of the experience with the Enact software and in particular that the scenarios were considered as near as possible to real life situations. The attention paid during the development to the design of the scenarios is evident and helps the users feel comfortable while playing their role inside the game, even though the participants demographics has a wide age range and includes children and adults. Q7, which was designed to allow the respondent to sum up their experience with the platform and provide a general rating and of the scores previously assigned to the other aspects of the videogame, is the question whose frequency reaches the highest concordance peak (53,4% on the score "Good") and therefore gives an insight about the interpretation of the scores of the other questions. While, in fact, score "Good" shows that the experience for the majority of the sample was perceived as pleasant, what negatively impacted on the final rating and whole consideration of the software was the interface and the graphics, aspects that at the moment are qualitatively weaker than the standards set by the videogame industry. This trend in the results is consistent with the data obtained in UK. It is possible to infer that what weighted on the final question and therefore on the whole experience was the content, since the mean score of question Q7 (mean(Q7) = 3.79) is nearer to that of the questions with regard the content (mean(Q2, Q4, Q6) = 3.69), than that of the questions which concern the graphics (mean(Q1, Q3, Q5) = 3.34). Q8, the only item which requested the participant to respond with a Yes or No obtained a 95.89% of "Yes", implying a high willingness of the participants to play the game even after receiving their negotiation profile during this data collection. This result is also in line with data collected in UK.

Since the age variance and mean of the sample was comparable across the two genders, additional analyses were carried out on the results. In particular, an analysis was conducted in order to investigate if there was an effect of the gender on the distribution of the answers 1-7. A significant effect of the gender has been measured with the Mann-Whitney U test in the case of questions concerning the interface (Q3, p = 0.006) and the graphics (Q5, p = 0.012), while for the other 5 questions there was no differ-

ence among the distributions in regards to males and females. Regarding the effect of the age on the answers, the correlation between the point on the scale assigned to each of the Likert items and the age has been conducted. Data showed that there is a strong negative correlation between the age and the score assigned to Q7 (measured with Pearson's r, corr. = -0.193, p = 0.02), item which concerned the overall experience with the Enact game. In order to further investigate the effect of the variable age on the points assigned to each question, the sample was divided into three age groups: age range 10-17, from 18 to 40 and from 41 to 64. Analysing the differences between the means of the scores assigned to the questions, as displayed by the graph in Figure 5.6, it is easy to notice that the question which has been answered in the most consistent way among all age groups is Q6, which concerns how useful this platform can be for transferring the information to real life situations; other than that, its general mean across all groups is the highest. Another visual information given from the figure is that except for Q5, concerning the graphics of the software, for all the other questions the age range 40+ reported the lowest means for all the questions. In general, motivations to play games has been showed to decrease with age (Yee 2006), but what appears to remain constant is the perceived usefulness of the platform as a valid training and assessment ground for everyday negotiations. A MANOVA has been conducted in order to investigate if there was a significant difference among the distributions of scores among these age groups. Significant differences were found in Q4 between the age groups 18-40 and 41+ (p = 0.037), concerning the usefulness of the information at the end, and Q7 between the age groups 18-40 and 41+ (p = 0.011), concerning the overall satisfaction on the experience with the Enact platform. Among the additional comments that were collected, however, it is important to note that one male participant, age 32, reported that the situation in which he tried the software was not optimal, mentioning "noises and chaos which did not allow to conduct a deep evaluation".

Questions	Average points / 5
How appropriate and natural were the emotions expressed by the artificial bot?	3.3
How much realistic did you find the conversation with our artificial bot?	3.53
How would you rate the user interface (the controls, buttons, settings, etc.) of the ENACT Game?	3.43
How interesting were the information and the profile given at the end of your game session?	3.68
How would you rate the graphics and video quality of the ENACT Game?	3.28
How would you rate the usefulness of such game in dealing with real life situations?	3.86
Can you rate your overall experience with the ENACT Game?	3.79
Would you play this game again with different scenarios and characters?	95.9 % Yes, 4.1% No

Table 5.4: Means of the participant score obtained by each question in the second demo evaluation - Italian sample

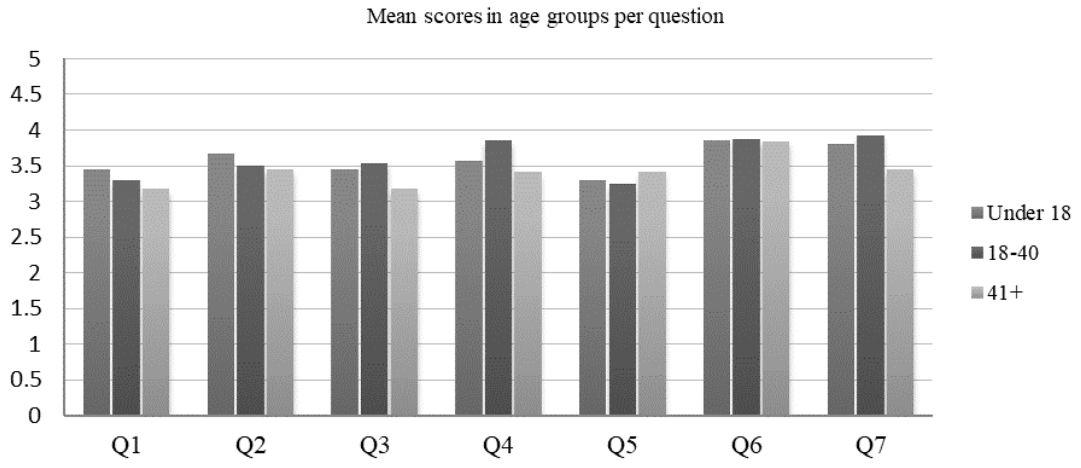

Figure 5.6: Bar chart of the difference in age groups obtained during Enact second evaluation in the Italian sample.

5.4 Second release validation: Materials and Methods

The second validation of the platform aimed at investigating the effect of adding the Compromising style to the accuracy of the Enact negotiation style prediction in comparison with the ROCI-II test only. This investigation was conducted within the scope of the ENACT project and in collaboration with all the project partners. All scenarios, feedback and introductions were translated from English to Italian, Spanish and Turkish using the back translation protocol. The back translation protocol (Hambleton & Bollwark 1991, Potaka & Cochrane 2004) is usually adopted as a practice in test adaptation, and consists in a continuous translations from original to target language and vice versa, made by several different translators. This process allows to control that important test features are retained from language to language.

The procedure by which the participants were selected and invited to participate to the experiment is the same as the previous validation process: they could be recruited both in person or online and received the same instructions as explained in Chapter 4 Section 5.2. The only difference in the protocol was that participants were not presented a battery of tests but only the ROCI-II and always before to interact with the Enact platform. The sample of the validation of this second release consisted of 130 individuals, 104 males and 26 females, aged between 13 and 60 (mean = 21.44 and st.d. = 9.02) and gathered by all partner projects. Participants were drawn from the 5 target populations: high school students (Mondo Digitale, Italy), researchers in formation (University of Naples Federico II, Italy), SME entrepreneurs (Fundetec, Spain), undergraduate students (Plymouth University, UK) and athletes (The Turkish Ministry of Youth and Sport, Turkey). The participants recruited from Plymouth, in particular, were 72, aged between 13 and 60 (mean = 23.58 and st.d. = 9.45). Also in this case, no significant statistical difference has been found between the distributions, so the sample was analysed as a whole.

Style as measured through Enact	Cronbach's alpha	Correlation with ROCI-II counterpart	Correlation p value
Dominating	0.524*	0.13*	0.008*
Integrating	0.543	0.23	0.051
Obliging	0.361*	0.18*	0.013*
Avoiding	0.249	0.04	0.320
Compromising	0.071	-0.17	0.121

Table 5.5: Cronbach's Alpha, Correlation with ROCI-II and relative p value computed across scenarios. Values marked with asterisk are statistically significant ($p<0.05$).

5.5 Second release validation: Results and discussion

The values of the reliability obtained by Enact scores in the new scenarios are shown in Table 5.5. Even though the reliability measured by Cronbach's alpha is lower than that measured on the first version of the platform, this is a direct consequence of halving the number of items inside the Enact platform (5 scenarios instead of 8, 5 interactions instead of 7), since as it is known in literature this measure (either Cronbach's or standardised) is a function of the number of items in a scale and is affected by the number of items, item intercorrelations and dimensionality (Cortina 1993). What it is instead to be stressed is that a significant correlation between the ROCI-II and the Enact test has been found in 2 styles on 5 (Dominating and Obliging), number that increases to 3 - adding the Integrating style - after the scores are corrected and optimised discarding the scenarios which do not provide informative data, as shown by Table 5.6. Interestingly, correlations computed using only the last answer given in each scenario - that asked the user to rethink of the whole interaction - showed similar results. Measuring Avoiding and Compromising styles remains still problematic, and this can be related to the fact that they presents a noticeable correlations with Obliging and Integrating styles respectively and are probably not well identified by the respective items within the scenarios.

Style as measured through Enact	Cronbach's alpha	Correlation with ROCI-II counterpart	Correlation p value
Dominating (scenarios 3 and 4)	0.472*	0.182*	0.047*
Integrating (scenarios 1 and 2)	0.583*	0.212*	0.022*
Obliging (scenarios 3 and 5)	0.415*	0.221*	0.011*
Avoiding	0.249	0.04	0.32
Compromising	0.071	-0.17	0.121

Table 5.6: Corrected Cronbach's Alpha, Correlation with ROCI-II and relative p values computed across scenarios to optimize reliability and correlation with ROCI-II. Values marked with asterisk are statistically significant ($p<0.05$).

5.6 Exploring the differences between Enact and traditional testing

Apart from the validity and the quality of the interface, another interesting comparison between Enact and the ROCI-II lies in whether there are other advantages in terms of bias reduction, engagement and motivation when interacting with one or the other instrument. While serious games use advantages have been already demonstrated in some experiments (cfr. Chapter 2 Section 2), there is no direct comparison when those are used as replacement or support to psychological tests. For this reason, an experiment was designed to explore the further advantages in terms of perceived pleasantness, usefulness, motivation and time spent between the two testing methodologies.

5.6.1 Advantages of innovative methodologies

Traditional psychological testing presents problems that may threaten the validity of the tests used. In fact, we can find external or internal disturbance variables and/or biases that can affect the performance of the participants, such as other people entering the testing room, or internal psychological processes, like boredom or poor motivation, which may be due to the excessive length of a questionnaire or its repetitiveness. We can also notice subjective disturbance factors such as the "Hawthorne effect" (McCarney et al. 2007), which causes a modification of the participant behaviour caused by he or she becoming aware of being the object of observation, the "acquiescence"

bias, that leads the subject to give positive responses to questionnaires so to confirm the expectations of the experimenter, or "social desirability", that leads the participants to respond according to what they think is socially accepted regardless of what they actually believe (Furnham 1986).

The emergence of innovative methodologies that use virtual platforms to measure psychological variables, like serious games and in particular Enact for the field of negotiation and conflict management, could solve or reduce some of the mentioned problems of traditional methodologies, or make sure that technological testing's limits are compensated by the advantages ofthese methodologies, and vice versa. The following experiment aims to demonstrate that using serious games in psychological testing can be advantageous enough to avoid the onset of certain threat variables that arise in the performance of traditional tests, especially in the area of involvement and personal motivation to perform a given task.

5.6.2 Implicit and explicit motivation

The research hypothesis behind this experiment is that, being the serious game Enact an interactive methodology that has game-like sources of internal motivation, it can be more involving and motivating for users, compared to the ROCI-II, a traditional paper and pencil test. Therefore, using serious games in psychological testing could help solve the problem of boredom and the lack of motivation to complete the test. To investigate explicit motivation an ad hoc questionnaire was created that will be described in the next section, while implicit motivation and pleasantness was operationalized as time perception. A positive correlation between Enact and the variables analyzed - motivation and pleasantness - in the researched sample, would show that using new technological methodologies for psychological assessment, such as serious games, could:

- improve engagement and personal motivation to complete a given task;
- provide new information to allow the development new scenarios for serious

games and applications such as Enact.

5.6.3 The psychological construct of time perception

Fechner is a pioneering experimental psychologist who started to analyze the changes in the individual perception of time, trying to evaluate the relationships between objective time and subjective time, and the current state of art in literature comprises countless proofs that this phenomenon occurs and is related to several internal and external conditions as well as subjective distortions (Grondin 2010). Time perception has already been used as a measurement of implicit pleasantness or motivation. Thayer & Schiff (1975) investigated how pleasantness impacted time perception by asking participants to estimate how long they spent in an eye-contact social task and manipulated the facial expression of the individual they were looking at, a critical variable that can affect the perceived valence of the social interaction. The eye-contact was estimated as longer when the stimulus (facial expression) was negative-unpleasant (scowling-angry) than when it was positive-pleasant (smiling-friendly). Concerning time perception and motivation, one example is Watt (1991), who investigated the effect of boredom proneness on perception of time. In the experiment, undergraduate students were divided into two groups, highly boredom-prone and low boredom-prone and were asked to complete a repetitive number circling-task. Highly boredome prone individuals reported subjective time as passing more slowly during the task.

5.7 Materials and Methods

For the purpose of the research, an ad hoc paper questionnaire was created, as a methodology for assessing the explicit and implicit levels of motivation and pleasantness of the participants. The questionnaire is composed as follows:

- 1st page: information sheet with informed consent
- 2nd, 3rd and 4th page: pleasantness self-report and demographic questionnaire
- 5th and 6th page: debriefing sheet

The demographic questions were placed in the last sheet to avoid that data collected, being personal data, influenced the participants in their responses during the administration process. Each sheet of the questionnaire shows at the top right an identification code used to guarantee the anonymity of the individuals participating in the research. Before starting the administration, students were asked to identify themselves through a code provided by the experimenter, a unique code reported on the information sheet, on the self-report and on the demographic questionnaire. In the self-report questionnaire, participants were asked to respond to items through a three or five-point Likert scale. The questionnaire evaluated:

1. the personal perception of time, interest in performing the two tests, pleasantness and attention;
2. the perception of having acquired greater awareness of their style of managing conflict after each of the tools, and the consequent motivation and will to deepen or not the subject;
3. estimation of the time spent in the administration of the two assessment tools. Through the demographic questionnaire, each participant was asked to report: their year of birth, their employment and type of work, the qualification and the current attendance, if applicable, of a course of study, marital status, current accommodation (and possible cohabitation), and nationality.

All the sheets composing the questionnaire are included in Appendix K. The participants had the task of completing the two assessment tools, i.e. the ROCI-II and Enact, and then completing the questionnaire.

The research was conducted at the Natural and Artificial Cognition (NAC) Lab at the University of Naples Federico II and the population of the sample is Italian. The sample consists of 100 participants, randomly selected in the Department of Humanistic Studies, and equally distributed between males and females (N = 100 of which M = 50 and F = 50). The age of the sample is between 18 and 35 (mean = 23.98, st.d. = 2.749,

Item	Enact	ROCI-II	I don't know
You took more time to finish	87%	9%	4%
You found more interesting	91%	5%	4%
You were more keen on completing	84%	10%	6%
You paid more attention more completing	77%	20%	3%

Table 5.7: Frequencies of the answers in Section 1 of the self-report questionnaire

female mean = 24, male mean = 23.96).

5.8 Results

77% of the sample is made up of non-working students, 21% of working students (of which 80.95% work in contact with people, 4.7% work in groups and 14.28% work not in contact with people) and 2% of the sample consists of interns and PhD students. 49% of the sample consists of BSc or MSc psychology students. The remaining 51% is made up of other BSc or MSc students in 9% language, 7% law, 7% philosophy, 5% literature, 2% history, 2% physical education, 2% biology, and 1% from each of the following courses: history of art, phonology, physics, geology, speech therapy, engineering, digital cultures, political science, music, economics. Finally, 2% are interns and PhD students, and 5% is not specified. For convenience, the sample was divided into two groups: a group formed by graduates and/or postgraduates in psychology and one formed by graduates and / or postgraduates in other addresses.

5.8.1 Results on the explicit pleasantness

Regarding the explicit perception of pleasantness, data from the answers given to the self-report was analyzed. The results of the first three sections of the self-report are shown respectively in Table 5.7, Table 5.8 and Table 5.9

From the analysis of the data of the first section, the following results emerged:

1. Most of the sample (87%) believe that it took longer to complete Enact than ROCI-II.

2. Although students perceived that more time was spent on Enact, most of them

Item	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
I think it took me too much time to finish the Enact game, compared to the ROCI-II	6%	37%	28%	26%	3%
I felt bored completing the ROCI-II test compared to the Enact game	4%	27%	29%	34%	6%
I felt more pleased completing ROCI-II questions than those of the Enact game	19%	53%	18%	10%	0%
I got distracted several times while completing the ROCI-II compared to the Enact game.	14%	38%	24%	19%	5%

Table 5.8: Frequencies of the answers in Section 2 of the self-report questionnaire

Item	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
While completing the Enact game, I gained more knowledge concerning my ability to handle conflicts	1%	23%	26%	48%	2%
I am motivated to further deepen this new knowledge in Enact (concerning my ability to manage conflicts)	0%	11%	30%	48%	11%
While completing the ROCI-II test, I learned more information concerning my ability to handle conflicts	8%	45%	29%	17%	1%
I am motivated to further deepen this new knowledge in ROCI-II (concerning my ability to manage conflicts)	1%	20%	31%	40%	8%

Table 5.9: Frequencies of the answers in Section 3 of the self-report questionnaire

found Enact more interesting (91%) and pleasant (84%) than the ROCI-II. Also the level of attention was considered by most students (77%), higher while completing the serious game compared to the traditional test.

From the analysis of the data of the second section of the questionnaire, it is possible to evidence that:

1. 43% of the sample, thus the relative majority, do not believe that it takes too much time to finish Enact compared to ROCI II, while only 29% believe the opposite is true.
2. 40% of the sample declares that completing the ROCI-II was a more boring task compared to Enact.
3. The majority of the sample (72% of the pool) do not prefer to complete the ROCI-II rather than Enact.
4. More than half (52%) of the sample declares that they were not distracted during the execution of ROCI-II compared to Enact, and only 24% declares the opposite.

From the analysis of the data of the third section, the following results emerged:

1. The majority (51%) of the sample declares that, during the interaction with Enact, new knowledge regarding the ability to manage conflict was acquired and even more (59%) declares that it is motivated to deepen this knowledge throughout the platform.
2. During the execution of ROCI-II, on the other hand, 53% of the sample states that they have not learned anything new.

To verify the presence of significant differences in the distribution of the gender subsamples, the Mann-Whitney U test was used. The following gender differences were found statistically significant:

- the serious game Enact is evaluated differently (in terms of the perception of learning new knowledge concerning conflict management skills) based on gender ($p = 0.001$). Specifically, most female participants (66%) positively assess Enact, (22% do not take a position on this and 12% evaluate it negatively); while male students are almost equally divided between those who positively rate it (34%), those who do not take a position (30%) and those who evaluate it negatively (36%);

- the ROCI-II test is evaluated differently (in terms of perception of new acquired knowledge regarding their conflict management skills) by participants based on their gender ($p = 0.038$). Specifically, almost half of the sub-sample of female students (46%) negatively assesses ROCI II, while the other half of the sub-sample either takes no position (30%) or evaluates it positively (26%); the percentage of male students that rate it negatively instead is much higher (60%), 28% do not take a position on it and only 12% rate it positively.

Summarizing the results, it is possible to declare that most of the sample in this particular section of the questionnaire correctly estimated that it took more time to complete Enact rather than the ROCI-II. However, the serious game was perceived as more interesting and enjoyable to play, while the traditional test was less appreciated. More than half of the entire sample also declares to have learned new knowledge about their way of managing conflict with others thanks to the use of the serious game Enact (while the opposite is true about the traditional test) and is motivated to deepen this knowledge.

5.8.2 Results on the implicit pleasantness

Regarding the implicit perception of pleasantness, the first thing we wanted to analyze is the comparison between the real time used to finish the two assessment tools, and the subjective time perceived by the participants. Table 5.10 shows the statistical averages and the standard deviations referring to the real time (expressed in minutes) used to perform the ROCI-II test and the Enact serious game, in the whole sample and

Gender	Stat	Real Time Enact	Real Time ROCI-II
F	Average	12.38	5.50
	St. Dev.	35.57	17.41
M	Average	12.00	6.42
	St. Dev.	33.38	17.27
TOT	Average	12.19	5.96
	St. Dev.	34.37	17.86

Table 5.10: Average time (minutes) taken by the participants to complete the two assessment tools divided by gender.

Item	Less than 10 minutes	Between 10 and 15 minutes	More than 15 minutes
Enact	18%	76%	6%
ROCI-II	88%	12%	0%

Table 5.11: Time estimated spent on the two assessment tools by the participants

in the sub-samples emerged from the classification of the participants based on their gender.

From the analysis of the sub-samples, it emerged that there are no significant differences between: duration of execution (real time) of Enact in male and female participants ($p = 0.697$); duration of execution (real time) of ROCI-II in male and female participants ($p = 0.931$). Regardless of their gender, participants spent about 6 minutes for completing the ROCI-II and 12 for completing Enact.

To express their perception of time, participants responded to the items in section 4 of the self-report by estimating the time spent on each assessment tool among: less than 10 minutes, between 10 and 15 minutes and more than 15 minutes. The results are summarized in Table 5.11.

In contrast to the lack on any effect on real time spent, the gender variable determines effects on the perceived duration of the two administrations: the time used to complete Enact, in fact, is estimated differently according to the gender ($p = 0.034$). In particular female students tend to more accurately estimate the execution time of Enact, while male students tend to slightly underestimate it, as shown in Figure 5.7. Time perception seems to be influenced by another variables, the type of degree (gradu-

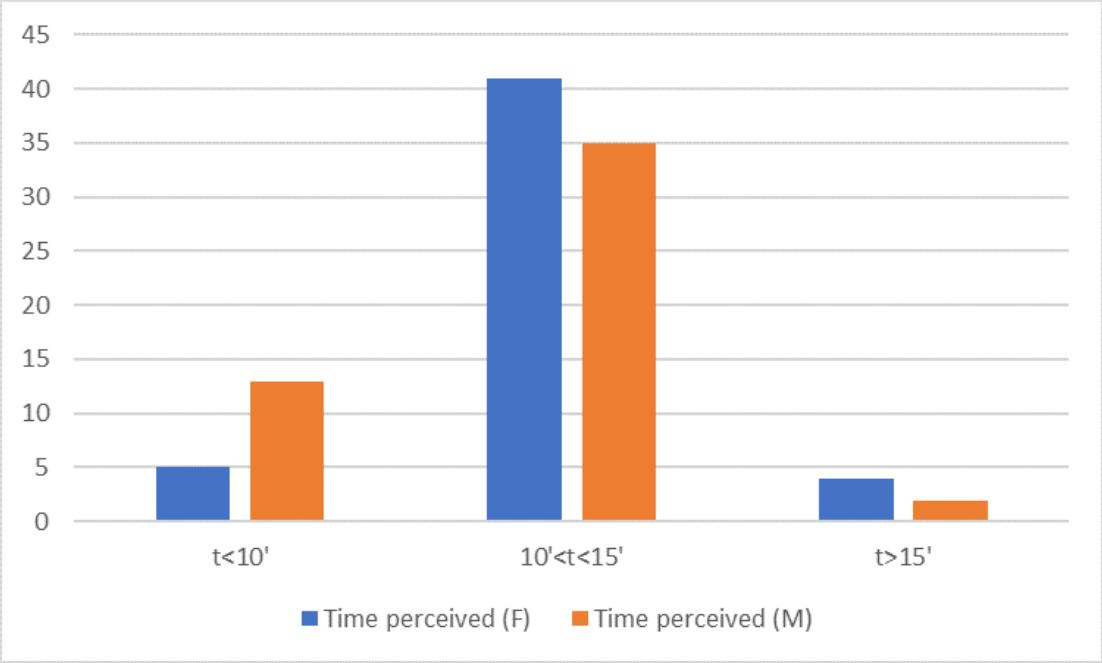

Figure 5.7: Gender differences in the perception of time spent completing Enact. Frequencies are displayed.

ate/postgraduate in psychology versus graduate/postgraduate in other subjects). The two groups are significantly different (p = 0.015, measured with the Mann-Whitney U test). Figure 5.8 shows the distribution of the answers given to the self-report by the two groups, concerning the time perceived in executing Enact.

It emerged that participants in the group "Psychology graduates/postgraduates" were able to correctly estimate the time spent completing Enact, while those enrolled or graduated in other courses did present a slight bias towards the underestimation of this time. The following variables were not analysable, since the variables were not equally distributed (not being subject to experimental control): marital status; employment; qualification; current accommodation; nationality; the five negotiation styles. The negotiation styles most used by the participants turned out to be the Integrating and the Compromising ones, used by 55% and 24% of the sample respectively; while only 7%, 5% and 2% of the sample use predominantly a Avoiding, Dominating and Obliging style. Moreover, some subjects are predominantly adopting two negotiation styles: 1% of the sample uses a Dominating/Compromising style and 6% of the sample an

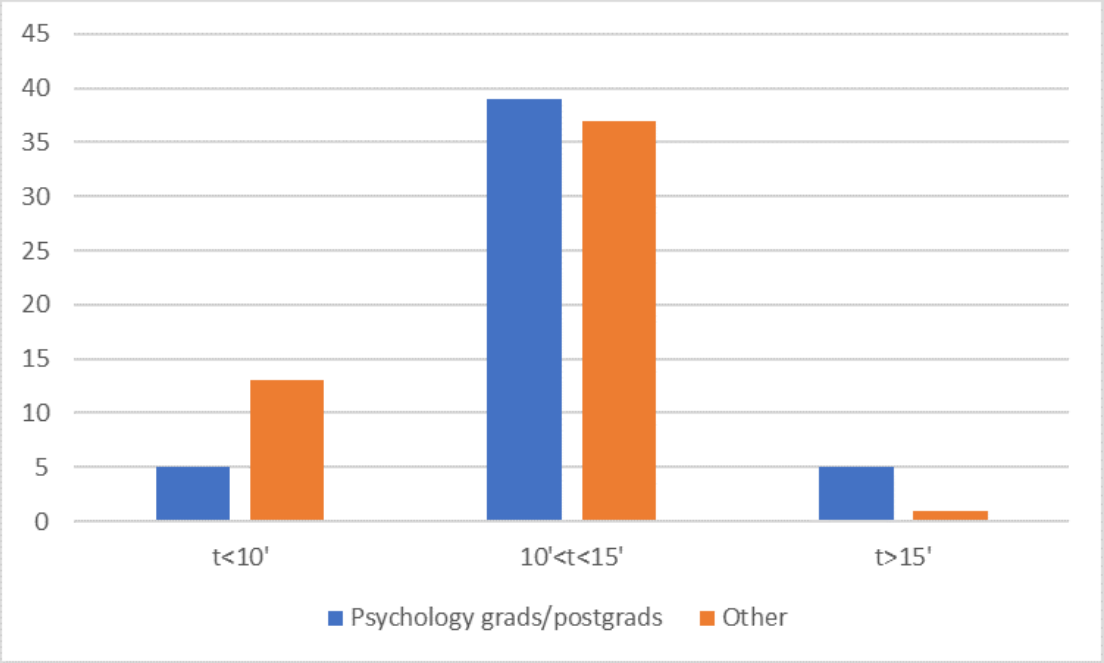

Figure 5.8: Course of study differences in the perception of time spent completing Enact. Frequencies are displayed.

Integrating/Compromising one.

5.9 Discussion

From the results, it emerged that most of the sample has correctly estimated the time taken to perform the two evaluation tools, recognizing that the serious game needs more time. However, the serious game was perceived as more interesting and enjoyable to play, while the traditional test was less liked. This data confirms the hypothesis that the perception of explicit pleasantness of the participants, in reference to the two evaluation tools, is higher towards the serious game Enact.

The gender variable does not determine effects on the actual duration of Enact and ROCI II, but it significantly determines the subjective perception of time: female subjects tended to estimate more correctly the execution time of Enact, while males tended to slightly underestimate it. Females are also more likely to positively evaluate Enact (in terms of perception of acquiring new knowledge regarding their ability to manage conflict), while males evaluate ROCI-II, more negatively than females. The lack of data

consistency could be due to a problematic in section 3 of the questionnaire, which may not be completely clear and easy to understand. Data related to gender differences sees the female sample having a significantly higher score in the items that express a greater appreciation explicit towards Enact (unlike males who maintain on average more neutral responses. This is in accordance with data collected on the Italian sample in the last Enact evaluation and reported in Chapter 5 Section 4, where it was found that the female sample, compared to the male one, gave significantly higher scores in items that concur to confirm an explicit high level of appreciation towards Enact.

Interesting is the interaction that emerged between the sample group that was graduate/postgraduate in psychology and the perception of time in reference to Enact: participants belonging to the group graduate/postgraduate in psychology correctly estimated the time taken, while the other participants tended to underestimate it, evidently incurring in an evaluation bias that can be interpreted as an influence given by their interest in the game. But why graduates/postgraduates in psychology were not affected? Literature researches show that graduates/postgraduates in psychology are more aware of the subjective aspects that influence our perception of time, as they, during their academic journey, come into contact with psychological constructs such as human metacognition and the biases associated with it. In literature there are theories on metacognition that argue that there is a causal link between metacognitive reflections, control processes and cognitive performance (Deroy et al. 2016), also concerning time perception (Chakroun et al. 2014). A metacognitive intervention may be suggested for example for the treatment of children with specific learning disorders (Wong 1991) or adults with personality disorders (Dimaggio & Lysaker 2015).

5.10 Conclusion

The Chapter described the improvements and the changes made on the Enact platform and the subsequent evaluations and validation phase of the second release. The changes on the platform concerned first of all the interface, with a renewed graphics, more graphically clear buttons, scenario introduction, animations and visually dis-

tinctive choices. Also, distinctive shape bubbles and gibberish sound was added to make the tone of the virtual agents' sentences less ambiguous. The Compromising style was added as separate style both in the Assessment and in the Training Environment, and a new Compromising scenario was added. Scenarios were reduced in their number, shortened and made more interactive. These changes were released and two evaluations were conducted on two different samples, one from UK and one from Italy, to assess the participants' perception of the Enact interface, usefulness and user-friendliness in comparison with that of the first release. Results were consistent and a slight improvement emerged concerning the evaluation of the game graphics. After the evaluation, a validation phase was conducted by all project partner to quantify the changes in terms of reliability of the tool after the introduction of the fifth style. While Avoiding and Obliging still remain problematic to predict (also due to a much smaller sample), the ability to correctly assess the other styles remains high. Finally, a research to explore the implicit and explicit pleasantness toward the Enact game in comparison with the ROCI-II was conducted, using a questionnaire created ad hoc. Results show that Enact was assessed as significantly more pleasant and interesting, and participants who were not graduates/postgraduates in psychology also underestimated the time spent interacting with the platform. Also, the majority of the sample reported that new knowledge regarding negotiation and conflict management was acquired during the Enact assessment, while the opposite was true for the ROCI-II.

Chapter 6

A Natural Language Processing model for a statistical prediction of negotiation styles

After the validation and the evaluation phase, the next aim of this research was to investigate a way to allow the platform to be more adaptive and interactive so to include, on one hand, a wider student model that took into account natural language information, and on the other hand a more direct and immediate way for the user to make their negotiation style choice more deliberately within the scenarios. For this purpose, a natural language processing statistical model to predict the user's negotiation style during the interaction with virtual humans in Enact is presented. Other than that, a technology-enhanced methodology for the collection of a dataset of natural language negotiations is shown. Sentences written during the interactions of the users with virtual artificial agents inside the ENACT 3D game platform were collected, along with the associated negotiation style of the user as measured by the standardized test ROCI-II. A corpus of documents was built tagging the sentences and their related user negotiation style, and an analysis was performed to highlight consistencies and differences in the term frequency, use of personal pronouns and virtual agent style. The documents were analysed applying text mining techniques and consistent statistical differences among the styles in agreement with their theoretical definitions were found. These results allow both to validate Rahim and Bonoma's negotiation styles classification as statistically relevant and solid as well as to provide a deeper understanding of the model in relation to the words and pronouns usage.

ONE of the main challenges of providing a reliable assessment for ill-defined domains concerns the need to structure platforms able to face the problem of the multitude of possible student behaviours and responses with different shades of correctness. In fact, in the domain of soft skills, what can be acceptable in some situations may not be adequate in others, as there are variables like context, the characteristics of the individual or the interpersonal attitude of the involved agents that influence the learner's perception of the environment. Other differences from hard skills domains lie

in that a certain amount of information is needed about the student in order to correctly categorize each user. In particular, the system's feedback on the correctness of the answer is delayed until a partially complete student model has been made. Additionally, in ill-defined domains, most of the tasks posed to the student do not have a single correct answer that applies to all contexts, and this ambiguity is reflected during the interaction between the tutor and the learner, raising the need of shaping the user model through time. In most serious games and game-based simulations, data provided by the user is collected in the form of multiple choices (Dell'Aquila et al. 2017), like in the first and second releases of Enact, or non-verbal information like eye-gaze and gestures (Baur et al. 2013), or facial expression (Becker-Asano & Wachsmuth 2009). While natural language processing techniques have been used to generate human-like negotiations, via Wizard-of-Oz or machine learning algorithms (e.g. Gratch et al. 2016), the information about the sentences provided by the learner have never been included in the user model and used as indicators and variables to categorize soft skills, and in particular negotiation styles, and build a personalized environment. The present study aims at presenting the first step to fill this gap by designing a natural language processing model that uses statistical inference to map the features of each user sentences and utterances used during a virtual negotiation to his or her predominant conflict management style. In particular, this work aims at investigating the statistical relationship between the users' predominant negotiation style and the structure of the sentences they are more likely to use by mining a corpus of documents collected during simulated negotiations. The simulation environment used within this experiment is taken from the Enact platform (Marocco et al. 2015).

6.1 Natural language and student modelling

Considering the importance of adding the analysis of natural language to the student model in serious games, ITSs and generally technology-enhanced learning tools, one of the challenges is represented by the necessity of finding the correct way to measure variables inside the language that can correctly represent the investigated constructs.

Operationalizing negotiation, bargain, conflict handling in terms of language is a hard task which is still unexplored. The most common approach requires the analysis of the parties' behaviour in terms of linguistic blocks, which can help interpret the concern and the power of one party on the other. These include blocks that can describe stalling or making threats or demands (De Dreu et al. 1998, Pruitt & Lewis 1975) or persuasion (Adair et al. 2001, Putnam & Jones 1982), or relational versus rational power (Giebels & Taylor 2009). These blocks or coding schemes can be theorized and refined so that it is possible to investigate when they tend to occur and how frequently they occur. Once these blocks are extracted from documents, one approach to studying them is using statistical methods, for example log-linear analysis to examine the probability of different coding blocks occurring after a given stimulus (Olekalns & Smith 2000) or Markov chain models (Weingart et al. 1999). Talking It Through: Communication Sequences in Negotiation

Rather than identifying whole sequences, another effective way to explore negotiation utterances, especially in relation with Rahim and Bonoma's five style model, is personal pronoun analysis. In literature, many researchers investigated the relationship between personal pronoun use and the level of engagement compared to the other party (e.g. Camden & Verba 1986, Fraser 1980). The level of engagement and detachment is related to concern and involvement in the interaction. In parallel with Rahim and Bonoma's dimensions of concern for self and for others, highly engaged agents are cognitively involved with the other agent of the negotiation, while low-engaged agents show detachment from their own as well as others' conversation subject and arguments. In particular, one pioneering study in communication (Cegala 1989) takes into account three types of pronouns (first person, second person and relational pronouns) and uses the ratios second/first person and relational/first person pronouns to estimate the level of cognitive and communicative involvement the speaker has with the other party. The theory hypothesises that second/first person ratio indicates the position of the conversation on a continuum from concern for the other to concern for self, while the relational/first person ratio indicates how integrative the approach is.

In the present work, a methodology for collecting natural language negotiations is presented. Also, a statistical perspective that investigates both the documents as a whole, analyzing the similarities and the differences in the pattern, word position and frequencies, sentence length and width of vocabulary and a focus on the use of personal pronouns will be taken, as explained in the next sections.

6.2 Materials and Methods

In order to create a natural language model of the distribution of the negotiation style dimensions according to the words used by the users, screenshots of the introductions and of the states were taken from the Enact game platform for each of its five scenarios, and the user was asked to respond to the virtual agents in the scenarios writing his or her own words. Then, the participant was asked to fill the ROCI-II psychological test, in order to obtain the width of each psychological dimension of his or her own negotiation style. The output of a ROCI-II, in fact, is represented by 5 scores, each representing an independent negotiation style dimension (Integrating, Avoiding, Obliging, Dominating, Compromising).

The requisite for participating to the experiment was exclusively the ability to speak fluent English. The experiment designed for the sentence collection was divided into two parts: the first part included the collection of the sentences, while the second part concerned the administration of the ROCI-II questionnaire. In the first part, the participant was asked to fill the consent form and then was presented a series of screenshots drawn from the Enact scenarios, for a total of 20 images. In detail, for each of the five scenarios, the participant was presented with two introductory scenes containing information about the situation in which the negotiation took place, and two screenshots of the interactive scenes with the virtual agent, including facial expression, gesture and tone. The two scenes were extracted randomly among the five possible for each scenario. Under the images of the interactive scenes a box was placed, and the user was asked to write the sentences that he or she would say in that situation. An example of the interface is shown in 6.1.

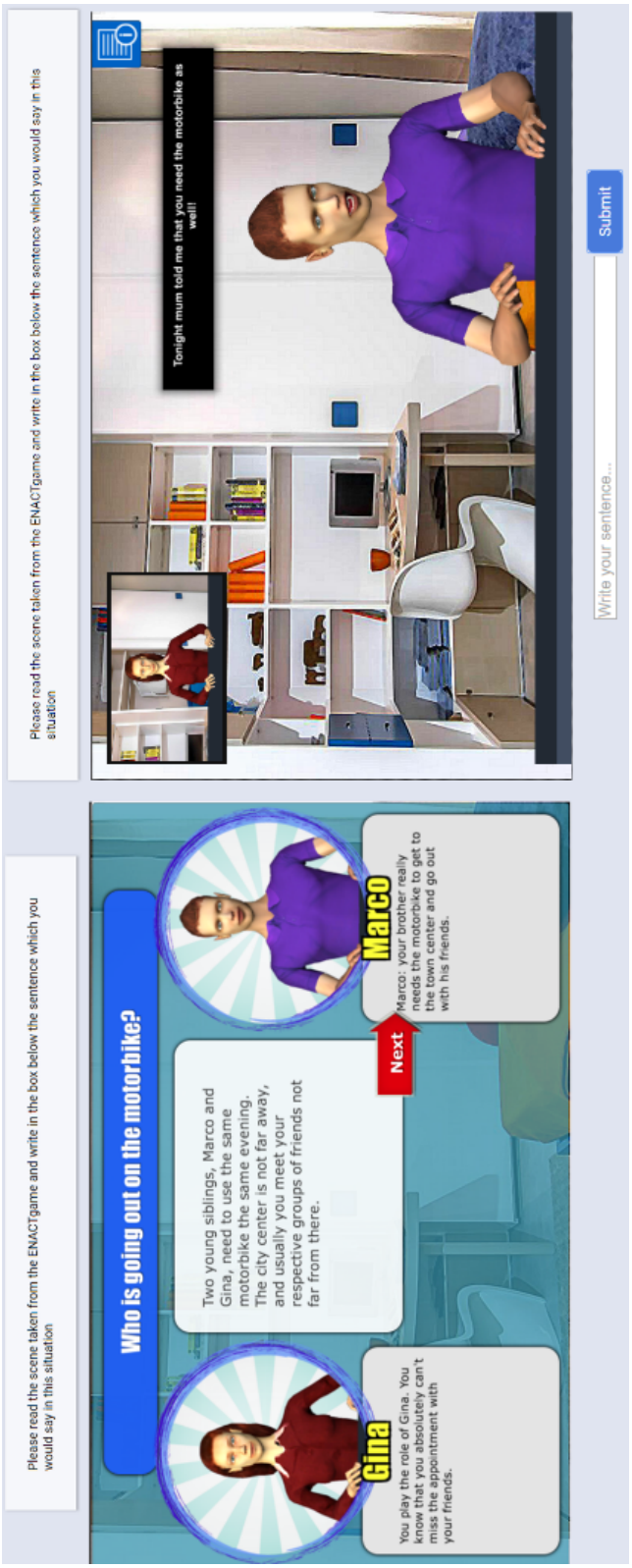

Figure 6.1: Sequence of the scenes in the natural language experiment. From the left, screenshot of the first introduction scene and screenshot of one interaction state of the first scenario.

Since the screenshots of the interactions had to maintain their chronological order to follow the flow of the scenarios, not all the random combinations were meaningful. To rule out those cases, the random presentation of these two screenshots was organized into 5 random seeds that initialized the number of both the interactions extracted and presented for each scenario. For example, seed number 1 contained the interaction 1 and 2 for each scenario, therefore if a user was randomly assigned to seed number 1 he or she will be shown screenshots of interactions 1 and 2 for the scenario. The sentence written in the box could be maximum 100 characters long. At the end of this first part of the experiment, the participant was asked to proceed to the second part of the experiment, and was asked to complete the ROCI-II questionnaire. The whole procedure took on average 30 minutes.

A total of 425 subjects (age mean = 22.84, st.d. = 7.47) participated in the experiment and were recruited as specified in the previous chapter. A total of 4250 sentences were collected. Each sentence was labelled according to the five continuous values indicating the five dimensional scores obtained by its author. ROCI-II scores for each style can have values between 1 and 5, as per Appendix C, and were normalized to a range between 0 and 1. A reliability test on the ROCI-II items' scores was conducted to confirm that the validity of the test administered to the sample is comparable to that obtained by Rahim's ROCI-II validation (Cronbach's alpha = 0.776, Standardized item's Cronbach's alpha = 0.797). The distributions of the original scores of the ROCI-II questionnaire obtained by the participants for each style dimension is shown in the boxplots in Figure 6.2.

6.3 Data preparation and pre-processing

Several analyses were conducted on the dataset formed by all the sentences collected in the experiments. First of all, an analysis involving the pronouns used in relation to the score obtained on each negotiation style dimension was conducted, taking into account the five values obtained from the ROCI-II. The hypothesis was that pronouns' use can be correlated with the two dimensions that identify the five styles as modelled

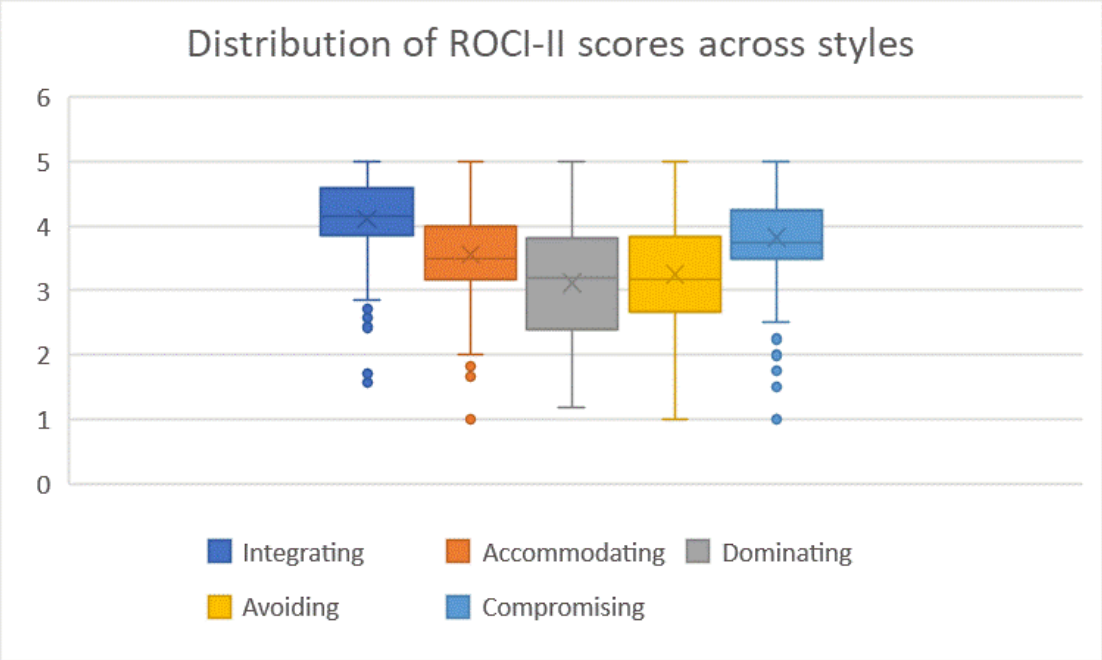

Figure 6.2: Distribution of the ROCI-II scores obtained by the participants for each negotiation style.

in Rahim and Bonoma's theory: concern for self and concern for others. In particular, styles with a high concern for self (Integrating, Dominating) would use first and second person pronouns more frequently than styles with a low concern for self (Avoiding, Obliging), and styles with a high concern for others (Integrating, would show a higher use of relational pronouns. According to the theory, the style used is also dependent on the situation, and therefore the virtual agent's negotiation style was expected to also influence the use of the pronouns, regardless of the ROCI-II scores. Then, based on the results of this investigation, a clustering of the sentences in five groups was conducted, and a further analysis on the pronoun use and clustered documents was performed.

In order to proceed with the whole analysis showed in the next two paragraphs, the sentences underwent a series of pre-processing steps which are shown in Figure 6.3 that aimed at tokenizing the words. Tokenization is a process that targets natural language text and reduces each meaningful and feature-rich word within a document to the smallest analysable unit so that words with the same root and/or derivation can be

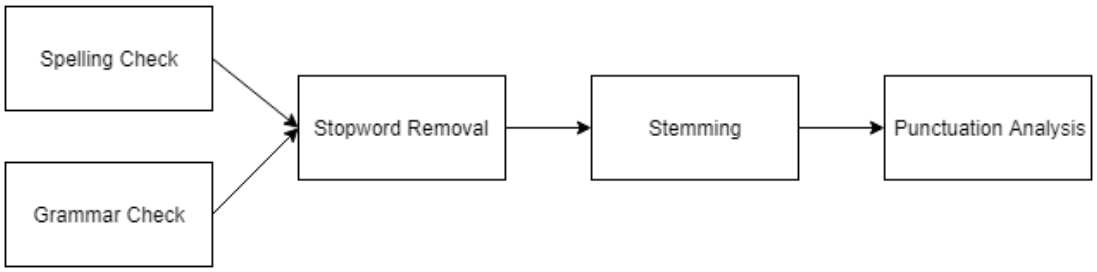

Figure 6.3: Diagram that shows the preprocessing steps operated on the natural language dataset prior to the analysis. All operations have been performed with the NLTK library, and an additional human spelling and grammar check was conducted afterwards.

considered similarly. In particular, the following steps were followed:

1. Grammar and spelling check. In order to improve the accuracy of the categorization and of the data analysis, and considering the dataset size, all the sentences were subjected to a thorough spelling and grammar correction. The corrections only included typical misspellings (e.g. Paollo/Paolo) and grammar mistakes (e.g. your/you're) and did not aim to achieve a perfect British English (for example, juvenile and slang expressions were maintained) or to alter the content in any way;

2. Stopword removal. In text-mining and information retrieval, stopwords are the most common words that do not help extract features from the dataset - since their frequency, usage and meaning remain the same across all documents - and could therefore negatively affect the results. Prior to the analysis that considered the frequency of the words in the dictionary, the common stopwords were excluded. The list is taken from the Natural Language ToolKit (NLTK) open source library, which is based on the results of Van Rijsbergen & Van (2016), with the exception of personal pronouns, which were deleted from the list so to be included in the analysis. The complete list of the stopwords excluded is shown in Table 6.1.

3. Stemming. In information retrieval, stemming is the process of reducing all derived or declined words to their base or root form. Of course, this process is language specific and can be subject to error as the English language contains

List of Stopwords
what, which, who, whom, this, that, these, those, am, is, are, was, wer, be, been, being, have, has, had, having, do, does, did, doing, a, an, the, and, but, if, or, because, as, until, while, of, at, by, for, with, about, against, between, into, through, during, before, after, above, below, to, from, up, down, in, out, on, off, over, under, again, further, then, once, here, there, when, where, why, how, all, any, both, each, few, more, most, other, some, such, no, nor, not, only, own, same, so, than, too, very, s, t, can, will, just, don, should, now

Table 6.1: List of stopwords used to pre-process the dataset

Part of speech	Original	Stemmed
Adverb	hopefully	hope
Verb	activate	activ
Verb	tanned	tan
Noun	adoption	adopt
Noun	inference	infer

Table 6.2: Examples of words reduced using the Porter stemmer

exceptions and words that may have similar roots but different meaning. Many efficient algorithms have been theorized in order to extract the words' base form, of which the most widely known and currently used is described in Porter (1980). The algorithm works like a finite state machine and contains several rules that can be applied to single words. The resulting stemmings usually do not constitute meaningful words. Examples of output words provided by the Porter stemmer is presented in Table 6.2.

4. Punctuation analysis. An analysis of the usage of punctuation was conducted in order to identify smiles and/or emoticons as well as the usage of dots. Smiles/emoticons were treated as separate words, while punctuation within the sentence was not retained in this natural language analysis since it was not part of the hypothesis.

6.4 Analysis of the relation between ROCI-II scores, Enact scenarios and personal pronouns

The analysis of personal pronouns (in all their forms, i.e. nominative, objective, possessive and reflexive), is crucial to investigate the level of engagement and concern shown

Type	Pronouns
1st Person	I, Me, Mine, My, Myself
2nd Person	You, Yours, Yourself, Yourselves
3rd Person	He, Him, His, Himself, It, Itself Her, Hers, Herself, She, Their, Theirs Them, Themself, Themselves, They
Relational	Our, Ours, Ourselves, Us, We

Table 6.3: List of pronouns used for the natural language analysis

by parties during negotiations (Cegala 1989). This section analyses the correlation in all sentences between the use of personal pronouns and the ROCI-II dimensional scores obtained by the user. All the personal pronouns used for this analysis are listed in Table 6.3.

The results of the Pearson correlation are shown in 6.4. All ROCI-II dimensions show a very unique correlation pattern and sign, whereas the Integrating and Dominating have all correlations respectively positive and negative. The correlation sign and significance level indicate that the Dominating component in the ROCI-II scores negatively correlates with the use of all pronouns, and significantly with the use of the first personal pronouns. This evidence, that seems to contradict Rahim's theory that states that highly Dominating individuals make prominent use of "I statements" (Table 4.1), can however be interpreted in an alternative way, that is in comparison with the pronoun use of highly Integrating individuals, placed on the opposite continuous in terms of concern for self and concern for other. In fact, the Integrating component, instead, positively correlates with the use of all pronouns and significantly with the amount of relational pronouns. This can be explained by the fact that highly Integrating individuals try to increase the level of mutual engagement in the negotiation and highlight possible decision that can satisfy both parties, therefore proposing and involving both self and other's centred solution. An example of this strategy is the following sentence extracted from an individual with a high Integrating dimension:

> "We can share. Bikes can handle 2 people and we can coordinate our simultaneous return".

The lack of pronoun's usage shown in relation with the Dominating component can be the consequence of not discussing different approaches to solve the situation that involves the two parties nor trying to balance the person's own will and aim with the one of the other. Examples taken from highly Dominating individuals in the dataset are

"Let's settle it now."

or

"Eat your pizza with chopsticks."

Individuals with a high Avoiding and Obliging dimensions use also less pronouns than those with a higher Compromising and Integrating dimension, leading to a further step in the interpretation: dimensions with a theoretical high concern for self and for others show a generally positive correlation with the use of personal pronouns, while dimensions with a low concern show the opposite pattern. The following sentence displays a highly Avoiding individual's strategy taken from the dataset:

"Okay whatever floats your boat."

while highly Obliging participants, for example, proposed

"Sure thing!"

and

"Sounds like a good idea!"

Concerning the analysis of the use of the pronouns in relation to the virtual agent negotiation style and to the dimensional scores of the participant as measured by the ROCI-II, the results of a Pearson correlation run for each pronoun type taken into consideration are presented in Tables 6.5, 6.7, 6.6, 6.8 and 6.9. It is possible to notice that there are only few highly significant correlations: the amount of relational pronouns

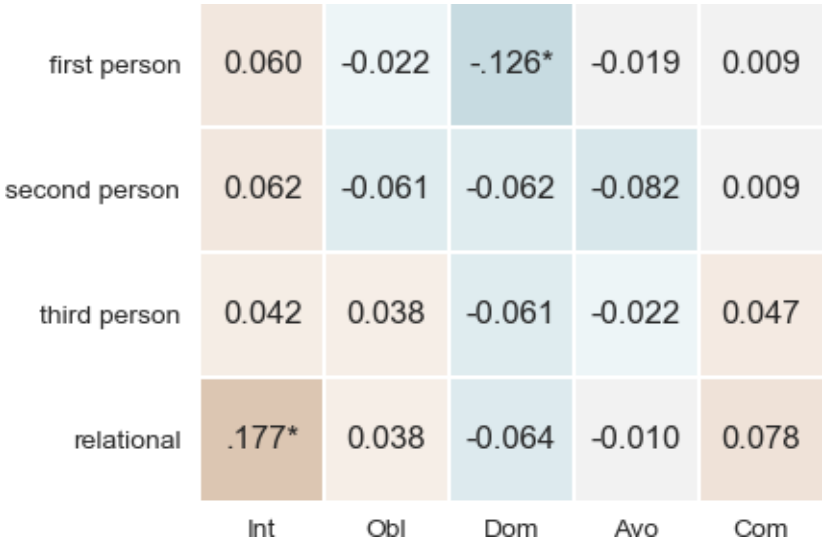

Figure 6.4: Correlation between the use of first person, second person, third person and relational pronouns on the y axis, and the ROCI-II scores of the user, placed on the x axis. Numbers and colours represent the Pearson correlation coefficients. Significant values ($p<0.05$) are marked with asterisks.

used during the interaction with an Avoiding, Compromising and Obliging virtual agent correlates with the Integrating score obtained by participants in the ROCI-II. In agreement with the definition of the Integrating style, highly integrating individuals try to balance between self and other's interests during a negotiation, and the use of relational pronouns is the most evident linguistic expression of this attempt. Even though, as it was observed, Compromising individuals may decide to solve the situation proposing a deal that is very similar to that of Integrating ones, the depth of the interest they have in others' is largely different, as evidenced by the following example taken from the dataset:

> "I can watch Masterchef with you. You should allow me to watch Report later on.".

The use of relational pronouns only correlates with the Compromising score when interacting with an Obliging virtual agent, and this could be explained by the fact that the Obliging style is the one that has the lowest concern for self, and a highly Compromising peer may wish to equally split the resources and would not feel comfortable if the

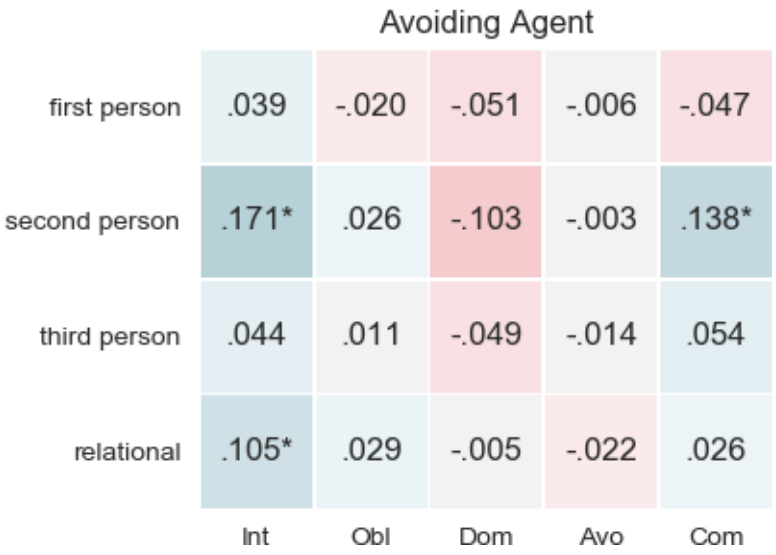

Figure 6.5: Correlation between first person, third person and relational pronouns used during the interaction with the Avoiding virtual agent and the ROCI-II scores Colours and numbers represent the Pearson correlation coefficients. Significant values ($p<0.05$) are marked with asterisks.

other party ends up with an unbalanced deal, as evidenced by the following example taken from the dataset:

> "It will be good for us to work together.".

The amount of first person pronouns used during the interaction with an Integrating virtual agent negatively correlates with the Dominating score obtained at the ROCI-II, and this can be interpreted as an effect that the integrative approach used by the other party influences - and decreases - the need to impose one's interest in general, as the Dominating individual feels that his or her interests are already taken into account. Examples of highly Dominating participants responses to Integrating agents are:

> "Why do you want Chinese?"

and

> "Do you hate Italian restaurants so much?"

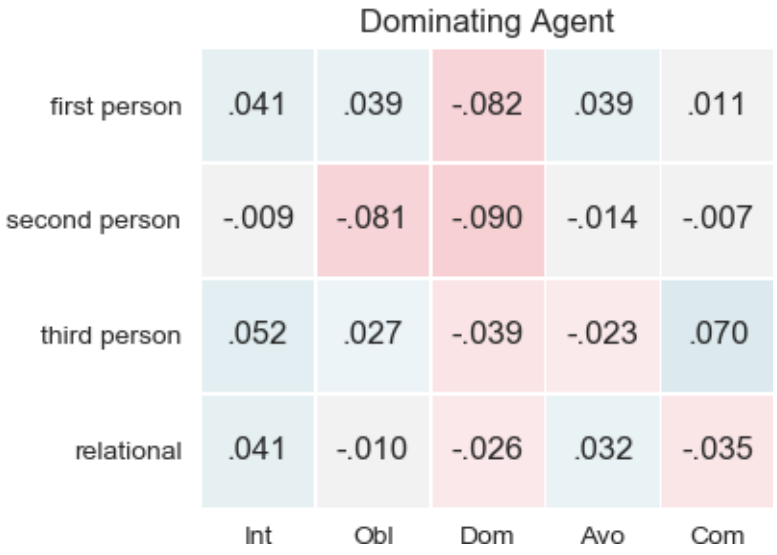

Figure 6.6: Correlation between first person, third person and relational pronouns used during the interaction with the Dominating virtual agent and the ROCI-II scores Colours and numbers represent the Pearson correlation coefficients.

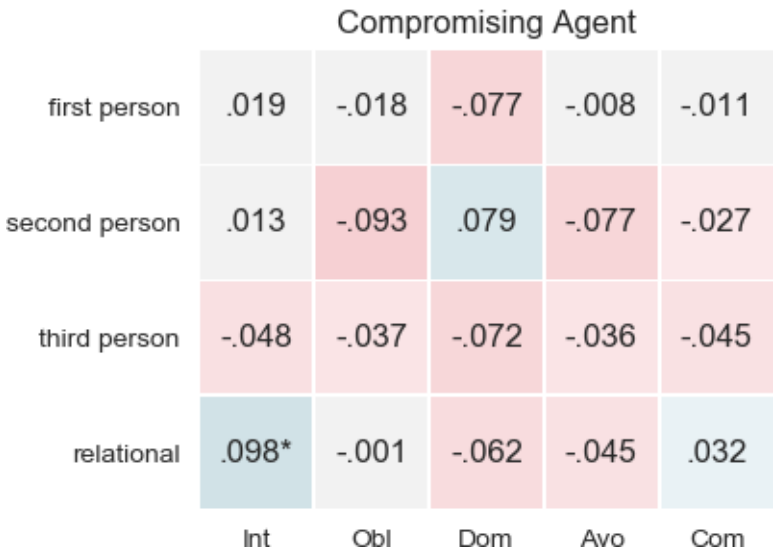

Figure 6.7: Correlation between first person, third person and relational pronouns used during the interaction with the Compromising virtual agent and the ROCI-II scores Colours and numbers represent the Pearson correlation coefficients. Significant values ($p<0.05$) are marked with asterisks.

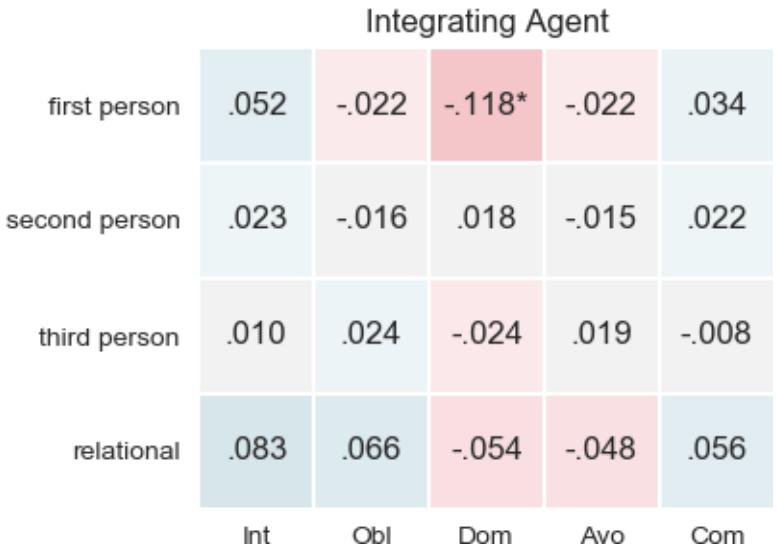

Figure 6.8: Correlation between first person, third person and relational pronouns used during the interaction with the Integrating virtual agent and the ROCI-II scores Colours and numbers represent the Pearson correlation coefficients. Significant values ($p<0.05$) are marked with asterisks.

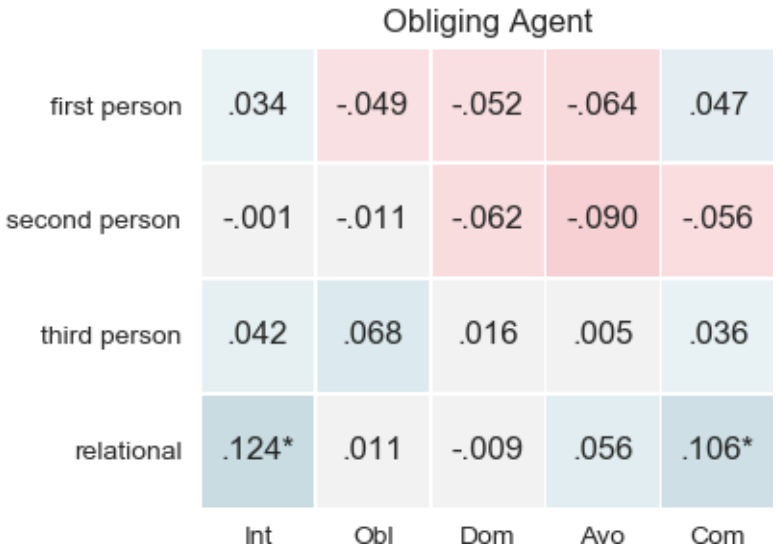

Figure 6.9: Correlation between first person, third person and relational pronouns used during the interaction with the Obliging virtual agent and the ROCI-II scores Colours and numbers represent the Pearson correlation coefficients. Significant values ($p<0.05$) are marked with asterisks.

A particular case, instead, concerns the use of second person pronouns. As shown in the tables, the use of second person pronouns with an Avoiding virtual agent strongly correlates with the Integrating and Compromising scores in the ROCI-II and, in accordance with the results of the relational pronouns, this could be explained by an attempt of the party to discuss the deal further and to increase the other's interest, as in

"Are you sure there isn't a shop in South Africa that also has the CD?"

The amount of second person pronouns in Avoiding scenarios, instead, negatively correlates with a Dominating score, and this can be interpreted by a retreat of the Dominating individual on their position and deal, rather than attempting what can be felt as an unnecessary involvement of the other.

Considering the relation between ROCI-II negotiation classification (i.e. the prevalent style) and the use of the four types of pronouns, means and standard deviations were analysed and then a ANOVA with LSD post-hoc was conducted for each of the pronoun type to investigate the significant differences. Means and standard deviations are shown in Table 6.4. Results of the ANOVA for each of the pronoun types are shown in Tables 6.5, 6.6, 6.7 and 6.8. As shown in the first table, there are many significant differences in the use of the first personal pronouns, where Integrating individuals use them significantly more than the others and Obliging and Avoiding significantly less. In general, as defined by Rahim's negotiation theory, styles with a high concern for self show a higher use of first person pronouns overall than styles with a low concern for self. Concerning the use of second person pronouns, the higher mean is still achieved by Integrating individuals, while Avoiding individuals, with a low concern for self and for others, show significantly lower values. Third person pronouns are generally much less used than the first and second person pronouns, and values show Dominating and Obliging as those that use them less, as they are styles that are particularly prone on reaching an agreement without switching the topic or the subject of the negotiation. Individuals with a prevalent Avoiding style, instead, which are the less engaged parties and had shown a general low use of personal pronouns in the other considered

		Avo	Com	Dom	Int	Obl
First Person	Mean	5.518519	5.518519	6.022727	6.306818	4.861111
	St. Dev.	3.289438	3.574632	3.372175	3.574059	2.641999
Second Person	Mean	4.962963	5.670886	5.727273	5.948864	5.472222
	St. Dev.	2.801367	3.058077	3.949469	3.151906	3.056479
Third Person	Mean	3.666667	3.696203	3.568182	3.744318	3.388889
	St. Dev.	2.44949	2.967147	2.296587	2.198044	1.916511
Relational	Mean	1.851852	2.037975	1.613636	2.392045	1.583333
	St. Dev.	1.709103	1.931142	1.587977	1.88217	1.518928

Table 6.4: Means and standard deviations between the number of personal pronouns and the ROCI-II style class.

cases, show a general higher use of third person pronouns. Relational pronouns, that are those used much less than all the others, show even more significant differences in their use among the different ROCI-II prevalent styles: Integrating still holds the highest mean, while both Obliging and Dominating, the styles which have the most unbalanced difference between their concern for self and for others, score the lowest, followed by Avoiding. Across all types of pronouns the Compromising style, consistently with its definition, shows a usage that is in the middle of all the other styles, with very few significant differences.

6.5 Analysis on the sentences classified by the user's highest negotiation style dimension

The ROCI-II scores obtained by every participant represent a unique pattern of style dimensions independent from each other. However, the previous analyses clearly evidenced that the quality of each of these dimensions is meaningful, i.e. the dimensions and their proportions are consistent with their theoretical definition and participants can be compared across dimensions after normalization. Since participants with high dimensional values showed unique words and pronouns usage regardless of their scores on the other dimensions, an assumption that can be made on the dataset is therefore that the prevalent negotiation style as measured by the ROCI-II - that will be now referred to as "class" - can be a robust predictor and can be meaningfully used to discretize into categories the dimensional continuous of the ROCI-II scores obtained by

ANOVA LSD post-hoc - First Person			
I	J	Mean Difference (J-I)	Sig.
avo	com	-0.4038*	0.000
	dom	-0.1854*	0.020
	int	0.3357*	0.000
	obl	0.0657	0.408
com	avo	0.4038*	0.000
	dom	0.2183*	0.006
	int	0.7394*	0.000
	obl	0.4695*	0.000
dom	avo	0.1854*	0.020
	com	-0.2183*	0.006
	int	0.5211*	0.000
	obl	0.2512*	0.002
int	avo	-0.3357*	0.000
	com	-0.7394*	0.000
	dom	-0.5211*	0.000
	obl	-0.2700*	0.001
obl	avo	-0.0657	0.408
	com	-0.4695*	0.000
	dom	-0.2512*	0.002
	int	0.2700*	0.001

Table 6.5: Post-hoc significance level between the number of first person pronouns and the ROCI-II prevalent style. Significant differences ($p<0.05$) are marked with asterisks.

ANOVA LSD post-hoc - Second Person			
I	J	Mean Difference (J-I)	Sig.
avo	com	0.0376	0.631
	dom	0.2606*	0.001
	int	0.8216*	0.000
	obl	-0.9343*	0.000
com	avo	-0.0376	0.631
	dom	0.2230*	0.004
	int	0.7840*	0.000
	obl	-0.9718*	0.000
dom	avo	-0.2606*	0.001
	com	-0.2230*	0.004
	int	0.5610*	0.000
	obl	-1.1948*	0.000
int	avo	-0.8216*	0.000
	com	-0.7840*	0.000
	dom	-0.5610*	0.000
	obl	-1.7559*	0.000
obl	avo	0.9343*	0.000
	com	0.9718*	0.000
	dom	1.1948*	0.000
	int	1.7559*	0.000

Table 6.6: Post-hoc significance level between the number of second person pronouns and the ROCI-II prevalent style. Significant differences ($p<0.05$) are marked with asterisks.

ANOVA LSD post-hoc - Third Person			
I	J	Mean Difference (J-I)	Sig.
avo	com	0.4038*	0.000
	dom	0.3803*	0.000
	int	0.7606*	0.000
	obl	-0.1667*	0.008
com	avo	-0.4038*	0.000
	dom	-0.0235	0.706
	int	0.3568*	0.000
	obl	-0.5704*	0.000
dom	avo	-0.3803*	0.000
	com	0.0235	0.706
	int	0.3803*	0.000
	obl	-0.5469*	0.000
int	avo	-0.7606*	0.000
	com	-0.3568*	0.000
	dom	-0.3803*	0.000
	obl	-0.9272*	0.000
obl	avo	0.1667*	0.008
	com	0.5704*	0.000
	dom	0.5469*	0.000
	int	0.9272*	0.000

Table 6.7: Post-hoc significance level between the number of third person pronouns and the ROCI-II prevalent style. Significant differences ($p<0.05$) are marked with asterisks.

ANOVA LSD post-hoc - Relational			
I	J	Mean Difference (J-I)	Sig.
avo	com	-0.1197*	0.016
	dom	-0.1995*	0.000
	int	-0.2770*	0.000
	obl	-0.1714*	0.001
com	avo	0.1197*	0.016
	dom	-0.0798	0.107
	int	-0.1573*	0.001
	obl	-0.0516	0.297
dom	avo	0.1995*	0.000
	com	0.0798	0.107
	int	-0.0775	0.117
	obl	0.0282	0.569
int	avo	0.2770*	0.000
	com	0.1573*	0.001
	dom	0.0775	0.117
	obl	0.1056*	0.033
obl	avo	0.1714*	0.001
	com	0.0516	0.297
	dom	-0.0282	0.569
	int	-0.1056*	0.033

Table 6.8: Post-hoc significance level between the number of relational pronouns and the ROCI-II prevalent style. Significant differences ($p<0.05$) are marked with asterisks.

each participant. For this reason, rather than using all the five dimensional scores of each style output from the ROCI-II classification and grouping sentences by this label, the analysis of the personal pronouns that follows, instead, takes into account the negotiation class of the user - i.e. the highest dimensional value - obtained in the ROCI-II scores. This categorization allows to deepen the investigation and divide the dataset into different documents that can be easily used for further natural language analysis. Since, according to the structure of the ROCI-II, the styles are not mutually exclusive, participants were able to be classified as having two or more predominant styles. These cases (n = 37) were excluded from the sample and therefore five documents were built. In order to mine the semantic differences between the documents, a series of analyses were conducted.

All of the sentences collected in the experiment have been re-labelled and classified according to the predominant style obtained by the user in the ROCI-II questionnaire. Considering only the highest dimensional value of the five negotiation styles obtained by each user, the distribution of the user's highest ROCI-II style value is summarized in Figure 6.10. The complete dataset consisted of documents distributed as specified in Table 6.9. It is important to notice the difference in the average number of words per sentence (average sentence length) in Table 6.9. In order to investigate the difference, a ANOVA with LSD post-hoc analysis was conducted on the distributions of the sentence lengths. Significant differences were found, and the summary of the results of the post-hoc analysis is shown in Figure 6.11. Participants classified as predominantly Avoiding, Dominating and Obliging produced significantly shorter sentences than Integrating ones, while sentences classified as Compromising did not significantly differ than the other styles' sentences. Avoiding, Dominating and Obliging sentence lengths all resulted with a very similar average length, as can be seen with the p level of 1.0 in the Table. This is again consistent with the hypothesis that while predominantly Compromising and Integrating styles will tend to integrate the perspectives of both the parties during the negotiation and consider multiple options, a highly Dominating party will try to reach his or her objective in the shortest time possible, as well as Obliging

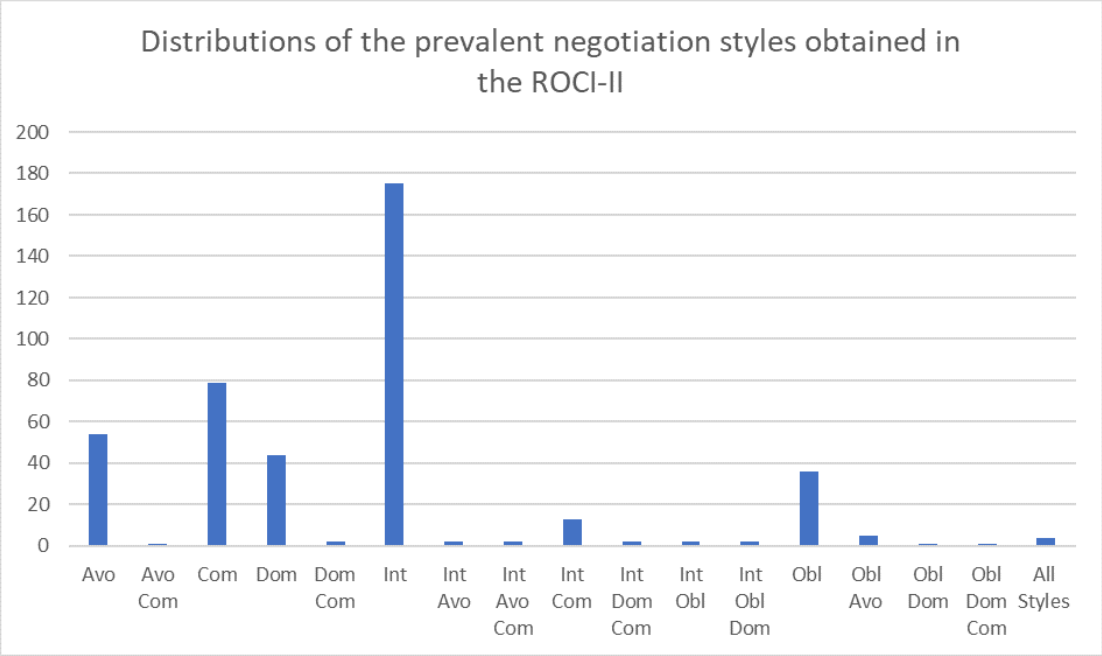

Figure 6.10: Distribution of the negotiation class obtained by the participants in the ROCI-II.

and Avoiding parties will not explore all the possible outcomes and will have a reduced concern towards the object of the negotiation and/or self or other's interest. This is also consistent, as previously explained, with the Integrating and Compromising styles being the negotiation styles that require the highest amount of time to reach a mutual decision.

Concerning the term frequency, an analysis of the occurrences of the words and pronouns was performed. All the results from this section were processed from stemmed

	Avo	Com	Dom	Int	Obl
N	540	790	440	1760	360
Tot words	4421	6978	3690	1761	2949
Tot words (no stopwords)	4146	6556	3466	15645	2771
Avg. sentence length (words)	8.19	8.83	8.39	9.34	8.19
Vocabulary	796	1074	780	1745	642

Table 6.9: Number, number of words, sentence length and vocabulary of each document

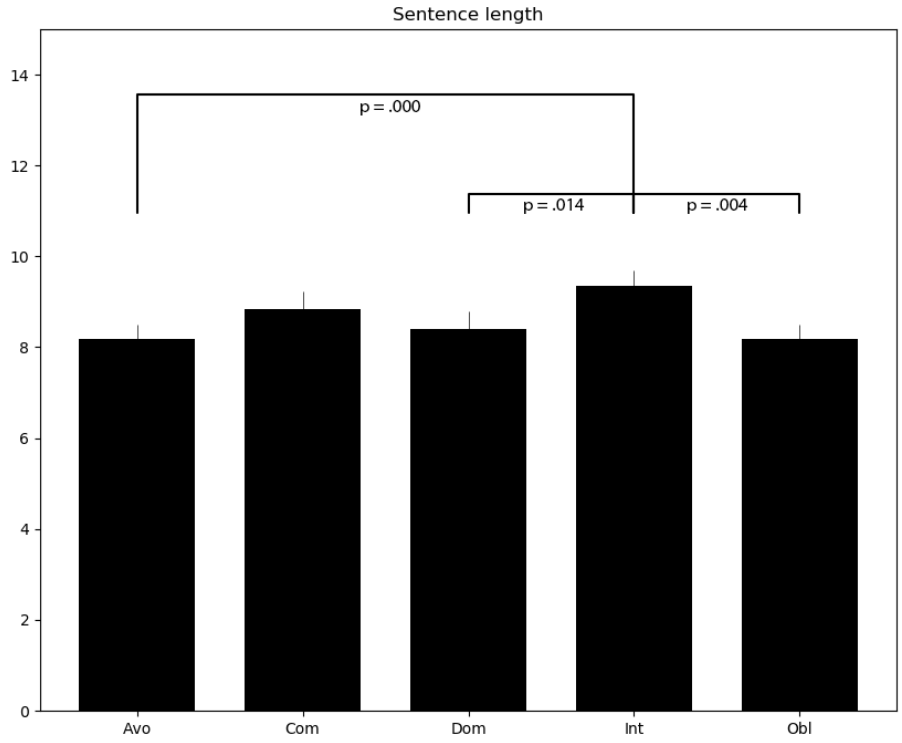

Figure 6.11: Mean, standard deviations and ANOVA post-hoc analysis of the sentence length among the documents. Significant results are marked and annotated on top with the p level.

	Most frequent pronouns (Abs. count)			**Most frequent pronouns (Normalized)**		
Avo	You (192)	It (152)	I (138)	You (0.068)	It (0.054)	I (0.049)
Com	You (314)	I (231)	It (200)	You (0.070)	I (0.051)	It (0.044)
Dom	You (193)	It (134)	I (120)	You (0.079)	It (0.055)	I (0.049)
Int	You (743)	I (535)	It (531)	You (0.070)	I (0.050)	It (0.050)
Obl	You (135)	I (100)	It (92)	You (0.071)	I (0.053)	It (0.049)

Table 6.10: Most frequent personal pronouns (absolute counts and normalized by the total number of words) that occurred in each of the documents

	Most frequent terms (Abs. count)			**Most frequent terms (Normalized)**		
Avo	Watch (60)	Take (48)	Go (38)	Watch (0.021)	Take (0.017)	Go (0.013)
Com	Watch (97)	Take (69)	Go (60)	Watch (0.022)	Take (0.015)	Go (0.013)
Dom	Watch (51)	Take (46)	Let (40)	Watch (0.021)	Take (0.019)	Let (0.016)
Int	Watch (230)	Take (212)	Let (150)	Watch (0.022)	Take (0.020)	Let (0.014)
Obl	Watch (43)	Let (40)	Take (29)	Watch (0.023)	Let (0.021)	Take (0.015)

Table 6.11: Most frequent terms (absolute counts and normalized by the total number of words) that occurred in each of the documents

tokens excluding the stopwords. Table 6.10 displays the most frequent words in their absolute counts that occurred in each of the documents and their relative occurrence regarding their proportion on the whole dataset (normalized from 0 to 1). Table 6.11 displays the most frequent personal pronouns and their absolute counts and normalized frequency.

As it can be seen, "You" was by far the most frequent pronoun in each document, while the occurrence of "It" and "I" is similar in proportion across the documents, not showing any significant difference. The most frequent terms, instead, were mainly verbs connected the specific scenario they were used in ("watch" in reference to the TV program Masterchef within the scenario in which the virtual agent was Dominating, in Appendix J, "take" in reference to the motorbike to share and the CD, scenarios in which the virtual agent was Compromising and Avoiding, in Appendix F and I). The differences among the documents, instead, is evidenced by the third most frequent term. Avoiding and Compromising, the negotiation styles that are less engaged and motivated to continue the negotiation, show a higher frequency of the verb "Go", that

	Context similarity "You"	Context similarity "I"
Avo	I, We, Together	You, Early, We
Com	We, I, Me	You, We, It
Dom	Could, Go, Take	You, Help, Pizza
Int	I, We, It	You, We, It
Obl	Me, That, I	Let, We, Upon

Table 6.12: Words that are used in similar contexts as the pronouns "You" and "I" across the documents measured with the Leacock-Chodorow algorithm.

is semantically related to a tendency to put off the discussion or take a immediate decision, while Dominating, Integrating and Obliging styles show "Let" as the third most frequent term, a word associated with a direct proposal and therefore a wish to continue negotiating.

Considering that the personal pronouns with the highest amount of occurrences, as shown in 6.10, were "You" and "I", the next step is trying to understand if their use was similar across documents. For this purpose, the context - defined as semantical concordance, i.e. the tokens that are more frequently associated and used before or after each word (Miller et al. 1994) - in which these personal pronouns were used in each document was measured with the Leacock-Chodorow algorithm (Leacock & Chodorow 1998). The pronoun "It" was not included since it is semantically ambiguous (Cegala 1989). The results of the analysis, conducted in order to spot the context concordance between the same pronouns in different documents, are presented in Table 6.12. While Avoiding, Compromising and Integrating display the use of "You" and "I" similarly with each other and with the pronoun "We", a peculiar difference is with the styles Obliging and Dominating. In the Dominating document, in fact "You" is used in context with imperative verbs or proposals ("Could", "Go" and "Take"), while "I" is associated with "Help" and, more importantly, with "Pizza", that is the object of negotiation the party would like to obtain in the scenario in which he or she deals with an Integrating agent (Appendix G), displaying a typically Dominating "I statement" (Chapter 4, Table 4.1).

Finally, Table 6.13 shows if there are shared similar contexts between the pronouns "I"

	Contexts shared by "I" and "You"
Avo	Know, Get, Good, Time
Com	Happy, Watch, Need, Yeah, Say
Dom	Don't, Take
Int	Want, Don't, Think, Could, Please
Obl	Yeah, Need

Table 6.13: Words that are used in shared similar contexts as the pronouns "You" and "I" across the documents.

and "You" across the documents. These results highlight that predominantly Avoiding participants again make references to a temporal dimension ("Time") while negotiating, aiming to postpone the discussion, as in

"Next time this comes up I'll give it to you."

while Obliging classified sentences (and largely less Compromising) shows "Yeah" as a shared context, which confirms the theory that sees it as accommodating the other parties' proposals, as in

"Anything to make you happy, yeah.".

In the Dominating document there are only two words that are shared between the contexts of "You" and "I", the most prevalent being "Don't", that clearly shows the attempt to impose one's will on the other party. Integrating, instead, consistently with the theoretical definition, shows also words that are semantically related to proposals and exploration, such as "Think", "Could" or "Please".

Finally, a natural language processing analysis of the dictionary and of the five documents containing the sentences in relation with the ROCI-II prevalent dimension (class) of each negotiation style of the participant was conducted. The frequency of the words, the width of the vocabulary, the similarity between the documents and the term frequency - inverse document frequency (TF-IDF) scores were calculated (Sparck Jones 1972). The TF-IDF can be measured for each word w in a document d with vocabulary V_d and weighs the frequency of that word in a document by the occurrence of it across

	Avo	Com	Dom	Int	Obl
Avo	1	0.9664	0.9541	0.973	0.9437
Com	0.9664	1	0.958	0.9799	0.9546
Dom	0.9541	0.958	1	0.9688	0.9472
Int	0.973	0.9799	0.9688	1	0.9613
Obl	0.9437	0.9546	0.9472	0.9613	1

Table 6.14: Cosine similarity values between each of the TF-IDF matrices generated from the documents

all documents D, as shown below:

$$TF = \frac{count(w \in d)}{V_d} \tag{6.1}$$

$$IDF = \log \frac{N_D}{count(d \in D : w \in d)} \tag{6.2}$$

$$TFIDF = TF * IDF \tag{6.3}$$

The TF-IDF value, performed on all words of each document, allows to obtain a vectorized form of each document, that can be used to measure the similarity with standard geometrical operations like cosine similarity. Cosine similarity can be used to find the width of the angle θ between the two vectors representing the documents (Salton & Buckley 1988), regardless of their length, and is defined as:

$$cos\theta = \frac{\vec{d_1} \cdot \vec{d_2}}{|d_1||d_2|} \tag{6.4}$$

where d_1 and d_2 are the TF-IDF values of two documents.

TF-IDF was used to vectorize each term in the vocabulary of the documents. TF-IDF matrices were computed for all the terms in each of the five documents and were compared using cosine similarity. Results of the computations are shown in Table 6.14. As it can be seen, the highest similarity values stand between the document containing the sentences classified as Integrating and all the other documents. This can be partly explained by the fact that the Integrating style is the one that tries to include all the perspectives and holds both a high interest for self and for other. Also, this

could be a consequence of the fact that, as previously evidenced, sentence length was higher, and so was the number of words in the vocabulary. The couple with the highest similarity stands, confirming the initial hypothesis, between the documents containing Compromising and Integrating sentences - and this is in accordance with the theoretical definitions of these styles - while, still consistently with Rahim's theory, the couple of documents classified as predominantly Dominating and Obliging holds the lowest score in terms of cosine similarity, since these styles are placed opposite in the concern for self's and concern for other's conflict handling spectrums.

6.6 Conclusion

A natural language analysis of the pronouns, word frequency and term usage was conducted to investigate the interaction between the negotiation style and the user natural language when interacting with virtual agents in the Enact game. The sentences used by participants in the interactions with Enact game scenarios were collected and labelled with the ROCI-II dimension scores representing the negotiation styles. An analysis of the frequency and usage of the first person, second person, third person and relational pronouns was performed on the dataset, taking into account the style of the virtual agent the user negotiated with. In general, results indicated that the Integrating dimension has been found to correlate with a higher number of all pronouns, and a significantly higher number of relational pronouns. The Dominating component has been found to correlate with the use of more second person pronouns and significantly less relational pronouns. Highly Avoiding and Obliging individuals were those that used the lowest amount of pronouns where Avoiding ones showed a higher amount of third person pronouns. The values in the pronouns usage of individuals with a high Compromising dimension are very close to the average of all the other styles across all the pronouns. Since these results were consistent with the theoretical definitions of Rahim's negotiation styles, and in particular, Integrating style sought to engage the other party using personal pronouns with a predominance of relational once, with an opposite pattern used by highly Dominating individuals and with styles less involved

in the negotiation as Avoiding and Obliging using less relational and more third person pronouns, a corpus containing the five documents of sentences labelled with the prevalent ROCI-II dimension obtained by the participant was analysed applying text mining techniques and differences among the sentence length, vocabulary and words usage in the styles consistently with the theoretical framework were found.

Chapter 7

Machine learning for multi-class text classification of negotiation styles

In literature, technology-based methodologies for psychological assessment, such as serious games, MOOCs and e-learning platforms, have mostly based the user model on multiple choice responses, reaction times, eye-tracking or other physiological responses with low emphasis on speech and natural language. While few conversational diagnostic tools were able to record natural language, this data has been mostly analysed and included in the user model using rule-based or probabilistic models, overlooking machine learning methods that represent the current trend in NLP. This chapter will present a possible machine learning approach to computational psychometrics, psychological user modelling and specifically for automated multi-class classification of the user negotiation skills based on documents. The dataset of negotiations utterances between players and virtual agents, collected through the game scenarios of the Enact platform, was used to train a vocabulary of word embeddings and fed into state-of-the-art machine learning classifiers (MNB and SVM) and deep learning networks (CNN) to predict the negotiation style of the users as measured by the psychological test ROCI-II. Results show that the ROCI-II five-factor negotiation score can be predicted reliably by the models considered and, in particular, the use of pre-trained word embeddings increased the accuracy compared to one-hot vectors.

THE last phase of the present work concerns the proposal of a new methodology for a possible enhancement of the Enact platform in the attempt to integrate a system for the automatic classification of the user negotiation style based on natural language texts. The dataset of natural utterances occurred in the interactions between the Enact users and the virtual agents in the scenarios, paired with the ROCI-II negotiation scores of each user, constitutes in fact a supervised training dataset that was used to fit several classification models to help reliably profiling the users' negotiation style. The idea behind modelling an automatic text classification and prediction is twofold: on

one side, the aim is to map the ROCI-II negotiation style scores to full sentences, with a broader width than multiple choice answers, and extract possible non-linear features that pertain to all aspects of language, improving the accuracy of the user model; on the other, the aim is to implement shorter virtual interactions in the scenarios, allowing natural language as a mean of communication with the platform, and to provide more targeted user choices, with the future possibility of sentence generation from the agent side.

While previous studies on communication engagement - such as (Cegala 1989), as well as previous analysis conducted on the dataset (see Chapter 6) - have identified the key role of personal pronouns for predicting the level of engagement during discussions and negotiations, in association with each of the considered negotiation styles of Rahim and Bonoma's theory, the contributions of the modelling approach presented in this section lie in four main points:

- previous natural language studies have overlooked the importance of using standardized psychometric ways to classify or assess their human data, while in this study the ROCI-II scores for each negotiation style has been collected and used for training supervised algorithms;
- negotiating with participants and collecting natural language sentences was conducted in an automated process and with the use of virtual agents rather than humans; this procedure brings several advantages, among which the possibility to have a consistent methodology virtually free of the biases connected to the interaction of participants with human experimenters/other users;
- the collected dataset is currently the biggest one in literature that concerns natural language negotiations currently available and it is potentially expansible using the same automated procedures;
- using machine learning on the whole dataset rather than analysing only the personal pronouns or other specific parts of speech allows to extract and include

in one classification all the possible non-linear features that can be relevant to the categorization, such as the temporal aspects of the word sequences, word synonyms, broader word contexts and sentence length and to convert identified patterns into possible generative algorithms.

7.1 Natural language processing methodologies for psychological assessment

7.1.1 Non-linear classification and artificial intelligence in computational psychometrics

In literature, it is not common to find machine learning methods applied to the classification or prediction of text to psychological variables, especially if collected without human experimenters, using a technology developed for this specific purpose. The first attempts to use machine learning algorithms and artificial intelligence for psychological and cognitive assessment in the field of psychometrics and neuroscience date back to the late '80s (Garfield et al. 1992), shortly after the first practical applications of the multilayer perceptron (MLP) and around the last years of cybernetics. However, due to the high human cost of developing software on purpose for data collection, or manually process huge amount of self-reported questionnaires, this approach has not been widely adopted in quantitative psychology and for automated diagnosis until the more recent rise of deep learning, virtual reality and simulative technologies applied to clinical settings (Cipresso & Immekus 2017, Gamito et al. 2017).

Among the most recent studies that tried to map human behaviours to traditional psychological assessment, Maldonato et al. (2018) employed a shallow Artificial Neural Network (ANN) to predict borderline personality disorder, measured via SCID-II (a structured interview with binary answers) using the scores of the DES (Dissociative Experiences Scale) and the TCI (Temperament and Character Inventory), while Polyak et al. (2017) collected user interactions with virtual agents in a scenario-based serious game to investigate collaborative skills, although without using natural language. Concerning natural language analysis, psycholinguistics and computational linguistics

studies have rarely tried to integrate machine learning for natural language understanding with traditional psychometric methods. The MOST system, a recently developed "therapeutic" social network, includes chatbots and automated suggestions based on the user behaviour, but does not include psychological assessment in the user model (D'Alfonso et al. 2017). Park et al. (2015), instead, tried to predict personality factors measured with the Big Five inventory analysing the words of the participants' Facebook posts and used information retrieval (IR) methods such as Linguistic Inquiry and Word Count (LIWC), linear regression and Principal Component Analysis (PCA) for the analysis, finding a robust reliability.

7.1.2 State-of-the-art machine learning for multi-class text classification

The design of the Natural Language Processing architecture is thought to research the algorithm that can provide the best feature mapping between the numerical representation of the sentences and either the prevalent style measured by the ROCI-II or the five scores for each of the negotiation style. In order to achieve this task, a comparison among several computational methodologies will be provided.

Text classification using supervised methods is generally deemed as a complex categorization problem that has no unique solution, in that set labelling itself can be an ambiguous task even for humans (Sebastiani 2001). This is in contrast with image classification problems, which can reach even 100% accuracy on the training set on specific datasets (e.g. Marmanis et al. (2016)). Multi-class text classification can be defined as the task to find the function Φ that approximates the relationship between each document $d_i \in D$ and a category or class $c_j \in C$, where $D = \{d_1, d_2, d_3 ... d_n\}$ and $C = \{c_1, c_2, c_3 ... c_m\}$. In this case, the documents are the user sentences, while the classes are the five negotiation styles. The first approaches in machine learning that tried to model this relation are based on linear classifiers like Support Vector Machines (SVM) (Joachims 1998) or the probabilistic Multinomial Naive Bayes (MNB) (Kibriya et al. 2004). The current state-of-the-art deep learning approaches to natural language processing and multi-class classification propose the use of recurrent neural networks

(RNNs) Elman type (Mikolov et al. 2011), RNNs with long short-term memory (LSTM) (Graves 2013, Mirowski & Vlachos 2015) and more recently CNNs (Johnson & Zhang 2014, Zhang et al. 2015). In the next sections, a comparison of SVM, MNB and CNNs and their performance on the dataset is conducted.

7.2 Materials and Methods

The dataset used for the analysis is the same collected with the methods explained in 6.2. Considering the novelty of the dataset and its limited size (4250 sentences), it is not meaningful to use a comparison with benchmark datasets in terms of accuracy. Instead, a comparison will be performed between the performances of a linear classifier (SVM), a probabilistic classifier (MNB) and deep neural networks (CNNs). The soft-margin SVM is an algorithm widely applied to text categorization and is particularly useful when the problem is not entirely linearly separable. As it was originally conceived to be applied to binary categorization, the one-against-one model allows to apply the SVM to multiclass classification, as will be explained in detail in Section 7.2.2.1. The Multinomial Naive Bayes, instead, is another widely used technique that uses the Bayes theorem to estimate the probability of a text to fall within one of the classes provided, as shown in detail in Section 7.2.2.2. Finally, Convolutional Neural Networks, a kind of deep neural network composed of multiple layers and filters generally trained with backpropagation and whose structure is detailed in Section 7.2.2.3, was also tested on the dataset.

7.2.1 Word representations and word embeddings

7.2.1.1 Skip-gram and Continuous bag of words model

In literature, given a vocabulary V the most frequent way used to get a numeric vector $v_w \forall w \in V$ composed by all the words used in the collection of documents D was to represent them as unidimensional binary vectors (one-hot vectors) of length V in which all values are zeroes except for the index in the vector representing that word in the vocabulary. However, this raw encoding is not frequently used anymore as a

direct input of classifiers whereas other encoding that are more dense and include in their representation other features such as context, temporal occurrence and frequency are the preferred solution. To transform the words into their numeric representation, a modified version of the Google software Word2vec (https://code.google.com/p/word2vec/) was used. Word2Vec is an open source project which represents an implementation of the skip-gram model and the continuous bag-of-words model (CBOW) (Mikolov, Chen, Corrado & Dean 2013) and includes two different training algorithms, hierarchical softmax and negative sampling (Rong 2014, Goldberg & Levy 2014). Both hierarchical softmax and negative sampling algorithms represent an optimization in terms of speed of the classical softmax, whose computation on each word of the dataset at the output layer of the Word2Vec network would be overly long. Hierarchical softmax makes use of binary trees, while negative sampling uses noise estimation methods to compute the probability of a pair of words being in the dataset and, where extremely low, they discard it from the model. The negative sampling algorithm, used in the present work, is explained in section 7.2.1.2. Both the skip-gram model and the CBOW model allow to map any collection of words and/or documents into vectors and take into account the similarity and proximity of words. The proximity of words is defined by their context, that is a specific window of words before and after the target word. In particular, the skip-gram model, for each word in the document and each context in which the word occurs, tries to maximize the log probability of the occurrence of the current context ω given the word w and the model parameters θ, as shown below:

$$\underset{\theta}{\operatorname{argmax}} \sum_{w \in D} [\sum_{\omega \in \omega_w} \log p(\omega \mid w; \theta)] \tag{7.1}$$

The CBOW model builds the embeddings of the vectors by instead maximizing the opposite, i.e. the probability of the target word w (in the output layer) given the context ω (in the input):

$$\underset{\theta}{\operatorname{argmax}} \sum_{w \in D} [\sum_{\omega \in \omega_w} \log p(w \mid \omega; \theta)] \tag{7.2}$$

7.2.1.2 Negative sampling

Modelling the log probability $\log p(w \mid \omega;\theta)$ part of the equation 7.2 of the previous section can be done with a training algorithm that represents the loss function of a shallow neural network whose weights are trained with a standard stochastic gradient descent (SGD) and a specific learning rate. The log probability, i.e. the logarithm of the probability, is used because it usually scales better with all machine learning gradient methods as the one used in Word2Vec. However, while the original formulation of the skip-gram model uses softmax as training function, its two approximations hierarchical softmax and negative sampling were proven to be both more cost efficient and as accurate (Mikolov, Sutskever, Chen, Corrado & Dean 2013). In particular, the Word2Vec embeddings used for the present experiment are taken from the pre-trained vectors with a dimensionality of 300 provided by Google and trained using negative sampling on Google News containing a vocabulary size of over 3 billion words. Negative sampling allows to reduce the dimensionality of the vectors by discarding from the computation all the pairs (w, ω) that are less likely to be in the training set. Defining the presence of a pair in the dataset as $(w, \omega \in D) = 1$ and the probability that a pair is in the dataset as $p(1 \mid w, \omega, \theta)$ then the probability that a pair is not in the dataset will be the inverse $1 - p(1 \mid w, \omega, \theta)$. In order to model this probability, a new set of documents D' is generated containing k random pairs which are not in the training set for each pair which is in the dataset. The new function to optimize now becomes:

$$\underset{\theta}{\operatorname{argmax}} \sum_{w,c \in D} \log p(1 \mid w, \omega, \theta) + \sum_{w,c \in D'} \log(1 - p(1 \mid w, \omega, \theta)) \tag{7.3}$$

Using softmax, the probability can be redefined as:

$$p(1 \mid w, \omega, \theta) = \frac{1}{1 - e^{-v_\omega \cdot v_w}} \tag{7.4}$$

And the loss function can be written, as derived in (Goldberg & Levy 2014), as:

$$\underset{\theta}{\operatorname{argmax}} \sum_{w,c \in D} \log \frac{1}{1 - e^{-v_\omega \cdot v_w}} + \sum_{w,c \in D'} \log(1 - \frac{1}{1 - e^{-v_\omega \cdot v_w}}) \tag{7.5}$$

The words of the documents used for training the skip-gram model are represented by one-hot vectors. Using this training model, the results obtained show that similar vectors represent similar words, and semantic relationships between the words have been shown to emerge using just basic arithmetic operations on the vectors, like sum and element-by-element multiplication, and verified with cosine similarity (Demeester et al. 2016), obtaining an accuracy between 60% and 65%.

7.2.1.3 The GloVe algorithm

The GloVe algorithm describes another method to model word embeddings that reached an accuracy as high as negative sampling and a higher performance on word similarity tasks computed with arithmetic vector operations and is described in (Pennington et al. 2014). Rather than trying to model the probability of the presence of each word in the context, the model is based on the idea of modelling the probability of the co-occurrence of the words in the context. In particular, the function F of the co-occurrence of the words w_i and w_j in a context and of the other words w_ω in the context can be defined as:

$$F(w_i, w_j, w_\omega) = \frac{p_{ij}}{p_{j\omega}} \tag{7.6}$$

and extracted from the corpus. Given a constraint to the function so that the probability is only dependent on the difference between w_i and w_i, and given that the probability is a scalar while the terms of F are scalars, the multiplication can be expressed as a dot product:

$$F((w_i - w_j) \cdot w_\omega) = \frac{p_{ij}}{p_{j\omega}} \tag{7.7}$$

Given that all words in the context can be exchangeable in the equation, then the function F must be symmetric. The authors choose it to be a log and define it as:

$$w_i \cdot w_j + b_i + b_j = log(X_{ij}) \tag{7.8}$$

where b is a bias for the main and context word and X_{ij} is the co-occurrence matrix of the two words w_i and w_j. For all considered words in the vocabulary V, then the function to be optimized can be rewritten as:

$$\sum_{i,j=1}^{V} f(X_{ij})(w_i \cdot w_j + b_i + b_j - log(X_{ij}))^2 \tag{7.9}$$

where $f(X_{ij})$ is a non-decreasing weighing function for the co-occurrences that authors choose as:

$$f(n) = (x/x_{max})^{\alpha} \quad if \quad x < x_{max}, \quad 1 \ \ otherwise \tag{7.10}$$

7.2.2 Classification training algorithms

Once the whole set of vectors in the vocabulary is created from the training set, the vectors can be used as input for the different architectures. In particular in the present work the input consists of the trained vectors, and the output is the prevalent ROCI-II style obtained by the participant. While the SVM and MNB classifiers take as training examples simple labelled pairs $(d,c)\forall d \in D$ presenting therefore all words in each document at once, the neural networks' training examples consists of a n-gram, that is word and its context. In particular, for training the CNN the context ω is defined by tri-grams (triplets) of words.

7.2.2.1 Soft margin Support Vector Machine

A Support Vector Machine is a classification algorithm that finds the linear separator, constituted by a set of hyperplanes, between all data points of two classes. This separator will be defined by a hyperplane and its two margins - two other hyperplanes - whose boundaries are defined by the so-called support vectors. Considering that a

hyperplane can be defined as:

$$v \cdot x - b = 0 \tag{7.11}$$

where v is the normal vector to the hyperplane, b is the intercept and x are all the data points belonging to the hyperplane. Considering the binary classes c_1 as -1 and c_2 as +1 then the two hyperplanes linearly separating the data are

$$\vec{d} \cdot w - b = 1 \tag{7.12}$$

and

$$\vec{v} \cdot d - b = -1 \tag{7.13}$$

where d is the document, i.e. the collection of word vectors for each sentence in the training example, with the hard constraints that

$$\vec{v} \cdot d_i - b \geq 1, \; if \; c_i = 1 \tag{7.14}$$

$$\vec{v} \cdot d_i - b \leq -1, \; if \; c_i = -1 \tag{7.15}$$

if all the points in the dataset fall outside of the margins. The bigger the distance between the hyperplanes, the more accurate the classification is. Considering that the geometric distance between these two hyperplanes is, using the distance from a point to a plane, defined as $\frac{2}{\|\vec{v}\|}$, maximizing the distance is the same as minimizing $\|\vec{v}\|$. Considering that in most real cases it is not possible to perfectly separate the two classes, constraints are softened adding a penalty function, that is a measure of how much the constraints are violated and it is defined as:

$$k \sum_{i=1}^{n} \xi_i \tag{7.16}$$

where ξ is how much each outlier data point d is distant from the margin and k is the penalty parameter, that controls the trade-off between this distance and the cost on the classification error. Putting these terms together, the linear primal formulation is then

defined as the following optimization problem:

$$\min_{v,\xi} \|v^2\| + k\sum_{i=1}^{n} \xi_i \tag{7.17}$$

with the constraints

$$\vec{v}\cdot d_i - b \geq 1,\ if\ c_i = 1 - \xi \tag{7.18}$$

$$\vec{v}\cdot d_i - b \leq -1,\ if\ c_i = -1 + \xi \tag{7.19}$$

In order to perform multi-class text classification the optimization problem must be adapted to support more than 2 classes. There are several approaches to this problem, of which the most performing for text classification has been shown to be the one-against-one model (Knerr et al. 1990, Hsu & Lin 2002, Colas & Brazdil 2006). The model has this name because is based on the idea of creating $m(m-1)/2$ classifiers where m is the length of total categories C each trained to classify a pair of classes. Given each pair of classes, denoted as c_i, c_j the formulation becomes:

$$\min_{v_{i,j},\xi_{i,j}} \|v_{i,j}^2\| + k\sum_{l=1}^{n_{i,j}} \xi_l \tag{7.20}$$

with the constraints

$$\vec{v_{i,j}}\cdot w_l - b_{i,j} \geq 1,\ if\ c_l = 1 - \xi_{i,j} \tag{7.21}$$

$$\vec{v_{i,j}}\cdot w_l - b_{i,j} \leq -1,\ if\ c_l = -1 + \xi_{i,j} \tag{7.22}$$

Then the class that is selected by most classifiers is chosen.

7.2.2.2 Multinomial Naive Bayes

The Multinomial Naive Bayes is a variant of the Naive Bayes classifier for multi-class classification and estimates the probability that a document d falls within the class c, i.e. $p(c \mid d)$ (Kibriya et al. 2004). For the Bayes rule it can be defined as:

$$p(c \mid d) = \frac{p(d \mid c)p(c)}{p(d)} \tag{7.23}$$

Considering that the probability of a document $p(d)$ is constant and the same across all classes, and that, assuming independence among the words in the document, $p(d \mid c)$ can be defined as $p(w_1, w_2, w_3 ... w_n \mid c)$, the optimization problem can be defined as:

$$\underset{c \in C}{\operatorname{argmax}}\, p(c) \prod_{w \in d} p(w \mid c) \tag{7.24}$$

in which the probability of each class c_j is defined as:

$$p(c_j) = \frac{count(d \in c_j)}{count(d \in D)} \tag{7.25}$$

and the probability of a word w_i given a class c_j as:

$$p(w_i \mid c_j) = \frac{count(w_i \in c_j)}{\sum_{w \in V} count(w \in c_j)} \tag{7.26}$$

7.2.2.3 Convolutional neural networks

Following the good performances on image classification and segmentation, CNNs have been applied to text classification with excellent results (Kim 2014, Zhang et al. 2015). The architecture of the network is structured with a fully connected convolution layer, a max-pooling layer and a softmax output layer. Given a document d (in this case a sentence) and its words $x_{1:n}$, a convolution layer is defined as a set of f feature extractors (filters) composed by unique trainable random weights W fully connected and multiplied to $c_{1:p}$ windows of k words that slides across the sentence with a stride size δ. In particular, for one feature f_i generated from a window of words $x_{i:k}$:

$$f_i = sigmoid(W \cdot x_k + b) \tag{7.27}$$

where b is a bias unit. Then, for each set of features belonging to a specific window size, a max-pooling layer follows taking the maximum value among the outputs such

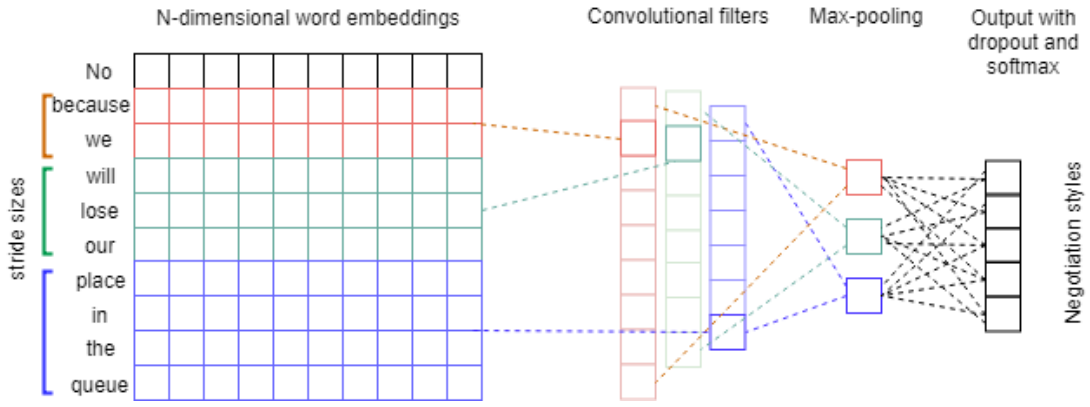

Figure 7.1: Diagram of the Convolutional neural network used for the negotiation style classification

as:

$$F = \underset{f}{\operatorname{argmax}}(f_{i:p}) \tag{7.28}$$

All the layers are fully connected and trained with backpropagation using SGD. The final layer is constituted by a softmax output and its connections from the weights W and the max-pooling outputs $F_{i:c}$ to the outputs y are controlled by a dropout parameter r to prevent co-adaptation within the units as follows:

$$y = W \cdot (F_{i:c} \odot r) + b \tag{7.29}$$

with $\odot$ being the elementwise multiplication. The structure of the full CNN used for the classification is presented in Figure 7.1

7.3 Results

7.3.1 Datasets and training methodologies

As a baseline, two different datasets were used. The first dataset contained the whole dataset of collected sentences grouped by the negotiation style class (i.e. the prevalent ROCI-II score) of the user who provided them. The information about the scenario in which the sentence was written was excluded from the classification, in order to generate a more situation-invariant model. As explained in Chapter 6, the classes

are extremely imbalanced. In order to account for this problem and prevent the models from overfitting on one of the classes, the sentences from each of the lower represented classes were repeated entirely to match as much as possible the length of the most represented class. Among all the class balancing approaches used in literature, this has been found the most effective in a very recent comparison (Buda et al. 2017). From this dataset, which will be referred to as Full Dataset, all the sentences provided by users who scored on their predominant style at least .4 points above any other style score were further extracted. This restricted dataset will be called Rest Dataset. The Rest Dataset, after class balancing, consisted of five documents of similar length, in particular:

- n = 230 Avoiding sentences (total words count = 1670, average sentence length = 16.40)
- n = 210 Compromising sentences (total words count = 1885, average sentence length = 17.61)
- n = 220 Dominating sentences (total words count = 1665, average sentence length = 16.33)
- n = 230 Integrating sentences (total words count = 1703, average sentence length = 17.42)
- n = 240 Obliging sentences (total words count = 1994, average sentence length = 18.94)

In all cases training and test split was performed using 10-fold cross validation.

7.3.1.1 Training pipeline for SVM

The soft margin SVM was trained on both the Full and the Rest datasets. Since all word vectors in each sentence need to be presented at once to the SVM, each matrix of word vectors was condensed into one vector of the same length of the word embedding. This was done with two different approaches:

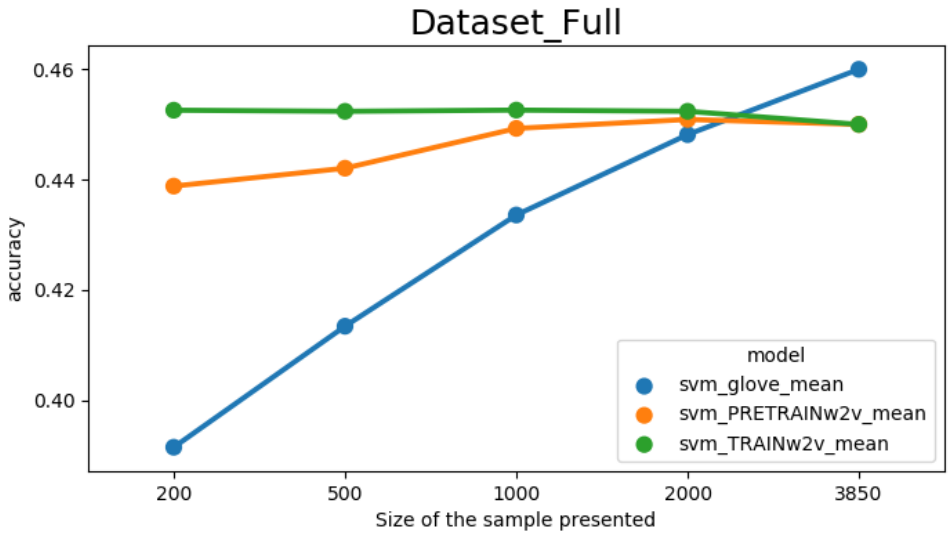

Figure 7.2: Results of SVM training on the Full Dataset using the "mean" model

- by averaging the values of the word vectors in a sentence (referred to as "mean" model);
- by weighting the sum of all word vectors in the sentence by their TF-IDF value (referred to as "tf-idf" model).

Also, for each dataset, the SVM was trained using one of three different word representations (Pacella, Dell'Aquila, Marocco & Furnell 2017):

- Word2Vec pre-trained 300-dimensional embeddings;
- Word2Vec 300-dimensional embeddings trained on the dataset dictionary;
- GloVe pre-trained 300-dimensional embeddings.

If a word was not found in the GloVe or Word2Vec pre-trained vocabulary, it was added to the model as a vector of numbers drawn randomly from a uniform distribution between 0 and 1, having the same length as the other word embeddings. The results of training the model on the Full Dataset are shown in Figure 7.2, that shows the training performance on the "mean" model and in Figure 7.3, that shows the performance of

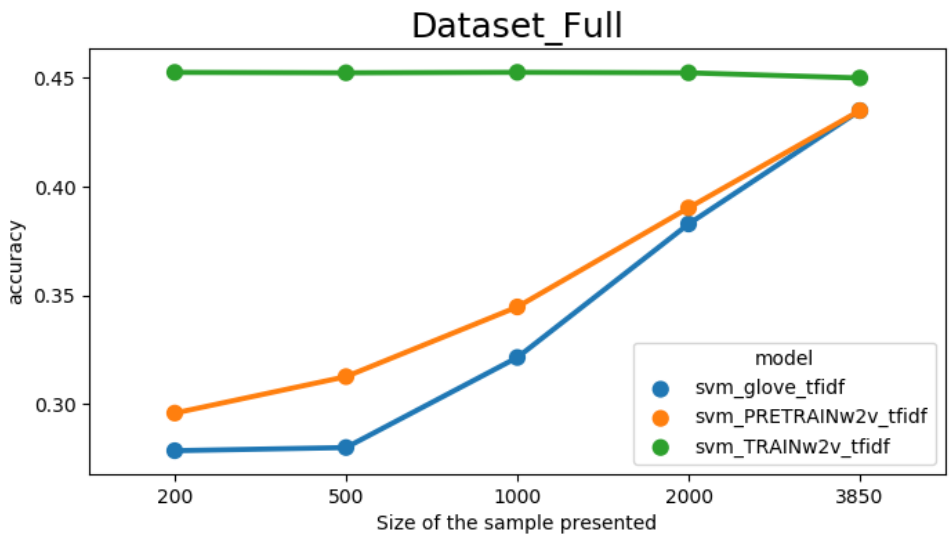

Figure 7.3: Results of SVM training on the Full Dataset using the "tf-idf" model

the "tf-idf" model. The average accuracy on the test set on all classes is summarized in Table 7.1.

Regarding the difference between the word embeddings, it is evident that the Word2Vec model trained on the documents' dictionary, even if it has a very high training accuracy, becomes saturated and overfits the dataset already at the beginning of the training presentations of the sentences (at 100 examples). This is consistent with the idea that the SVM, in fact, benefits from large and sparse matrices as input, and a smaller feature space results in a lower accuracy, as shown particularly in the test set table. Between the GloVe and the W2V pretrained embeddings, the GloVe vectors seem to retain a very high accuracy on the test set, while the W2V ones, even if their accuracy shows an increase during training similar to the GloVe model, have a much lower accuracy on the test set for both the "mean" and the "tf-idf" model. Comparing the "mean" and the "tf-idf" models, the training accuracy of the two models have very small differences across all the embeddings with the "mean" model being slightly above. On the test set it seems that using the "mean" model holds a quite higher score in all cases. This could be due to the fact that word embeddings already have a trained representation of the geometric relationship between vectors and the tf-idf weighting distorts this space

Full Dataset

Accuracy	**Glove**	**W2V pre-train**	**W2V train**
"mean" model	0.424	0.365	0.344
"tf-idf" model	0.393	0.349	0.322

Table 7.1: Accuracy of the SVM models on the test set of the Full Dataset

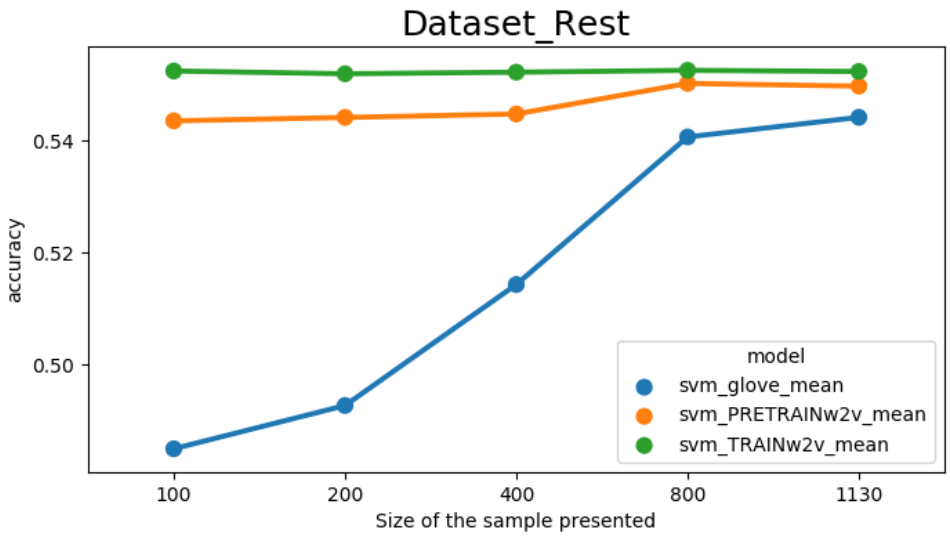

Figure 7.4: Results of SVM training on the Rest Dataset using the "mean" model

losing consistent information.

Concerning the Rest Dataset, the results are shown in Figure 7.4 for the "mean" model and in Figure 7.5 for the "tf-idf" model. The average accuracy is shown in 7.2. In general, the training and test accuracy on this dataset is much higher, reaching a maximum of 0.48 in the case of the GloVe vectors using the "mean" model. However, the differences in the models' accuracy hold the same relations that were evidenced in the case of the Full Dataset: the W2V embeddings trained on the dataset overfit in both the "mean" and the "tf-idf" model, with the GloVe vectors having a slightly lower accuracy during training but a much higher accuracy on the test set. Again, all the models scored less with the "tf-idf" model with the GloVe model and Word2Vec trained on the dataset scoring 6% less and Word2Vec pretrained scoring roughly 8% less.

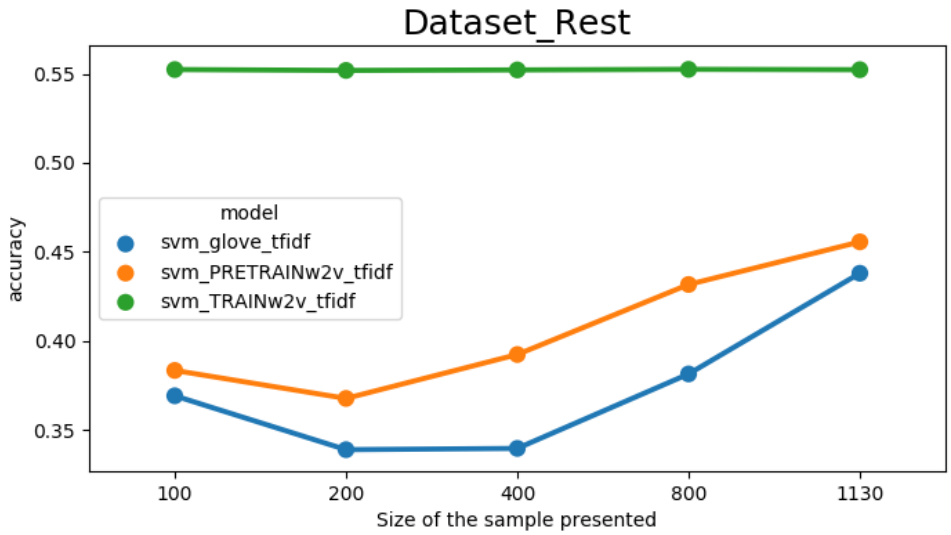

Figure 7.5: Results of SVM training on the Rest Dataset using the "tf-idf" model

Rest Dataset			
Accuracy	**Glove**	**W2V pre-train**	**W2V train**
"mean" model	0.481	0.472	0.457
"tf-idf" model	0.423	0.395	0.392

Table 7.2: Accuracy of the SVM models on the test set of the Rest Dataset

7.3.1.2 Training pipeline for MNB

As explained above (7.2.2.2), the MNB is based on the assumption of independence among the words in the document. For this reason, it is not meaningful to make use of any form of denser vectorization that is based on semantic word representation. However, for a fair comparison, using the one-hot vectors for the words encoding was measured against the same one-hot vectors weighted by their tf-idf value. The results of the accuracy on the training set for the Full Dataset are shown in Figure 7.6, while the results for the Rest Dataset are shown in Figure 7.7. The accuracy on the test set is summarized in Table 7.3. Concerning the word models, the tf-idf features did not

Accuracy	**Full Dataset**	**Rest Dataset**
one-hot vectors	0.401	0.477
one-hot * tf-idf	0.363	0.397

Table 7.3: Accuracy of the MNB models on the test set of the Full and Rest Dataset

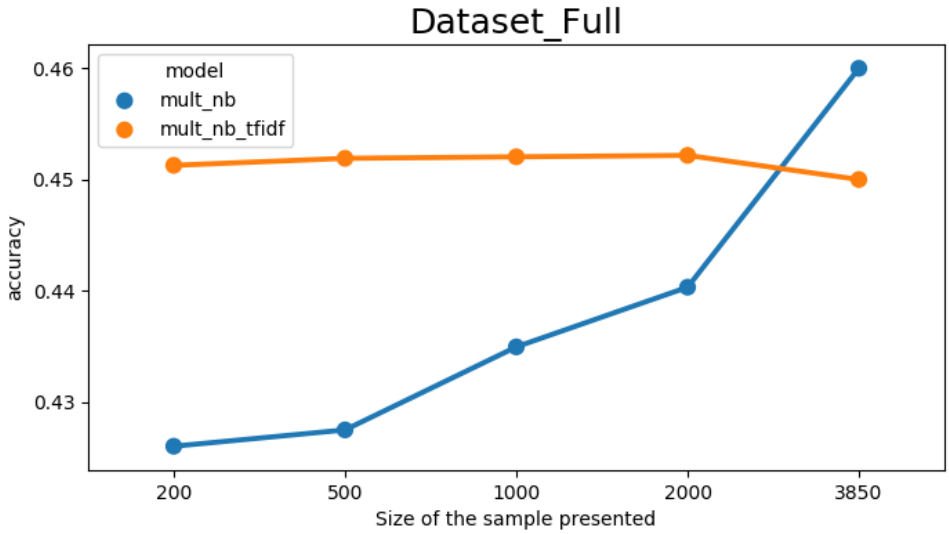

Figure 7.6: Results of MNB training on the Full Dataset

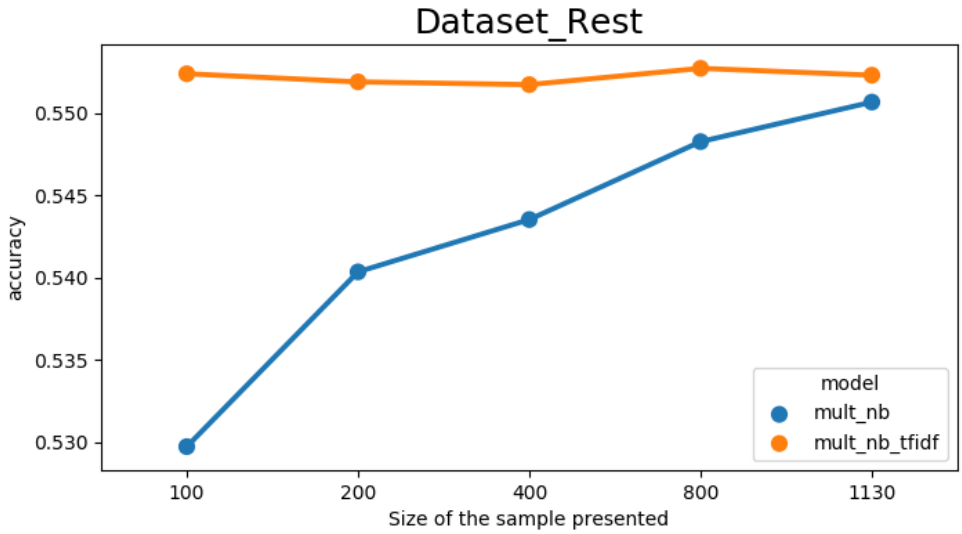

Figure 7.7: Results of MNB training on the Rest Dataset

improve the accuracy during training, and even led to overfitting on the Rest Dataset as can be seen by the fact that the training accuracy was high at only 100 examples presented and does not reflect the test set accuracy. Consistently, the accuracy on the test set is much higher for simple one-hot vectors, that are completely independent from each other. Also in this case, like for the SVM, the Rest Dataset shows a much higher accuracy.

7.3.1.3 Training pipeline for CNN

Since the CNN is the deepest model taken into consideration, and given the size of the datasets, the network was only trained on the Full Dataset, as the Rest Dataset would have been too small. The CNN was tested using as starting parameters those set in Kim (2014), Zhang et al. (2015), in particular:

- number of training epochs: early stopping.
- optimizer: Adam, with .0001 learning rate
- stride window: 1

while the parameters that were manipulated according to the dataset were:

- batch size;
- stride sizes;
- dropout probability during training;
- number of convolution filters;

The word embeddings that were used for encoding the sentences were the same as those used for the SVM, respectively the GloVe pre-trained vectors, the W2V pre-trained vectors and W2V vectors trained on the documents' dictionary. The parameters found to achieve the highest training and test results on the dataset with all types of word embeddings were the following: batch size = 20, stride sizes = (1,2,3), dropout

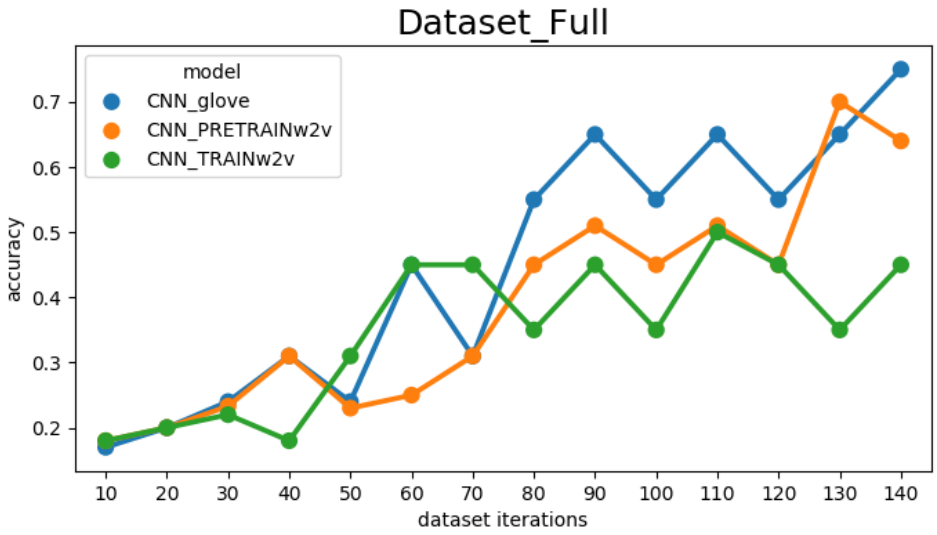

Figure 7.8: Results of CNN trained on the Full Dataset.

	Glove	**W2V pre-train**	**W2V train**
Accuracy	0.38	0.371	0.342

Table 7.4: Accuracy of the CNN models on the test set of the Full Dataset

probability = 0.3, number of convolution filters = 256. Figure 7.8 shows the training learning accuracy of the CNN across dataset iterations, while Table 7.4 shows the performances results on the test set. In general, it is possible to notice that despite the dropout layer the network tends to greatly overfit. In fact, in all cases, the training accuracy is much higher than the one on the test set. However, the performance of this deep model is still better with the GloVe word vectors rather than with the other two W2V vectors, consistently with the other models.

7.4 Discussion

Before discussing the differences among the performance of the models themselves, it is important to underline that in all machine learning models considered and all word vectors used the accuracy showed to be much higher on the Rest Dataset than on the Full Dataset. Considering that the only criterion by which the sentences for the Rest Dataset were selected is the difference between the predominant negotiation style dimension and the other negotiation scores, this performance difference confirms that

Full_Dataset			
	SVM_Glove_mean	**MNB_Glove**	**CNN_Glove**
Accuracy	*0.424*	0.401	0.38

Table 7.5: Test accuracy of the best performing models on the Full Dataset

there are unique and evident language features that emerge the more predominant each of the negotiation profile becomes.

Concerning the models, the accuracy of all models scored significantly higher than chance, with it being 0.20. Also, all of them hold a score comparable to the state-of-the-art literature for five-class classification across datasets applied to psychometric problems (Park et al. 2015, Maldonato et al. 2018) and it is important to stress that text is widely considered as being even more difficult to categorize than images or other types of data since its ambiguousness. Table 7.5 summarizes the performance on the test set of the best for each of the model considered. The CNN, contrary to expectations, scored the lowest, but this is due to the size of the dataset, which is way too small compared to benchmarks for deep neural networks. However, level of accuracy of the MNB on both training and testing for one-hot vectors in the Full and Rest Dataset is fully comparable with that of the SVM with the GloVe word embeddings, although the SVM held a higher accuracy in all cases. A possible interpretation is that while the MNB is a more powerful tool for text categorization, the choice of the right word embeddings can help the SVM with its geometric separation dramatically improving its performance.

From the results, it is evident that Rahim's ROCI-II negotiation scores are robust, and that it is possible to map these dimensions to language features to a certain extent. Also, the higher the difference between the dimensional scores obtained by participants, the more evident these distinctive language features become; however, the differences and similarities, as well as possible relationships among these feature spaces is still to explore. Possible future directions of research include the collection of a wider dataset, to reduce the problems of overfitting across all models; the attempt to classify not only the prevalent ROCI-II score, but the classification of all negotiation dimensions;

the investigation of the classification accuracy class against class.

7.5 Conclusion

A comparison of three machine learning models for the classification of a newly created dataset of labelled negotiation sentences as defined by Rahim and Bonoma negotiation theory was presented. Sentences of participants negotiating with virtual artificial agents inside Enact Game were collected along with the negotiation style assessment provided by the psychological test ROCI-II. Each participant's sentences were labelled with the predominant negotiation style obtained in the test. After the data collection, two datasets were built, one containing all the sentences, and a smaller one containing only the sentences of the individuals that had received a score significantly higher for their predominant negotiation class. Soft margin SVM, Multinomial Naive Bayes and Convolutional Neural Networks were trained on the dataset and a comparison of word embeddings between the GloVe and Word2Vec models was also carried out. Results showed a greater accuracy in case of the smaller dataset, indicating that Rahim and Bonoma's model scales well with emergent language features extracted from the dataset. Also, all models showed an accuracy in line with the state-of-the-art multiclass text classification works, with SVM with GloVe embeddings showing the highest accuracy on the test set. The MNB scores are comparable during training to that of the SVM, but on the test set it scored 2.3 % less than the SVM. This can be interpreted as caused by the positive effect of the use of word embeddings with the SVM, that add relational properties to the words, something that a model that assumes independence such as the MNB cannot benefit. The CNN, the most powerful algorithm, scored lower, and this can be interpreted as due to the current size of the dataset, at the moment smaller than that of standard training sets for CNN.

Chapter 8

Conclusions

The work presented in this thesis represents a novel approach to computational modelling and statistical grounding of soft skills assessment and training and in particular of negotiation skills; this work begins with the development of an e-learning simulation game for a quick, reliable and automated evaluation and personalised training of the user conflict management abilities, and continues with the development a full language model for negotiation text classification based on machine learning and deep learning. The definition and establishment of a complete methodology for building, testing, validating, evaluating and releasing serious games for psychological assessment and training was described in the chapters.

8.1 Achievements of the research

The problem of state-of-the-art soft skills technology-enhanced assessment tools is that they are loosely based on psychological theories, and therefore the assessment they provide may not have validity in comparison with standardised traditional tests, and may not have commonalities with other platforms, or be inclusive towards different cultures, genders along with being inconsistent with literature findings. Other than that, most e-learning tools and serious games rarely include natural language in the user model, whereas natural language is usually either used for the user-agent interaction (e.g. to build responsive chatbots) or for detecting sentiment and/or other non-psychological aspect of the human behaviour. Also, when this occurs, the analyses and results are not grounded or compared with data collected through psychological tests, nor collected through a reliable controlled platform but inferred through spontaneous

social media posts, blogs, forums. The aim of the present work, instead, stems from the need of innovating psychological assessment and proposing a novel way of collecting behavioural and psychological data while integrating gaming motivational aspects and engaging techniques.

The first key step that has been followed was finding a robust and reliable theory that encompasses all aspects of conflict management, and Rahim and Bonoma's psychological model was the perfect candidate: based on dimensions and factors, highly operationalizable thanks to the authors' behavioural and language indicators of each negotiation style, and an internationally validated psychological tool, the ROCI-II, extensively tested on human participants. Moreover, negotiation styles are highly context-dependent, and while there are situational traits, other hidden stable behaviours need to be taken into account. The Enact game platform was born from the mutual collaboration of each ENACT Project partners. The University of Plymouth had the leading role of helping develop a scenario-based adaptive system, providing a reliable operationalization from the theoretical model, building the interaction flow for the intelligent tutoring system and modelling the natural language processing architecture, which was developed outside of the scope of the European Project. The Enact platform was designed to capture the same variables measured with the ROCI-II, substituting what was labelled in the standard test as "the most recent conflict situation" with live 3D scenarios where each user can experience in real-time the consequences of their choices, receive a feedback, interact with the tutor, see the history of each interaction with the virtual agents, receive a personalized training path, respond to virtual agents using natural language and test every situation in a safe environment. Thanks to the ENACT Project consortium, the game was validated in all the partner countries, i.e. UK, Italy, Spain and Turkey. The game underwent two cycles of development and release and each version was tested for its usability, validity and implicit and explicit pleasantness. The usability experiments aimed at evaluating how the serious games was perceived in age and gender groups, how realistic and useful the virtual agents were considered, and how willing the users were to play the scenarios again. Results indicated that not

only it is possible to reproduce verbal and non-verbal aspects of negotiation contexts in artificial settings and with virtual agents, but that this methodology is also capable to provide reliable assessment in comparison with traditional tests. From the user perspective, the game has been also considered pleasant, useful, realistic, educational, and both intrinsically and extrinsically motivating. The game has been perceived as shorter and less tiring than its standard pencil and paper counterpart, and a general interest in playing it again has been shown across different age and cultural groups. Concerning its reliability in comparison with the ROCI-II, the game has been shown to accurately measure in individuals at least two of the negotiation styles considered.

After the end of the validation phase, a natural language processing architecture to recognize the user negotiation style from the natural utterances recorded from each player was integrated with the platform. A series of experiments aimed at collecting a dataset of negotiation interactions between the users and virtual agents in the game scenarios were designed, and a comparison of these sentences and the ROCI-II scores obtained by each participant was carried out to provide an automatic classification that did not stand under the constraints of the previously designed multiple choice interaction. Participants of the study, in detail, were asked to interact using spontaneous sentences with virtual agents in the game, and later asked to fill the ROCI-II questionnaire. This allowed to construct a novel negotiation dataset, the current largest in literature, that includes more than four thousand sentences associated with the ROCI-II score obtained by each user. Natural language understanding was added in the User model of the game and used as a measure to predict the negotiation style from features extracted from the virtual negotiations, in that the collected sentences were pre-processed and corrected, and used for text mining and classification.

Two different approaches have been used in order to achieve this classification. First of all, a statistical analysis of the use of personal pronouns was conducted on the dataset to investigate possible differences in their usage across negotiation styles. Results showed that the use of the pronouns is consistent with the concern for self -

concern for others theory behind the definition of each negotiation style provided by Rahim and Bonoma. In particular, individuals classified as having a high Integrating and Compromising dimension, that have a balanced proportion between concern for self and concern for others (respectively high-high and medium-medium) during negotiations, show a similar usage of the first person and second person pronouns, and this frequency is significantly different from all other styles. Highly Obliging and Avoiding participants, instead, which are less engaged in the negotiation, use significantly less pronouns than the other styles. Predominantly Dominating participants, which show a high disproportion between concern for self and concern for others (high-low), show a very low usage of relational pronouns, whereas it uses significantly more second person pronouns in an attempt to take the lead on the decision.

The second approach to the analysis of the dataset was to use machine learning and deep learning algorithms to find a possible way to cluster the dataset and the best classifier for the dataset so to include it in the game. Word representations were created using word embeddings trained with Word2Vec or GloVe models. State-of-the-art natural language processing architectures were trained on the dataset, in particular Multinomial Naive Bayes, Support Vector Machine and Convolutional Neural Network. On a 5-class classification, all models scored above 0.32 accuracy and the use of word embeddings highly improved the general accuracy of the models.

8.2 Limitations of the present work

Limitations of the present study include methodological and experimental aspects. Concerning the methodology, one of the limitations of the Enact platform is that it is not designed to measure all the scores across all theorized negotiation styles as in the ROCI-II test. In fact, while the ROCI-II is able to capture each of the negotiation styles dimensions, the Enact game can only provide the prevalent style expressed by the user, even if it is able to keep track of the history of answers provided by each user. Another limitation is represented by the fact that the game does not include sounds, and although the tone of the utterances pronounced by the virtual agent and the user

character is represented by the shape of the speech bubble, this aspect could be better integrated and included in the user model and in the expert module. Concerning experimental aspects, one of the problems is represented by the sample, in particular:

- the sample size for the development of the natural language architecture, even if it comprised 425 participants, was not wide enough for training all the machine learning models considered. In fact, most of them and in particular deep networks, showed overfitting on the dataset;
- since the recruitment was conducted both online and offline, age and gender could not be fully controlled and balanced. Even if there is no proven difference in the way negotiation styles are expressed across these two dimensions, more attention on these two important aspects must be put in subsequent experiments;
- other text-categorization machine learning models, such as the emerging recurrent neural networks, could be used for a more thorough comparison.

8.3 Future work and open questions

The present work has given rise to many possible future research questions, that concern both the improvement of the current model and the development of other psychological assessment based serious games. The Enact game architecture, in particular, could be improved in the following ways:

- the number of scenarios could be increased and a methodology for the development of more scenarios could be designed and developed;
- the statistical analysis of the natural language sentences could target not only personal pronouns but also other parts-of-speech, so to make a more accurate prediction;
- the developed natural language architecture could be used to generate sentences that could be added to the model of the virtual agents, to allow a more adaptive flow of the dialogues inside the scenarios;

- the platform could include a speech detection system, so to start analysing and associating the tone with the words used in each utterance;
- the platform could include a facial detection system, to pick the user expression and transform it into a feature to use for the psychological assessment.
- the platform could be targeted not only at adults but also at children, by designing scenarios that can fit their everyday life but also focusing on training rather than assessment.

In the future, the aim is to start to develop more intelligent technology-enhanced platform and provide a full suite for soft skill assessment and training that could also look at the interaction among several psychological aspects and factors. Possible future implementations concern for example leadership skills, problem solving abilities and time management.

To summarize, the importance of the present work for the assessment and training of soft skills encompasses several aspects, in particular it describes:

- the introduction of a novel computational methodology for the development of a free, portable assessment and training game platform;
- the research and development of an adaptive tutoring system;
- the collection of a novel database of negotiation utterances, recorded within a technological game platform;
- the psychologically-grounded prediction of negotiation abilities using verbal and non-verbal data.

Therefore, thanks to the results presented in this thesis, it is possible to provide a new mean for the investigation of all aspects of a technology-enhanced psychological assessment and training, from user to data.

List of references

Adair, W., Okumura, T., Psychology, J. B. J. o. A. & 2001, U. (2001), 'Business Source: Negotiation Behavior When Cultures Collide: The United States and Japan.', *psycnet.apa.org* .

Afzalur Rahim, M., Antonioni, D. & Psenicka, C. (2001), 'A structural equations model of leader power, subordinates' styles of handling conflict, and job performance', *International Journal of Conflict Management* **12**(3), 191–211.

Akkoyunlu, B. (2002), 'Educational Technology in Turkey: Past, Present and Future', *Educational Media International* **39**(2), 165–174.

Aylett, R., Figueiredo, R., Louchart, S., Dias, J. & Paiva, A. (2006), Making it up as you go along-improvising stories for pedagogical purposes, *in* 'Intelligent Virtual Agents', pp. 304–315.

Bancino, R. & Zevalkink, C. (2007), 'Soft Skills: The New Curriculum for Hard- Core Technical Professionals', *Techniques* pp. 20–22.

Bar-On, R. (1997), *Bar-On emotional quotient inventory (EQ-I): A measure of emotional intelligence: Technical manual*.

Baur, T., Damian, I., Gebhard, P., Porayska-Pomsta, K. & André, E. (2013), A job interview simulation: Social cue-based interaction with a virtual character, *in* 'Proceedings - SocialCom/PASSAT/BigData/EconCom/BioMedCom 2013', IEEE, pp. 220–227.

Becker-Asano, C. & Wachsmuth, I. (2009), 'Affective computing with primary and secondary emotions in a virtual human', *Autonomous Agents and Multi-Agent Systems* **20**(1), 32–49.

Bell, M. (2010), 'Online Role-Play : Anonymity , Engagement and Risk Online Role-Play : Anonymity , Engagement and', *Educational Media International* **38**(4), 251–260.

Billingham, R. E. & Sack, A. R. (1987), 'Conflict Tactics and the Level of Emotional Commitment Among Unmarrieds', *Human Relations* **40**(1), 59–74.

Blake, R. R. & Mouton, J. S. (1970), 'The Fifth Achievement', *The Journal of Applied Behavioral Science* **6**(4), 413–426.

Boyatzis, R. E., Stubbs, E. & Taylor, S. (2002), 'Learning Cognitive and Emotional Intelligence Competencies Through Graduate Management Education', *Academy of Management Learning & Education* **1**(2), 150–162.

Boyatzis, R., Goleman, D. & Rhee, K. (2000), 'Clustering competence in emotional intelligence: Insights from the Emotional Competence Inventory (ECI)'.

Boyce, G., Williams, S., Kelly, A. & Yee, H. (2001), 'Fostering deep and elaborative learning and generic (soft) skill development: The strategic use of case studies in accounting education', *Accounting Education* **10**(1), 37–60.

Brusilovsky, P. (2012), 'Adaptive Hypermedia for Education and Training', *Adaptive Technologies for Training and Education* pp. 46–50.

Brusilovsky, P. L. (1992), 'a Framework for Intelligent Knowledge Sequencing and Task Sequencing', *Lecture Notes in Computer Science* **608**, 499–506.

Buabeng-Andoh Charles (2012), 'Factors influencing teachers ' adoption and integration of information and communication technology into teaching : A review of the literature Charles Buabeng-Andoh', *International Journal of Education and Development using Information and Communication Technology* **8**(1), 136–155.

Buda, M., Maki, A. & Mazurowski, M. A. (2017), 'A systematic study of the class imbalance problem in convolutional neural networks'.

Camden, C. & Verba, S. (1986), 'Communication and consciousness: Applications in marketing', *Western Journal of Speech Communication* **50**(1), 64–73.

Campuzano, F., Garcia-Valverde, T., Serrano, E. & Botía, J. c. (2014), 'Generation of human computational models with knowledge engineering', *Engineering Applications of Artificial Intelligence* **35**, 259–276.

Carbonell, J. R. (1970), 'AI in CAI: An Artificial-Intelligence Approach to Computer-Assisted Instruction', *IEEE Transactions on Man-Machine Systems* **11**(4), 190–202.

Cavallo, K. & Brienza, D. (2002), 'Emotional competence and leadership excellence at Johnson & Johnson: The emotional intelligence and leadership study', *for Research on Emotional Intelligence in* pp. 1–12.

Cegala, D. J. (1989), 'A study of selected linguistic components of involvement in interaction', *Western Journal of Speech Communication* **53**(3), 311–326.

Cha, H. J., Kim, Y. S., Park, S. H., Yoon, T. B., Jung, Y. M. & Lee, J. H. (2006), Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system, *in* 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)'.

Chakroun, N., Izaute, M., Lamotte, M. & Droit-Volet, S. (2014), 'Metacognitive Questionnaire on Time: Feeling of the Passage of Time', *Timing & Time Perception* **2**(3), 339–359.

Cipresso, P. & Immekus, J. C. (2017), 'Back to the future of quantitative psychology and measurement: Psychometrics in the twenty-first century', *Frontiers in Psychology* **8**(DEC), 2099.

Colas, F. & Brazdil, P. (2006), 'Comparison of SVM and some older classification algorithms in text classification tasks', *IFIP International Federation for Information Processing* **217**, 169–178.

Conati, C. (2009), Intelligent tutoring systems: new challenges and directions, *in* 'Proceedings of the 21st international jont conference on Artifical intelligence', Vol. 9, pp. 2–7.

Cortina, J. M. (1993), 'What is coefficient alpha? An examination of theory and applications.', *Journal of Applied Psychology* **78**(1), 98–104.

Cosmides, L. & Tooby, J. (2000), Evolutionary Psychology and the Emotions BT - Handbook of Emotions, *in* 'Handbook of Emotions', Vol. 2000, pp. 91–115.

Costa, E., Silva, P., Magalh??es, J. & Silva, M. (2012), An open and inspectable learner modeling with a negotiation mechanism to solve cognitive conflicts in an intelligent tutoring system, *in* 'CEUR Workshop Proceedings', Vol. 872.

D'Alfonso, S., Santesteban-Echarri, O., Rice, S., Wadley, G., Lederman, R., Miles, C., Gleeson, J. & Alvarez-Jimenez, M. (2017), 'Artificial intelligence-assisted online social therapy for youth mental health', *Frontiers in Psychology* **8**(JUN), 796.

De Dreu, C. K., Giebels, E. & Van De Vliert, E. (1998), 'Social motives and trust in integrative negotiation: The disruptive effects of punitive capability', *Journal of Applied Psychology* **83**(3), 408–422.

Deary, I. J., Gow, A. J., Taylor, M. D., Corley, J., Brett, C., Wilson, V., Campbell, H., Whalley, L. J., Visscher, P. M., Porteous, D. J. et al. (2007), 'The lothian birth cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond', *BMC geriatrics* **7**(1), 28.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J. & Fox, H. C. (2004), 'The impact of childhood intelligence on later life: following up the scottish mental surveys of 1932 and 1947.', *Journal of personality and social psychology* **86**(1), 130.

Dede, C. (1986), 'A review and synthesis of recent research in intelligent computer-assisted instruction'.

Deepa, S. & Seth, M. (2013), 'Do Soft Skills Matter? - Implications for Educators Based on Recruiters' Perspective', *IUP Journal of Soft Skills* **7**(1), 7–20.

Dell'Aquila, E., Marocco, D., Ponticorvo, M., Di Ferdinando, A., Schembri, M. & Miglino, O. (2017), *Educational Games for Soft-Skills Training in Digital Environments*, Springer International Publishing, Cham.

Demeester, T., Rocktäschel, T. & Riedel, S. (2016), 'Lifted Rule Injection for Relation Embeddings'.

Deroy, O., Spence, C. & Noppeney, U. (2016), 'Metacognition in Multisensory Perception', *Trends in Cognitive Sciences* **20**(10), 736–747.

Deutsch, M. (1949), 'An Experimental Study of the Effects of Co-Operation and Competition upon Group Process', *Human Relations* **2**(3), 199–231.

Dimaggio, G. & Lysaker, P. H. (2015), 'Metacognition and mentalizing in the psychotherapy of patients with psychosis and personality disorders', *Journal of Clinical Psychology* **71**(2), 117–124.

Dubois, D. (1998), 'The competency casebook: Twelve studies in competency-based performance improvement'.

Duffy, F. D., Gordon, G. H., Whelan, G., Cole-Kelly, K., Frankel, R., Buffone, N., Lofton, S., Wallace, M., Goode, L. & Langdon, L. (2004), 'Assessing competence in communication and interpersonal skills: the Kalamazoo II report.', *Academic medicine : journal of the Association of American Medical Colleges* **79**(6), 495–507.

Dulewicz, V. & Higgs, M. (2000), 'Emotional intelligence - A review and evaluation study', *Journal of Managerial Psychology* **15**(4), 341–372.

Elsom-Cook, M. (1993), 'Student modelling in intelligent tutoring systems', *Artificial Intelligence Review* **7**(3-4), 227–240.

Felder, R. & Silverman, L. (1988), 'Learning and teaching styles in engineering education', *Engineering education* **78**(June), 674–681.

Fisher, R., Ury, W. & Patton, B. (2011), 'Getting to yes: Negotiating agreement without giving in'.

Folkman, S. & Lazarus, R. S. (1988), 'The ways of coping questionnaire', *Null* .

Fraser, B. (1980), 'Conversational mitigation', *Journal of Pragmatics* **4**(4), 341–350.

Freedman, J. (1969), 'Role playing: Psychology by consensus.', *Psycnet.Apa.Org* **13**(2), 107–114.

Furnham, A. (1986), 'Response bias, social desirability and dissimulation', *Personality and Individual Differences* **7**(3), 385–400.

Gamito, P., Oliveira, J., Alghazzawi, D., Fardoun, H., Rosa, P., Sousa, T., Maia, I., Morais, D., Lopes, P. & Brito, R. (2017), 'The art gallery test: A preliminary comparison between traditional neuropsychological and ecological VR-based tests', *Frontiers in Psychology* **8**(NOV), 1911.

Garfield, D. A. S., Rapp, C. & Evens, M. (1992), 'Natural Language Processing in Psychiatry', *The Journal of Nervous and Mental Disease* **180**(4), 227–237.

Giebels, E. & Taylor, P. J. (2009), 'Interaction Patterns in Crisis Negotiations: Persuasive Arguments and Cultural Differences', *Journal of Applied Psychology* **94**(1), 5–19.

Giloth, R. (2000), 'Learning from the field: Economic growth and workforce development in the 1990s', *Economic Development Quarterly* .

Giuffra, C. & Silveria, R. (2013), 'A Multi-Agent System Model to Integrate Virtual Learning Environments and Intelligent Tutoring Systems', *International Journal of Interactive Multimedia and Artificial Intelligence* **2**(1), 51–58.

Goldberg, Y. & Levy, O. (2014), 'word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method'.

Goleman, D. (1995), 'Emotional intelligence', *Emotional intelligence* pp. 33–45.

Goleman, D. (1998), *Working with Emotional Intelligence*.

Gratch, J., Devault, D. & Lucas, G. (2016), The benefits of virtual humans for teaching negotiation, *in* 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)', Vol. 10011 LNAI, Springer, Cham, pp. 283–294.

Graves, A. (2013), *Supervised Sequence Labeling with Recurrent Neural Networks*, Vol. 12, Springer.

Grondin, S. (2010), 'Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions', *Attention, Perception, and Psychophysics* **72**(3), 561–582.

Hambleton, R. K. & Bollwark, J. (1991), 'Adapting tests for use in different cultures: Technical issues and methods'.

Hayes-Roth, F. (1985), 'Rule-based systems', *Communications of the ACM* .

He, L., Fu, M. & Hu, X. (2010), *To improve the social interaction of Web-based Collaborative Learning via online Educational Games for multi-player*, Vol. 2.

Heift, T. & Schulze, M. (2007), 'Errors and Intelligence in Computer-Assisted Language Learning'.

Hendarman, A. F. (2012), 'Relationship among Soft Skills, Hard Skills, and Innovativeness of Knowledge Workers in the Knowledge Economy Era', *Procedia - Social and Behavioral Sciences* **52**, 35–44.

Henderson, H. (1996), *Building a Win-win World: Life Beyond Global Economic Warfare*.

Hsieh, T. C., Wang, T. I., Su, C. Y. & Lee, M. C. (2012), 'A Fuzzy Logic-based Personalized Learning System for Supporting Adaptive English Learning', *Educational Technology and Society* **15**(1), 273–288.

Hsu, C. W. & Lin, C. J. (2002), 'A comparison of methods for multiclass support vector machines', *IEEE Trans Neural Netw* **13**(2), 415–425.

Jackson, D. (2010), 'An international profile of industry-relevant competencies and skill gaps in modern graduates', *The International Journal of Management Education* **8**(3), 29–58.

Jackson, G. T. & McNamara, D. S. (2013), 'Motivation and performance in a game-based intelligent tutoring system.', *Journal of Educational Psychology* **105**(4), 1036–1049.

James, R. & James, M. (2004), 'Teaching career and technical skills in a" mini" business world', *Business* .

Jaques, N., Conati, C., Harley, J. M. & Azevedo, R. (2014), Predicting affect from gaze data during interaction with an intelligent tutoring system, *in* 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)', Vol. 8474 LNCS, Springer, Cham, pp. 29–38.

Jeremić, Z., Jovanović, J. & Gašević, D. (2012), 'Student modeling and assessment in intelligent tutoring of software patterns', *Expert Systems with Applications* **39**(1), 210–222.

Joachims, T. (1998), Text categorization with Support Vector Machines: Learning with many relevant features, Springer, Berlin, Heidelberg, pp. 137–142.

Johnson, R. & Zhang, T. (2014), 'Effective Use of Word Order for Text Categorization with Convolutional Neural Networks'.

Jr, P. P., Brawner, K., Olney, A. & Mitrovic, A. (2013), 'A Review of Student Models Used in Intelligent Tutoring Systems'.

Juárez-Ramírez, R., Navarro-Almanza, R., Gomez-Tagle, Y., Licea, G., Huertas, C. & Quinto, G. (2013), 'Orchestrating an Adaptive Intelligent Tutoring System: Towards Integrating the User Profile for Learning Improvement', *Procedia - Social and Behavioral Sciences* **106**, 1986–1999.

Kadhim, M. A., Afshar Alam, M. & Kaur, H. (2014), 'A multi-intelligent agent architecture for knowledge extraction: Novel approaches for automatic production rules extraction', *International Journal of Multimedia and Ubiquitous Engineering* **9**(2), 95–114.

Kajnc, E. U. (2010), 'What it Takes to Run an EU Presidency: Study of Competences in Slovenia's Public Administration', *ceeol.com* .

Kantrowitz, T. M. (2005), Development and Construct Validation of a Measure of Sft Skills performance, PhD thesis.

Kenny, C. & Pahl, C. (2009), 'Intelligent and adaptive tutoring for active learning and training environments', *Interactive Learning Environments* **17**(2), 181–195.

Kibriya, A. M., Frank, E., Pfahringer, B. & Holmes, G. (2004), Multinomial Naive Bayes for Text Categorization Revisited, Springer, Berlin, Heidelberg, pp. 488–499.

Kim, Y. (2014), 'Convolutional Neural Networks for Sentence Classification'.

Klaus, P., Rohman, J. & Hamaker, M. (2007), 'The hard truth about soft skills: workplace lessons smart people wish they'd learned sooner'.

Knerr, S., Personnaz, L. & Dreyfus, G. (1990), Single-layer learning revisited: a stepwise procedure for building and training a neural network, *in* 'Neurocomputing', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 41–50.

Kuhn, T. & Poole, M. (2000), 'Do conflict management styles affect group decision making? Evidence from a longitudinal field study', *Human Communication Research* **26**(4), 558–590.

Kurdek, L. A. (1994), 'Areas of conflict for gay, lesbian, and heterosexual couples: What couples argue about influences relationship satisfaction', *Journal of Marriage and the Family* **56**(4), 923–934.

Latham, A., Crockett, K., McLean, D. & Edmonds, B. (2012), A conversational intelligent tutoring system to automatically predict learning styles, *in* 'Computers and Education', Vol. 59, pp. 95–109.

Lawrence, P. R. & Lorsch, J. W. (2009), 'Differentiation and Integration in Complex Organizations', *Science* **12**(1), 1–47.

Leacock, C. & Chodorow, M. (1998), 'Combining Local Context and WordNet Similarity for Word Sense Identification', *WordNet: An electronic lexical database.* (JANUARY 1998), 265–283.

LeDoux, J. (2003), 'The emotional brain, fear, and the amygdala'.

Lo, J. J., Chan, Y. C. & Yeh, S. W. (2012), 'Designing an adaptive web-based learning system based on students' cognitive styles identified online', *Computers and Education* **58**(1), 209–222.

Ma, W., Adesope, O. O., Nesbit, J. C. & Liu, Q. (2014), 'Intelligent tutoring systems and learning outcomes: A meta-analysis.', *Journal of Educational Psychology* **106**(4), 901–918.

Maldonato, N. M., Sperandeo, R., Moretto, E. & Dell'Orco, S. (2018), 'A Non-linear predictive model of Borderline Personality Disorder based on multilayer perceptron', *Frontiers in Psychology* **9**(APR), 447.

Marmanis, D., Datcu, M., Esch, T. & Stilla, U. (2016), 'Deep learning earth observation classification using imagenet pretrained networks', *IEEE Geoscience and Remote Sensing Letters* **13**(1), 105–109.

Marocco, D., Pacella, D., Dell'Aquila, E. & Di Ferdinando, A. (2015), Grounding serious game design on scientific findings: The case of ENACT on soft skills training and assessment, *in* 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)', Vol. 9307, Springer, Cham, pp. 441–446.

McCarney, R., Warner, J., Iliffe, S., Van Haselen, R., Griffin, M. & Fisher, P. (2007), 'The Hawthorne Effect: A randomised, controlled trial', *BMC Medical Research Methodology* **7**(1), 30.

McClelland, D. (1973), 'Testing for competence rather than for "intelligence".', *The American psychologist* **28**(1), 1–14.

Mengel, S. & Lively, W. (1992), 'Using a neural network to predict student responses', *Proceedings of the 1992 ACM/SIGAPP symposium* .

Miglino, O., Venditti, A., Delli Veneri, A. & Di Ferdinando, A. (2010), Eutopia-Mt. teaching mediation skills using multiplayer on-line role playing games, *in* 'Procedia - Social and Behavioral Sciences', Vol. 2, pp. 2469–2472.

Mikic Fonte, F. A., Burguillo, J. C. & Nistal, M. L. (2012), 'An intelligent tutoring module controlled by BDI agents for an e-learning platform', *Expert Systems with Applications* **39**(8), 7546–7554.

Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013), 'Efficient Estimation of Word Representations in Vector Space'.

Mikolov, T., Kombrink, S., Burget, L., Cernocky, J. & Khudanpur, S. (2011), Extensions of recurrent neural network language model, *in* '2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)', IEEE, pp. 5528–5531.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. (2013), 'Distributed Representations of Words and Phrases and their Compositionality'.

Miller, G. A., Chodorow, M., Landes, S., Leacock, C. & Thomas, R. G. (1994), Using a Semantic Concordance for Sense Identification, *in* 'Communications of the ACM', Vol. 18, Association for Computational Linguistics, Morristown, NJ, USA, pp. 240–243.

Ming Chia, Y. (2005), 'Job offers of multi-national accounting firms: the effects of emotional intelligence, extra-curricular activities, and academic performance', *Accounting Education* **14**(1), 75–93.

Mirowski, P. & Vlachos, A. (2015), 'Dependency Recurrent Neural Language Models for Sentence Completion'.

Moreno, J. (1946), 'Experimental psychodrama.'.

Moseley, D., Higgins, S., Bramald, R., Hardman, F., Miller, J., Mroz, M., Tse, H., Newton, D., Thompson, I., Williamson, J., Halligan, J., Bramald, S., Newton, L., Tymms, P., Henderson, B. & Stout, J. (1999), 'Ways forward with ICT: Effective pedagogy using Information and Communications Technology for Literacy and Numeracy in Primary Schools', p. 170.

Moss, P. & Tilly, C. (1996), '"Soft" skills and race: An investigation of black men's employment problems', *Work and Occupations* .

Murata, A. (2008), 'Human Error Management Paying Emphasis on Decision Making and Social Intelligence -Beyond the Framework of Man-Machine Interface Design-'.

Ohlsson, S. (1987), 'Truth versus appropriateness: Relating declarative to procedural knowledge', *Production system models of learning and* .

Olekalns, M. & Smith, L. (2000), 'Understanding optimal outcomes: The role of strategy sequences in competitive negotiations', *Human Communication Research* **26**(4), 527–557.

Ong, J. & Ramachandran, S. (2005), 'An Intelligent Tutoring System Approach to Adaptive Instructional Systems'.

Pacella, D., Dell'Aquila, E., Marocco, D. & Furnell, S. (2017), Toward an automatic classification of negotiation styles using natural language processing, Vol. 10498 LNAI, Springer, Cham, pp. 339–342.

Pacella, D. & López-Pérez, B. (2018), 'Assessing children's interpersonal emotion regulation with virtual agents: The serious game Emodiscovery', *Computers and Education* **123**, 1–12.

Pacella, D., Ponticorvo, M., Gigliotta, O. & Miglino, O. (2017), 'Basic emotions and adaptation. A computational and evolutionary model', *PLoS ONE* **12**(11), e0187463.

Panksepp, J. (1998), *Affective neuroscience: The foundations of human and animal emotions*, Vol. 4.

Panksepp, J. (2010), 'Affective consciousness in animals: perspectives on dimensional and primary process emotion approaches', *Proceedings of the Royal Society B: Biological Sciences* **277**(1696), 2905–2907.

Pant, I. & Baroudi, B. (2008), 'Project management education: The human skills imperative', *International Journal of Project Management* **26**(2), 124–128.

Park, G., Andrew Schwartz, H., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H. & Seligman, M. E. (2015), 'Automatic personality assessment through social media language', *Journal of Personality and Social Psychology* **108**(6), 934–952.

Pate, J., Martin, G. & Robertson, M. (2003), 'Accrediting Competencies: A Case of Scottish Vocational Qualifications', *Journal of European Industrial Training* **27**(2/3/4), 169–176.

Paviotti, G., Rossi, P. & Zarka, D. (2012), 'Intelligent tutoring systems: an overview', *Pensa Multimedia* .

Pennington, J., Socher, R. & Manning, C. (2014), Glove: Global Vectors for Word Representation, *in* 'Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)', Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 1532–1543.

Polizzi, G. (2011), 'Measuring School Principals' Support for ICT Integration in Palermo, Italy', *Journal of Media Literacy Education* **32**, 113–122.

Polson, M. & Richardson, J. (2013), 'Foundations of intelligent tutoring systems'.

Polyak, S. T., von Davier, A. A. & Peterschmidt, K. (2017), 'Computational psychometrics for the measurement of collaborative problem solving skills', *Frontiers in Psychology* **8**(NOV), 2029.

Porter, M. (1980), 'An algorithm for suffix stripping', *Program* **14**(3), 130–137.

Potaka, L. & Cochrane, S. (2004), 'Developing Bilingual Questionnaires : Experiences from New Zealand in the Development of ÂÍ ori Language Survey the 2001 Ma', *Journal of Official Statistics* **20**(2), 289–300.

Pratt, J. a., Claire, E., Hauser, K. & Ross, S. C. (2009), 'IS Staffing During a Recession: Comparing Student and IS Recruiter Perceptions', *Journal of Information Systems Education* **21**(1), 69–85.

Proctor, R. & Dutta, A. (1995), 'Skill acquisition and human performance.'.

Pruitt, D. & Carnevale, P. (1993), *Negotiation in social conflict.*

Pruitt, D. G. (1983), 'Strategic Choice in Negotiation', *American Behavioral Scientist* **27**(2), 167–194.

Pruitt, D. G. & Lewis, S. A. (1975), 'Development of integrative solutions in bilateral negotiation', *Journal of Personality and Social Psychology* **31**(4), 621–633.

Putnam, L. L. & Jones, T. S. (1982), 'Reciprocity in negotiations: An analysis of bargaining interaction', *Communication Monographs* **49**(3), 171–191.

Putnam, L. L. & Wilson, C. E. (1982), 'Communicative Strategies in Organizational Conflicts: Reliability and Validity of a Measurement Scale', *Annals of the International Communication Association* **6**(1), 629–652.

Rahim, A. & Bonoma, T. V. (1979), 'Managing Organizational Conflict: a Model for Diagnosis and Intervention', *Psychological Reports* **44**(3c), 1323–1344.

Rahim, M. (1983), 'A Measure of Handling Interpersonal Conflict', *Academy of Management Journal* **26**(2), 368–376.

Rahim, M. A. & Magner, N. R. (1995), 'Confirmatory factor analysis of the styles of handling interpersonal conflict: first-order faactor model and its invariance accross groups', *Journal of Applied Psychology* **80**, 122–132.

Rands, M., Levinger, G. & Mellinger, G. D. (1981), 'Patterns of Conflict Resolution and Marital Satisfaction', *Journal of Family Issues* **2**(3), 297–321.

Redondo, M. Á., Bravo, C., Ortega, M. & Verdejo, M. F. (2007), 'Providing adaptation and guidance for design learning by problem solving: The design planning approach in DomoSim-TPC environment', *Computers and Education* **48**(4), 642–657.

Robles, M. M. (2012), 'Executive Perceptions of the Top 10 Soft Skills Needed in Today's Workplace', *Business Communication Quarterly* **75**(4), 453–465.

Rogmann, J. J. (2008), 'Evaluating Computer-Supported Social Learning : An Integrative Study of the Development of Soft Skills in a University-Based Mediation Training Programme presented by'.

Rong, X. (2014), 'word2vec Parameter Learning Explained'.

Rudrauf, D., David, O., Lachaux, J.-P., Kovach, C. K., Martinerie, J., Renault, B. & Damasio, A. (2008), 'Rapid Interactions between the Ventral Visual Stream and Emotion-Related Structures Rely on a Two-Pathway Architecture', *Journal of Neuroscience* **28**(11), 2793–2803.

Saaty, T. L. (1980), 'The analytic hierarchy process (new york: Mcgrawhill, 1980)', *MATH Google Scholar* .

Salovey, P. & Mayer, J. D. (1989), 'Emotional Intelligence', *Imagination, Cognition and Personality* **9**(3), 185–211.

Salton, G. & Buckley, C. (1988), 'Term-weighting approaches in automatic text retrieval', *Information processing & management* **24**(5), 513–523.

Sandberg, J., Maris, M. & Hoogendoorn, P. (2014), 'The added value of a gaming context and intelligent adaptation for a mobile learning application for vocabulary learning', *Computers and Education* **76**, 119–130.

Schilling, M. A. (2000), 'Decades ahead of her time: advancing stakeholder theory through the ideas of Mary Parker Follett', *Journal of Management History (Archive)* **6**(5), 224–242.

Sebastiani, F. (2001), 'Machine Learning in Automated Text Categorization'.

Seymour, B. & Dolan, R. (2008), 'Emotion, Decision Making, and the Amygdala'.

Shute, V. & Psotka, J. (1995), Intelligent tutoring systems: Past, Present, and Future, *in* 'Handbook of Research on Educational Communications and Technology'.

Sjoberg, L. (2001), 'Emotional Intelligence: A Psychometric Analysis', *European Psychologist* **6**(2), 79–95.

Skinner, B. F. (1963), 'Operant behavior.', *American Psychologist* **18**(8), 503.

Sleeman, D. H. & Brown, J. S. (1982), 'Intelligent tutoring systems', *Science (New York, N.Y.)* **228**(4698), 456–462.

Snyder, A. V. & Ebeling, H. W. J. (1992), 'Targeting a company's real core competencies.', *The Journal of business strategy* **13**(6), 26–32.

Sparck Jones, K. (1972), 'A statistical interpretation of term specificity and its application in retrieval', *Journal of documentation* **28**(1), 11–21.

Steenbergen-Hu, S. & Cooper, H. (2014), 'A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning.', *Journal of Educational Psychology* **106**(2), 331–347.

Sternberg, R. (1997), 'Successful intelligence'.

Sukhoo, A., Barnard, A., Eloff, M. M. & Poll, J. A. V. D. (2005), 'Accommodating Soft Skills in Software Project Management Motivation for this Research', *Issues in Informing Science and Information Technology* **2**, 691–703.

Sultana, R. G. (2009), 'Competence and competence frameworks in career guidance: Complex and contested concepts', *International Journal for Educational and Vocational Guidance* **9**(1), 15–30.

Thayer, S. & Schiff, W. (1975), 'Eye-contact, facial expression, and the experience of time', *Journal of Social Psychology* **95**(1), 117–124.

Thomas, J., Divakaran, A. & Khan, S. (2013), 'Leveraging a generalized tutoring framework in exploratory simulations of ill-defined domains', *AIED 2013 Workshops* .

Thomas, K. W. (1992), 'Conflict and conflict management: Reflections and update', *Journal of Organizational Behavior* **13**(3), 265–274.

Thorndike, E. L. (1920), 'Intelligence and its uses.', *Harper magazine* .

Van De Vliert, E. & Hordijk, J. W. (1989), 'A Theoretical Position of Compromising Among Other Styles of Conflict Management', *The Journal of Social Psychology* **129**(5), 681–690.

Van de Vliert, E. & Kabanoff, B. (1990), 'Toward Theory-Based Measures of Conflict Management.', *Academy of Management Journal* **33**(1), 199–209.

Van Rijsbergen, C. J. & Van, C. J. (2016), *Information Retrieval*, Vol. 573, Butterworths.

Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F. & Mislevy, R. J. (2000), *Computerized adaptive testing: A primer*, Routledge.

Walton, R. & McKersie, R. (1965), *A behavioral theory of labor negotiations: An analysis of a social interaction system*.

Wang, T. & Mitrovic, A. (2002), 'Using neural networks to predict student's performance', *Computers in Education, 2002.* (October), 969–973.

Watson, C. M. (1983), 'Leadership, management, and the seven keys', *Business Horizons* **26**(2), 8–13.

Watson, J. B. (2017), *Behaviorism*, Routledge.

Watt, J. D. (1991), 'Effect of boredom proneness on time perception.', *Psychological reports* **69**(1), 323–327.

Weber, M. R., Finley, D. A., Crawford, A. & Rivera, D. (2009), 'An exploratory study identifying soft skill competencies in entry-level managers', *Tourism and Hospitality Research* **9**(4), 353–361.

Weider-Hatfield, D. (1988), 'Assessing the rahim organizational conflict inventory-II (ROCI-II)', *Management Communication Quarterly* **1**(3), 350–366.

Weingart, L. R., Prietula, M. J., Hyder, E. B. & Genovese, C. R. (1999), 'Knowledge and the Sequential Processes of Negotiation: A Markov Chain Analysis of Response-in-Kind', *Journal of Experimental Social Psychology* **35**(4), 366–393.

Wong, B. Y. L. (1991), The relevance of metacognition to learning disabilities, *in* 'Learning about learning disabilities.', Elsevier, pp. 231–258.

Woodruffe, C. (1993), 'What Is Meant by a Competency?', *Leadership & Organization Development Journal* **14**(1), 29–36.

Woolf, B. & Cunningham, P. A. (1987), 'Multiple Knowledge Sources in Intelligent Teaching Systems', *IEEE Expert-Intelligent Systems and their Applications* **2**(2), 41–54.

Xavier, D., Crespo, B. & Fuentes-Fernández, R. (2015), 'A rule-based expert system for inferring functional annotation', *Applied Soft Computing Journal* **35**, 373–385.

Yager, R. R. & Zadeh, L. A. (1992), *An Introduction to Fuzzy Logic Applications in Intelligent Systems*, Springer US.

Yamauchi, T. (2001), 'Learning from human tutoring', *Cognitive Science* pp. 471–553.

Yardley-Matwiejczuk, K. (1997), 'Role play: theory and practice'.

Yee, N. (2006), 'Motivations for Play in Online Games', *CyberPsychology & Behavior* **9**(6), 772–775.

Zartman, I. & Berman, M. (1982), *The practical negotiator*.

Zhang, X., Zhao, J. & LeCun, Y. (2015), 'Character-level Convolutional Networks for Text Classification'.

Zigler, E. F. (1973), 'Project Head Start: success or failure?'.

Appendix A

TNA Semi-structured interviews

A.1 General description

1. Could you please describe the organisation/enterprise/school (target group context) where you work/operate, your specific activities and role?

A.2 Participants' competencies/previous experiences/individual needs

2. Could you please describe technological competencies of the target group that will be involved into ENACT research study and trialling? Are they used to technology?

3. Is your school well technology equipped and has got Internet connection? (Only for teacher to verify technological availability)

4. Which kind of operating system and browser do you usually use? Do you normally accede to Internet high-speed connection?

5. In your school/enterprise/association (target group context), do you foresee mandatory or voluntary courses to develop social competencies (e.g. leadership, communication, conflict management, teamwork and so on)? If so, how are they organized and which activities are provided?

6. Are online role-playing or/and computer based simulation methods used in your school/enterprise/organization (target group context) as learning tools? If so, can you describe aims, objectives, and structure of the activities? For interviewee to share with interviewees: role-playing a game through which you can assume a

role and control a character according and within a specific situation; Computer based simulation concerns a reproduction of a specific aspect, behaviour, model, theory, experience you can observe, act or interact with, and reflect about.

7. In your experiences, have you ever used role-play and simulations, not necessarily computer-based? If so, in which way and why?

8. I would be grateful if you could give me a definition of what negotiation skill means to you? If possible, can you give me an example by recalling a situation you personally experienced that can clarify my understanding of what you mean? You can also decide to discuss something that is related to your personal experience and not necessarily to your work or study.

9. Are negotiation and communication skills important in your work/target group context as an employee/student, etc.? If so, why and what is the added value to develop them? Can you give me an example by recalling a situation you personally experienced that can clarify my understanding of what you mean?

10. In your opinion, how these competencies can be promoted in your work/study (target group context)?

A.3 Availability to participate in the project, expectations

11. Are you interested to learn more about these competencies with an online game?

12. In your opinion, what are the advantages and disadvantages of this kind of activities (online games in order to learn more about, and possibly further develop social competencies) in your specific work/activity(target group context)?

13. Considering your specific work/activity (target group context), what are your expectations about the project? In which way ENACT project can help yourself in your specific work?

14. Do you think that participating in the project can help you? Could you give me an example to understand what you mean?

Appendix B

TNA Online survey

There are many ways to approach negotiations. As a preliminary activity in the ENACT project, this questionnaire is designed to make you reflect on your own approach to negotiation and on the ways you might improve or round out your negotiating skills. We want to reassure you that this not an evaluative questionnaire, therefore there are no "right" or "wrong" answers to the questions below. We believe that you can really help us to gather valuable information regarding your experience and thoughts about negotiation.

1. Are you aware of your negotiation skills, and when you exercise and use those skills? On a scale of 1-4 how would you rate your level of awareness as 4 being extremely aware, and 1 not aware at all?

    - Not at all aware
    - Somewhat aware
    - Moderately aware
    - Extremely aware

2. How important do you think are negotiation skills in your professional/school life? On a scale of 1-4 how would you rate them as 4 being most important and 1 least important?

    - Not at all important
    - Somewhat important

- Moderately important
- Extremely important

3. How important do you think are negotiation skills in your personal life? On a scale of 1-4 how would you rate them as being 4 being most important and 1 least important?
    - Not at all important
    - Somewhat important
    - Moderately important
    - Extremely important
4. Thinking about your working/school context how do you think you interact with others when you are involved in a negotiation (e.g. making a decision, planning an activity, asking for a change at work, etc.?)
5. I try to understand my interlocutor's interests, and investigate an issue with my supervisor/teacher/line manager/trainer to find a solution acceptable to us.
6. I tend to satisfy the needs of my supervisor/teacher/trainer/coach/line manager, etc., as he/she understand what can be better for me.
7. I take the stock of the situation and I use my influence to get my ideas accepted
8. I try to avoid being in the situation of making a decision about something very quickly and I withhold my conflict with my supervisor/teacher/trainer/coach/ line manager, etc.
9. I attempt to find a middle course arrangement to resolve an impasse, or taking decisions.
10. Thinking about your personal life how do you think you interact with parents, friends, etc. when you are involved in a negotiation (e.g. making a decision, planning an activity, asserting your opinions or point of views, etc.?)

- I try to understand my interlocutor's interests, investigate an issue with him/her and to find a solution acceptable to us.
- I'm inclined to satisfy their needs, as they understand what can be better for me.
- I take the stock of the situation and I use my influence to get my ideas accepted
- I try to avoid being in the situation of making a decision about something very quickly and try to keep my conflict with them to myself.
- I attempt to find a middle course arrangement to resolve an impasse, or taking decisions.

11. During a negotiation session, what do you when listening parties or interlocutors? On a scale of 1-4, with 4 being almost always and 1 almost never, how would you rate yourself for each of the following statements?

    - I pay attention even if I am not directly interested in the immediate topic or in the person I am talking with 1 - 2 - 3 - 4
    - I stop myself from interrupting the person I am talking with 1 - 2 - 3 - 4
    - If in a large meeting, I find I do not speak up. 1 - 2 - 3 - 4
    - I maintain emotional control, regardless of what it is said 1 - 2 - 3 - 4
    - I find it difficult to say "no". 1 - 2 - 3 - 4
    - I pay attention to my body language as a listener, and I try to tune my communication style to my interlocutor's style 1 - 2 - 3 - 4

12. Do you think there is a specific moment to stop negotiations? And how do you know when walking away is appropriate? (You can select more than one option)

    - When the other side is entrenched in a position that is very far from my own.
    - When the parties are so far apart it would take a whole lot of work to get them agreeing.

- When I can try again later.
- When the agreement is established.
- When I have a better alternative that meets my interests better than what I have been offered
- When I understand that my proposal needs to be developed more carefully in order to meet parties' interests more fully than the alternative they might have.

13. Among the relational skills listed below, please tick only 4 skills that you believe are particularly involved in negotiation processes

    - Communication (and active listening)
    - Team working
    - Problem solving
    - Critical and creative thinking
    - Empathy
    - Result orientation
    - Judgment and decision making
    - Emotional stability
    - Self-awareness

14. Taking each of the soft skills listed below, how much do you think that being involved in the project could help in improving those skills for your personal and/or professional development? Tick one box on each line, where 4 = Very helpful; 3 = Fairly helpful; 2 = Not very helpful; 1 = Not at all helpful

    - Communication (and active listening) 1 - 2 - 3 - 4
    - Team working 1 - 2 - 3 - 4
    - Problem solving 1 - 2 - 3 - 4

- Critical and creative thinking 1 - 2 - 3 - 4
- Empathy 1 - 2 - 3 - 4
- Result orientation 1 - 2 - 3 - 4
- Judgment and decision making 1 - 2 - 3 - 4
- Emotional stability 1 - 2 - 3 - 4
- Self-awareness 1 - 2 - 3 - 4
- Other, specify ..... 1 - 2 - 3 - 4

15. I believe that training is required to help development of communication and negotiation skills. Please rate your level of agreement or disagreement (Please thick one box only)
    - Strongly disagree
    - Disagree
    - Agree
    - Strongly agree
16. I believe that specifically designed computer games supporting the development of negotiation skills can be very beneficial. Please rate your level of agreement or disagreement (Please tick one box only)
    - Strongly disagree
    - Disagree
    - Agree
    - Strongly agree
17. Please, feel free and add any comment and/or other expectation about the ENACT project.

Appendix C

ROCI II

Rahim Organizational Conflict Inventory–II, Form C

Strictly Confidential

Please check the appropriate box after each statement, to indicate *how you handle your disagreement or conflict with your peers*. Try to recall as many recent conflict situations as possible in ranking these statements.

	Strongly Disagree				*Strongly Agree*
	1	2	3	4	5
1. I try to investigate an issue with my peers to find a solution acceptable to us.	❑	❑	❑	❑	❑
2. I generally try to satisfy the needs of my peers.	❑	❑	❑	❑	❑
3. I attempt to avoid being "put on the spot" and try to keep my conflict with my peers to myself.	❑	❑	❑	❑	❑
4. I try to integrate my ideas with those of my peers to come up with a decision jointly.	❑	❑	❑	❑	❑
5. I try to work with my peers to find solution to a problem that satisfies our expectations.	❑	❑	❑	❑	❑
6. I usually avoid open discussion of my differences with my peers.	❑	❑	❑	❑	❑
7. I try to find a middle course to resolve an impasse.	❑	❑	❑	❑	❑
8. I use my influence to get my ideas accepted.	❑	❑	❑	❑	❑
9. I use my authority to make a decision in my favor.	❑	❑	❑	❑	❑
10. I usually accommodate the wishes of my peers.	❑	❑	❑	❑	❑
11. I give in to the wishes of my peers.	❑	❑	❑	❑	❑
12. I exchange accurate information with my peers to solve a problem together.	❑	❑	❑	❑	❑
13. I usually allow concessions to my peers.	❑	❑	❑	❑	❑
14. I usually propose a middle ground for breaking deadlocks.	❑	❑	❑	❑	❑
15. I negotiate with my peers so that a compromise can be reached.	❑	❑	❑	❑	❑
16. I try to stay away from disagreement with my peers.	❑	❑	❑	❑	❑
17. I avoid an encounter with my peers.	❑	❑	❑	❑	❑
18. I use my expertise to make a decision in my favor.	❑	❑	❑	❑	❑
19. I often go along with the suggestions of my peers.	❑	❑	❑	❑	❑
20. I use "give and take" so that a compromise can be made.	❑	❑	❑	❑	❑
21. I am generally firm in pursuing my side of the issue.	❑	❑	❑	❑	❑
22. I try to bring all our concerns out in the open so that the issues can be resolved in the best possible way.	❑	❑	❑	❑	❑
23. I collaborate with my peers to come up with decisions acceptable to us.	❑	❑	❑	❑	❑
24. I try to satisfy the expectations of my peers.	❑	❑	❑	❑	❑
25. I sometimes use my power to win a competitive situation.	❑	❑	❑	❑	❑
26. I try to keep my disagreement with my peers to myself in order to avoid hard feelings.	❑	❑	❑	❑	❑
27. I try to avoid unpleasant exchanges with my peers.	❑	❑	❑	❑	❑
28. I try to work with my peers for a proper understanding of a problem.	❑	❑	❑	❑	❑

Appendix D

Consent Form

PLYMOUTH UNIVERSITY

FACULTY OF SCIENCE AND ENGINEERING

CONSENT TO PARICIPATE IN RESEARCH PROJECT

Principal Investigator: Davide Marocco

Title of Research: ENACT project study on negotiation skills

Brief statement of purpose of work:
You are being invited to participate in a research study that attempts to investigate and learn more about negotiation skills in both personal life and business success. This research project is being coordinated by Dr. Davide Marocco, Reader at Plymouth University School of Computing, Electronics and Mathematics.
All the process is designed to be completely anonymous and all possible efforts for it to remain anonymous have and will be in place for the duration of the research process. No identifying information will be collected or published upon the completion of this research.
The study will consists of a series of non-evaluative questionnaires and an online game experience.
If you wish to stop taking the process for any reason you may.
This research is anonymous; please make every attempt to answer each question and to play as honestly as possible.
If you have any questions or concerns about being in this study, you may contact Dr. Davide Marocco at davide.marocco@plymouth.ac.uk.

The objectives of this research have been explained to me.

I understand that I am free to withdraw from the research at any stage, and ask for my data to be destroyed if I wish.

I understand that my anonymity is guaranteed, unless I expressly state otherwise.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities (e.g. under COSHH regulations)

Under these circumstances, I agree to participate in the research.

Appendix E

Information Sheet

PLYMOUTH UNIVERSITY

FACULTY OF SCIENCE AND ENGINEERING

INFORMATION SHEET

Principal Investigator: Davide Marocco

Title of Research: ENACT project study on negotiation skills

This research study attempts to investigate and learn more about negotiation skills in both personal life and business success. This research project is being coordinated by Dr. Davide Marocco, Reader at Plymouth University School of Computing, Electronics and Mathematics, and is funded by EACEA under the EU Leonardo da Vinci KA3 action.

ENACT implements:

- a training methodology, in the form of an online "game", which emphasizes practical skills by exploiting an innovative technology that uses autonomous intelligent agents as virtual interlocutors for different target users
- an innovative and scientifically sound assessment methodology, in the form of an intelligent tutor, capable of providing reliable analysis of the user's negotiation skills and communication style.

The distinct contribute that we would like the ENACT Platform to give to the serious game and e-learning community is dependent on three key concepts:

1. The gamification of standardised psychological tests and, on the other hand, the standardisation of a serious game;
2. The use of Artificial Intelligence for creating a unique system which can actively interact and guide the user for the purpose of training and assessment;
3. The possibility to improve the reliability of the platform with the collection of user data thanks to machine learning techniques.

If you have any questions or concerns about being in this study, or, if you wish to cancel your data, you may contact Dr. Davide Marocco at davide.marocco@plymouth.ac.uk.

Appendix F

Scenario 1: Compromising

F.1 Introductions

THIS scenario deals about two brothers of opposite genders, Gina and Marco, who have to reach a deal about how to share the motorbike on a Saturday night. The scenario options and details are reported as if the user gender was male.

F.1.0.1 Introduction to the scenario if Integrating style is the most appropriate

Situation description: Two young siblings, Marco and Gina, need to use the same motorbike the same evening. Usually you have very different schedules.

User description: You play the role of Marco. You know that you absolutely can't miss the appointment with your friends.

Virtual Agent description: Gina: your sister, really needs the motorbike to get to the centre and go out with her friends.

Additional details: Remember: - your appointment is very important. - If only you go out, Gina may not see her boyfriend. - You would like both of you to go out with friends.

F.1.0.2 Introduction to the scenario if Dominating style is the most appropriate

Situation description: Two young siblings, Marco and Gina, need to use the same motorbike the same evening.

User description: You play the role of Marco. You know that you absolutely can't miss the appointment with your friends.

Virtual Agent description: Gina: your sister, really needs the motorbike to get to the

centre and go out with her friends. The last time she took the motorbike, she brought it back without fuel and with one more scratch.

Additional details: Remember: - Because of Gina, the motorbike has two scratches on the bodywork.

F.1.0.3 Introduction to the scenario if Obliging style is the most appropriate

Situation description: Two young siblings, Marco and Gina, need to use the same motorbike the same evening.

User description: You play the role of Marco. You know that you absolutely can't miss the appointment with your friends.

Virtual Agent description: Gina: your sister, really needs the motorbike to get to the centre and go out with her friends. She has had back problems in recent weeks, and because of this she has only gone out very seldom.

Additional details: Remember: - Gina is really excited about going out. - Gina can't wait to go out after the back problems she has had.

F.1.0.4 Introduction to the scenario if Avoiding style is the most appropriate

Situation description: Two young siblings, Marco and Gina, need to use the same motorbike the same evening.

User description: You play the role of Marco. You know that you absolutely can't miss the appointment with your friends. You are almost done watching the exclusive of the latest episode of Game of Thrones that you don't want to miss at all

Virtual Agent description: Gina: your sister, really needs the motorbike to get to the centre and go out with her friends.

Additional details: Remember: - The episode you are watching, is live on an American channel, you will not be able to watch it again.

F.1.0.5 Introduction to the scenario if Compromising style is the most appropriate

Situation description: Two young siblings, Marco and Gina, need to use the same motorbike the same evening. The city centre is not far away, and usually you meet your respective groups of friends not far from there.

User description: You play the role of Marco. You know that you absolutely can't miss the appointment with your friends.

Virtual Agent description: Gina: your sister, really needs the motorbike to get to the centre and go out with her friends.

Additional details: Remember: - The important thing for you is to find a solution without too many sacrifices.

F.2 Interaction

F.2.0.1 First interaction

Virtual agent: Tonight mum told me that you need the motorbike as well!

User: Yes, I didn't know you needed it too!

Virtual agent: I would like to go out as well, what can we do?

Possible user choices:

1. Integrating: Where are you going tonight? Maybe we could go to the same area!
2. Dominating: You could take public transport tonight ... I asked mum for the motorbike two days ago!
3. Obliging: I wouldn't know, you are usually good at solving problems! Suggest a solution!
4. Avoiding: I understand that you are in a hurry, can we take a moment?
5. Compromising: Let's spend some time with my friends and some with your friends,

so we can both go out!

Virtual agent responses:

1. I was leaving now to go to the party I told you about, can we find a compromise?
2. I have a really important appointment and I can't go by public transport, can we find a compromise?
3. Let's find a compromise so neither of us is unhappy.
4. I should go out and I would like to talk about it now, what do you reckon on finding a quick compromise?
5. It is a good compromise, but your friends could keep us longer, we need another solution...

F.2.0.2 Second interaction

Possible user choices:

1. Integrating: First let us think of the timing. We could synchronize without overlapping, what do you think?
2. Dominating: I can go out after you! You go out now and when you come back I will take the motorbike, ok?
3. Obliging: You usually go out before me ... If you tell me what time you are going out, I will try to avoid overlaps.
4. Avoiding: I am still watching the episode of Game of Thrones, I am going out quite a lot later!
5. Compromising: You go out at 10 pm and come back before midnight. After midnight I will use the motorbike. What do you reckon?

Virtual agent responses:

1. Unfortunately I have to go out soon and I will come back very late. Why don't you ask dad for the car so we both have transport?
2. I have to go out soon and I will come back very late. Why don't you ask dad for the car so we both have transport?
3. Tonight I will be back late. Why don't you ask dad for the car so we both have transport?
4. I will be back late... Why don't you ask dad for the car before he leaves the house, so we both have transport?
5. Tonight I will be back later than usual. Why don't you ask dad for the car so we both have transport?

F.2.0.3 Third interaction

Possible user choices:

1. Integrating: Dad needs the car. We could find together a way to both go out with no worries, what do you think?
2. Dominating: Dad needs the car. Tonight you will have to go with public transport.
3. Obliging: Dad needs the car, but I know that perhaps this evening you need the motorbike more than I do.
4. Avoiding: I can talk to dad also after the episode, there are only 20 minutes to go till the end!
5. Compromising: Dad needs the car. We must understand how to share the motorbike in a fair manner.

Virtual agent responses:

1. It is important to find an acceptable compromise.

2. I do not think it is a good compromise.

3. Indeed, it would be very useful to me, but I am willing to find a compromise.

4. I have no time ... come on, let's look for a compromise.

5. If we find a way, it would be a good compromise.

F.2.0.4 Fourth interaction

Possible user choices:

1. Integrating: I could take you with myself on the motorbike and pick you up when you finish. What do you think?

2. Dominating: Tonight I want to go out and I do not want to change my plans.

3. Obliging: Ok, I can forgo going out tonight. Try to be careful with the motorbike and have fun!

4. Avoiding: There is now little time until the end of the episode, when it ends we can talk about it again ...

5. Compromising: The only solution is to use the motorbike for the same number of hours each.

Virtual agent responses:

1. It seems to be a good solution, will we manage our timekeeping?

2. Ok, you take it, but next time I will take it.

3. Ok, perfect, thanks! Maybe next time we can split it.

4. Ok, we can talk about it later even if there will be less time to decide.

5. It seems to be a good solution, will we manage our timekeeping?

F.2.0.5 Final interaction

Possible user choices:

1. Integrating: I will talk with [VirtualAgent] about the possibility of taking her on the motorbike and then going to collect her.
2. Dominating (if last choice was Dominating): As agreed with [VirtualAgent], I will take the motorbike tonight.
3. Dominating (if last choice was not Dominating): I will try to take the motorbike tonight.
4. Obliging (if last choice was Obliging): As agreed with [VirtualAgent], she will take the motorbike tonight.
5. Obliging (if last choice was not Obliging): I will let [VirtualAgent] take the motorbike tonight.
6. Avoiding (if last choice was Avoiding): As agreed with [VirtualAgent], the decision will be taken after the end of the episode.
7. Avoiding (if last choice was not Avoiding): I will try to postpone the decision until after the end of the episode.
8. Compromising: I will talk with [VirtualAgent] about the possibility of sharing the bike for the same number of hours.

Appendix G

Scenario 2: Integrating

G.1 Introductions

THIS scenario deals about two friends of the same gender who have to decide where to eat and what type of food to get. One would like to eat Chinese, while the other would prefer a pizza. The scenario options and details are reported as if the user gender was male.

G.1.0.1 Introduction to the scenario if Integrating style is the most appropriate

Situation description: Two friends Vittorio and Roberto must decide where to go for dinner. You are waiting for confirmation from your friends for the after dinner.

User description: You are Roberto, and you really want to eat pizza. You want it too much!

Virtual Agent description: Vittorio is passionate about Chinese cuisine and loves eating oriental food with chopsticks.

Additional details: Remember: - You would like to satisfy both your appetites. - There are many different options in the surrounding area.

G.1.0.2 Introduction to the scenario if Dominating style is the most appropriate

Situation description: Two friends Vittorio and Roberto must decide where to go for dinner. You are waiting for confirmation from your friends for the after dinner.

User description: You are Roberto, and you really want to eat pizza. Due to Vittorio's preferences you eat Chinese at least three times a week and you feel that you want to

eat something different this time.

Virtual Agent description: Vittorio is passionate about Chinese cuisine and loves eating oriental food with chopsticks.

Additional details: Remember: - You and Vittorio eat Chinese very often during the week. - You really want pizza.

G.1.0.3 Introduction to the scenario if Obliging style is the most appropriate

Situation description: Two friends Vittorio and Roberto must decide where to go for dinner. You are waiting for confirmation from your friends for the after dinner.

User description: You are Roberto, and you really want to eat pizza. You want it too much!

Virtual Agent description: Vittorio is passionate about Chinese cuisine and loves eating oriental food with chopsticks. He works all day in a pizzeria, and because of exams he has not gone out for a long time.

Additional details: Remember: - Vittorio can't wait to go out after the long time he has spent on books and work. - Vittorio works in a pizzeria.

G.1.0.4 Introduction to the scenario if Avoiding style is the most appropriate

Situation description: Two friends Vittorio and Roberto must decide where to go for dinner. You are waiting for confirmation from your friends for the after dinner.

User description: You are Roberto, and you really want to eat pizza. You want it too much!

Virtual Agent description: Vittorio is passionate about Chinese cuisine and loves eating oriental food with chopsticks.

Additional details: Remember: - You still don't know where you will go after dinner and you would rather not go too far. - You don't want to argue with your friend about food.

G.1.0.5 Introduction to the scenario if Compromising style is the most appropriate

Situation description: Two friends Vittorio and Roberto must decide where to go for dinner. You are waiting for confirmation from your friends for the after dinner.

User description: You are Roberto, and you really want to eat pizza. You want it too much!

Virtual Agent description: Vittorio is passionate about Chinese cuisine and loves eating oriental food with chopsticks.

Additional details: Remember: - You would like both of you to be happy, somehow. - There are many different options in the surrounding area.

G.2 Interaction

G.2.0.1 First interaction

Virtual agent: I am famished! I can't wait to go out! Will someone else come with us?

User: I am not so sure, I am famished as well though!

Virtual agent: The question is: pizza or Chinese?

Possible user choices:

1. Integrating: I'd like to go to the new pizzeria, I have been told that they don't make only pizza, would you like to see what they prepare?
2. Dominating: I have been told that there is a new pizzeria that makes a great pizza... Let's try it!
3. Obliging: I feel like pizza, but I know that you might prefer eating oriental food ... right?
4. Avoiding: We have just had lunch, how can we decide what we want tonight?
5. Compromising: I have been told that there is a new pizzeria that makes a great

pizza... We could eat oriental food next time!

Virtual agent responses:

1. I know it, and unfortunately they do not have Chinese cuisine ... I would really like to eat oriental food.
2. I really want Chinese tonight, we can try the new pizzeria in the week if you want, what do you think?
3. I miss oriental food, we could go in a pizzeria next time if you'd like, what do you think?
4. I would really love to eat at Chinese tonight, we could discuss this now, what do you think?
5. I have heard of this pizzeria, the pizza is very good actually, but this evening I would like to eat oriental food...

G.2.0.2 Second interaction

Possible user choices:

1. Integrating: Shall we look on the internet to see if there is a place that serves both pizza and oriental food?
2. Dominating: I find those chopsticks impossible to use! Better have pizza!
3. Obliging: If you really want Chinese, this time we will go there!
4. Avoiding: But we still do not know what to do afterwards... We should wait to decide!
5. Compromising: We could go for a Mexican this evening. It seems to be a good compromise!

Virtual agent responses:

1. I know one, just outside the centre. What do you say?
2. At the restaurant just outside the centre there is pizza and oriental food! No chopsticks for you! What do you think?
3. I am sorry if you eat something you don't want, the restaurant just outside the centre has pizza as well as oriental food. What do you say?
4. At the restaurant just outside the centre there is plenty of choice: pizza and oriental food, it could please us all. What do you say?
5. It could work... Or, at the restaurant just outside the centre there is both pizza and oriental food. What do you say?

G.2.0.3 Third interaction

Possible user choices:

1. Integrating: That would be great, but we do not have a car to get there. What can we do?
2. Dominating: We do not have the car tonight, and the pizzeria is the nearest from here.
3. Obliging: I know that you care and I would be pleased, but we do not have the car to get there...
4. Avoiding: And the car? Let's wait until we know what we are doing later, and we can talk about it again!
5. Compromising: We do not have the car unfortunately. Shall we go there the next week, and today we can go to the pizzeria?

Virtual agent responses:

1. I have an idea! There is a takeaway that would bring us both pizza and Chinese food. And then we could go out. What do you think?

2. I have a suggestion! There is a great takeaway that can bring us pizza and Chinese food at home for free! And then we can go out. What do you think?

3. I have an idea! There is a great takeaway that delivers pizza and Chinese food at home for free! What do you think?

4. I have an idea! There is a restaurant that has an excellent takeaway service; they can bring both pizza and Chinese food at home for free! What do you think?

5. I have an idea! This great takeaway would bring us a pizza and Chinese food at home for free, and then let's stroll around for a drink! What do you think?

G.2.0.4 Fourth interaction

Possible user choices:

1. Integrating: Great option! Before we decide, we could also have another look at the types of cuisines at the restaurants nearby!

2. Dominating: I don't feel like staying at home... With all the times that we had Chinese tonight you should try to please me!

3. Obliging: If this choice makes you happy, then it's fine for me! This evening I 'd like you to be happy in any case!

4. Avoiding: They might take too long for the delivery ... Let's watch TV for a bit and let's wait till we are hungry.

5. Compromising: And what if we wait too long? I would say that another type of cuisine tonight would be the best compromise!

Virtual agent responses:

1. Good idea, we could check on the Internet!

2. Let's go to the pizzeria, but let's make a deal: next time we eat oriental. Ok?

3. You are really nice. You'll see that this time you will love oriental food! And pizza the next time.

4. We could do, let's wait till we know what we are doing afterwards, but let's hope that it is not too late!

5. Yes, it is not a bad idea ... which restaurant are you thinking about in particular?

G.2.0.5 Final interaction

Possible user choices:

1. Integrating: I will discuss the possibility of finding a place nearby that prepares both cuisines with [VirtualAgent].

2. Dominating (if last choice was Dominating): As agreed with [VirtualAgent], we'll go for a pizza.

3. Dominating (if last choice was not Dominating): I will try to convince [VirtualAgent] to go to the pizzeria.

4. Obliging (if last choice was Obliging): As agreed with [VirtualAgent], we'll go for Chinese.

5. Obliging (if last choice was not Obliging): I will let [VirtualAgent] decide: we'll go for Chinese.

6. Avoiding (if last choice was Avoiding): As agreed with [VirtualAgent], we will decide after we know what we are doing afterwards.

7. Avoiding (if last choice was not Avoiding): I will try to postpone the decision until we know what we are doing.

8. Compromising: I will discuss the possibility of trying another type of cuisine that we could both like with [VirtualAgent].

Appendix H

Scenario 3: Obliging

H.1 Introductions

THIS scenario deals about two friends of the same gender who have to decide who is in charge of designing the logo of their sport team. The scenario options and details are reported as if the user gender was male.

H.1.0.1 Introduction to the scenario if Integrating style is the most appropriate

Situation description: To celebrate the transition to a higher series, the Plymouth team decided to redesign the logo and elect a person in charge of the design from among its members. You and your team-mate Giovanni received the same number of votes and you thought it would be a good idea for you to meet up to decide who should be in charge between the two of you.

User description: You are Paolo, a member of the Plymouth team and you have been voted together with Giovanni for the logo design. You are a pc and web wizard.

Virtual Agent description: Giovanni: Design expert, his assistance would be crucial in order to create a complete logo, but it seems that in this period he feels disheartened and discouraged.

Additional details: Remember: - Giovanni is really good, and feels very disheartened right now. - You are both skilful and together you would do a quality job.

H.1.0.2 Introduction to the scenario if Dominating style is the most appropriate

Situation description: To celebrate the transition to a higher series, the Plymouth team decided to redesign the logo and elect a person in charge of the design from among its members. You and your team-mate Giovanni received the same number of votes and you thought it would be a good idea for you to meet up to decide who should be in charge between the two of you. If you do not reach an agreement in a short time the coach will appoint an external designer.

User description: You are Paolo, a member of the Plymouth team and you have been voted together with Giovanni for the logo design. For you this role would be very important and you would also be interviewed by a broadcaster.

Virtual Agent description: Giovanni: He has an enviable experience in graphics, but he just had a long discussion with managers, and right now he is disheartened and discouraged, so that he has lost the motivation even for a simple task like restyling a logo.

Additional details: Remember: - Giovanni is really good, and feels very disheartened right now. - If you managed to take the role, you would get an important interview on TV.

H.1.0.3 Introduction to the scenario if Obliging style is the most appropriate

Situation description: To celebrate the transition to a higher series, the Plymouth team decided to redesign the logo and elect a person in charge of the design from among its members. You and your team-mate Giovanni received the same number of votes and you thought it would be a good idea for you to meet up to decide who should be in charge between the two of you.

User description: You are Paolo, a member of the Plymouth team and you have been voted together with Giovanni for the logo design. You would like the team to have a quality logo.

Virtual Agent description: Giovanni: Design expert, he is going through a difficult period and feels disheartened and discouraged, but if he was in charge of the design would have the opportunity to participate in an important graphic contest at national level.

Additional details: Remember: - Giovanni is really good, and feels very disheartened right now. - Giovanni is your dear friend, and you'd like to do everything you can to help him feel better.

H.1.0.4 Introduction to the scenario if Avoiding style is the most appropriate

Situation description: To celebrate the transition to a higher series, the Plymouth team decided to redesign the logo and elect a person in charge of the design from among its members. You and your team-mate Giovanni received the same number of votes and you thought it would be a good idea for you to meet up to decide who should be in charge between the two of you.

User description: You are Paolo, a member of the Plymouth team and you have been voted together with Giovanni for the logo design.

Virtual Agent description: Giovanni: He has an enviable experience in graphics, but he just had a long discussion with managers, and right now he is disheartened and discouraged, and feels inadequate even for a simple task like restyling a logo.

Additional details: Remember: - Giovanni is really good, and feels very disheartened right now. - You believe that Giovanni's problems are a priority compared to the logo decision.

H.1.0.5 Introduction to the scenario if Compromising style is the most appropriate

Situation description: To celebrate the transition to a higher series, the Plymouth team decided to redesign the logo and elect a person in charge of the design from among its members. You and your team-mate Giovanni received the same number of votes and you thought it would be a good idea for you to meet up to decide who should be in charge between the two of you. If you do not reach an agreement in a short time the

coach will appoint an external designer.

User description: You are Paolo, a member of the Plymouth team and you have been voted together with Giovanni for the logo design. You are a design expert, and this role would give you the opportunity to be interviewed by a broadcaster.

Virtual Agent description: Giovanni: as expert as you in design, this role would give him the opportunity to be interviewed by a broadcaster as well. He just had a long discussion with managers and is disheartened and discouraged right now.

Additional details: Remember: - Giovanni is really good, and feels very disheartened right now. - You believe that despite his problems, Giovanni should also contribute to the logo in equal parts.

H.2 Interaction

H.2.0.1 First interaction

Virtual agent: I really wanted this assignment! But now I am not sure I have the required focus...

User: It is a rather demanding job.

Virtual agent: Yes, and for me it would be a launch pad, but I don't know whether to take this responsibility ...

Possible user choices:

1. Integrating: If we cooperate and pool our ability together the result could be better!
2. Dominating: On the other hand, I am very motivated! It will be a nice challenge! I want this role to work on with commitment.
3. Obliging: I would like to have this role, but since it is so important to your career it is right for it to be assigned to you.
4. Avoiding: There is still so much time before the beginning of the championship,

let's talk about it after the exams ... Too much stress!

5. Compromising: We could divide the logo by inserting individually our two concepts. It would be creative!

Virtual agent responses:

1. With these negative thoughts I am afraid that I would only be in the way ... yet for me it would be a great opportunity ...

2. Yes, I think that you have a better thrust and approach ... but for me it would be a great opportunity ...

3. I feel so demoralised, probably you would do a better job ... However, it would be a great opportunity for me ...

4. I think it is better to decide now. For me it would be a great opportunity, but I know that you would be proud if you had the role.

5. With these negative thoughts I am afraid that I would only be in the way ... yet for me it would be a great opportunity ...

H.2.0.2 Second interaction

Possible user choices:

1. Integrating: We could think of a way to do it together. If we commit ourselves, we will find a solution.

2. Dominating: Your votes could be due to the fact that your father manages the funds of the school. Leave this assignment to me and you will feel better.

3. Obliging: Your father would be proud if you took this responsibility.

4. Avoiding: Do you know what? Let's let fate decide! Heads or tails?

5. Compromising: We received the same number of votes, and each of us should take 50% of the responsibility.

Virtual agent responses:

1. Perhaps you are right ... but I am afraid I was only voted because my father manages the funds of the school ...
2. Perhaps this influenced the votes ... it would be fairer for you to do it even if it's important to me ...
3. I really care about it, but it might be important for you as well...
4. Perhaps those who voted could be disappointed...
5. Perhaps I had votes only because my father manages the funds of the school ... Even if it means a lot to me would be fairer for you to do it...

H.2.0.3 Third interaction

Possible user choices:

1. Integrating: We could combine our forces and our skills to do a better job.
2. Dominating: You rest and recover, no worries! This assignment is perfect for me at this time.
3. Obliging: You look disheartened and I know that this logo is more important for you, perhaps you could find the motivation and take this responsibility.
4. Avoiding: Now we are too stressed from the exams, let's try some distraction with a game of PES!
5. Compromising: Let's give ourselves a week to work separately on the design elements. Would it be a good idea?

Virtual agent responses:

1. It would be nice, if you think that could work...
2. I would be delighted for you if you took this role ... maybe I could give you a hand if you want...
3. If you say so, perhaps I could try it ... But I do not want to wrong you...
4. In any case you always beat me at PES...
5. It would be nice, if you think that could work...

H.2.0.4 Fourth interaction

Possible user choices:

1. Integrating: I could focus more on the concept, while you could work on the graphic design. What do you think?
2. Dominating: I see that you still feel unsure about accepting this assignment, perhaps it would be better for me to do it myself!
3. Obliging: In fact, I don't feel like competing with you, we are friends ... I would be glad if you found the motivation if it is important for you.
4. Avoiding: It is the weekend, let's try to relax!
5. Compromising: We should both work the same number of hours on this logo.

Virtual agent responses:

1. I wouldn't know ... If I could get the motivation I could be a good team mate ...
2. You convinced me... maybe you are the person who is best suited to the task.
3. If you say so, then I could really do it myself...
4. Yes, maybe you are right, a bit of relaxation is what we need...
5. Actually you have proposed a good solution ...

H.2.0.5 Final interaction

Possible user choices:

1. Integrating: I will discuss the possibility of working together for the logo, dividing the roles based on skills with [VirtualAgent].
2. Dominating (if last choice was Dominating): As agreed with [VirtualAgent], I will take charge of the work.
3. Dominating (if last choice was not Dominating): I will try to take the assignment.
4. Obliging (if last choice was Obliging): As agreed with [VirtualAgent], he will deal with the logo.
5. Obliging (if last choice was not Obliging): I will let [VirtualAgent] deal with the logo.
6. Avoiding (if last choice was Avoiding): As agreed with [VirtualAgent], we will decide some other time.
7. Avoiding (if last choice was not Avoiding): I will try to postpone the decision until later.
8. Compromising: I will discuss the possibility of sharing the work in equal parts with [VirtualAgent].

Appendix I

Scenario 4: Avoiding

I.1 Introductions

THIS scenario deals about two friends of the same gender who have to decide who will take the last copy of their favourite CD which is available in the shop. The scenario options and details are reported as if the user gender was male.

I.1.0.1 Introduction to the scenario if Integrating style is the most appropriate

Situation description: It is morning, and for half an hour you have been with Michele waiting for your two copies of the CD of the concert of your favourite band, but on the shelf there is only one CD booked in your name; in addition, the shop is now really crowded.

User description: You are Gabriele, and you have been anxiously waiting for a month for the arrival of the autographed CD with the concert of your favourite band you went to see with Michele. You are a hardened collector of autographs and special editions.

Virtual Agent description: Michele: your flatmate. He has a habit of buying a CD and listening to it for a week to the point of exhaustion.

Additional details: Remember: - The autograph on the cover represents a real trophy for you. - Michael can't wait to listen to the CD for hours.

I.1.0.2 Introduction to the scenario if Dominating style is the most appropriate

Situation description: It is morning, and for half an hour you have been with Michele waiting for your two copies of the CD of the concert of your favourite band, but on the

shelf there is only one CD booked in your name; in addition, the shop is now really crowded.

User description: You are Gabriele, and you have been anxiously waiting for a month for the arrival of the autographed CD with the concert of your favourite band you went to see with Michele. The following day you will be leaving for one month abroad and you don't want to completely give up being able to hear the CD during the journey.

Virtual Agent description: Michele: your flatmate, apparently wants the CD as much as you do but you know that his CD player has been broken for some time.

Additional details: Remember: - Michele's CD player is broken. - You don't want to go for a month without your favourite CD.

I.1.0.3 Introduction to the scenario if Obliging style is the most appropriate

Situation description: It is morning, and for half an hour you have been with Michele waiting for your two copies of the CD of the concert of your favourite band, but on the shelf there is only one CD booked in your name; in addition, the shop is now really crowded.

User description: You are Gabriele, and you have been anxiously waiting for a month for the arrival of the CD with the concert of your favourite band you went to see with Michele.

Virtual Agent description: Michele: your flatmate. The following day he will be leaving for a year of volunteering in South Africa, he feels stressed by this event, and he would really like to bring this CD with him.

Additional details: Remember: - Michele will be leaving for South Africa and cannot wait for the arrival of another copy of the CD. - A new CD order could arrive at the shop.

I.1.0.4 Introduction to the scenario if Avoiding style is the most appropriate

Situation description: It is morning, and for half an hour you have been with Michele waiting for your two copies of the CD of the concert of your favourite band, but on the shelf there is only one CD booked in your name; in addition, the shop is now really crowded.

User description: You are Gabriele, and you have been anxiously waiting for a month for the arrival of the CD with the concert of your favourite band you went to see with Michele. You just got a call from your girlfriend; she landed a few moments ago at the airport and is waiting for you there.

Virtual Agent description: Michele: your flatmate, wants to get the CD as much as you do and he is a hard-core fan.

Additional details: Remember: - You hate to queue at shops. - You are hungry like the wolf. - Your girlfriend is waiting for you at the airport very soon.

I.1.0.5 Introduction to the scenario if Compromising style is the most appropriate

Situation description: It is morning, and for half an hour you have been with Michele waiting for your two copies of the CD of the concert of your favourite band, but on the shelf there is only one CD booked in your name; in addition, the shop is now really crowded. You don't want to give up the booking or you will lose the opportunity to have the limited edition.

User description: You are Gabriele, and you have been anxiously waiting for a month for the arrival of the CD with the concert of your favourite band you went to see with Michele. At home you have only one stereo in the lounge, and you try to take turns to use it.

Virtual Agent description: Michele: your flatmate, wants to get the CD as much as you do and he is a hard-core fan.

Additional details: Remember: - You only have one stereo to listen to your music. - The CD is a unique limited edition. - You would like to find the best solution without too many sacrifices.

I.2 Interaction

I.2.0.1 First interaction

Virtual agent: Crazy desire for that CD, but really bad organization! I am very disappointed!

User: Yes... Only one CD is left in spite of the reservation.

Virtual agent: How annoying! Who knows who they gave it to!

Possible user choices:

1. Integrating: Let's try to find a solution together that is good for both of us without losing our tempers!
2. Dominating: However, I want that CD absolutely at home today, I can't let it get away!
3. Obliging: Let's try not to get upset... Do you have any idea what to do? You always have great ideas.
4. Avoiding: So much stress for a CD! It is a decision we need to discuss with more time and calm.
5. Compromising: So let's try to find a compromise: well pleased is well served!

Virtual agent responses:

1. It is not easy! I am so unlucky, things never go well not even once!
2. I want it too! But I can't believe the bad luck that we got ...
3. I wouldn't know ... I 'm thinking that in this period I am terribly unlucky!

4. Time? Calm? It would take a bit of luck! In this period I have none really!

5. Before we find a compromise, I should ask my misfortune to stop persecuting me!

I.2.0.2 Second interaction

Possible user choices:

1. Integrating: Let's try not to think of bad luck and try to find a solution to our common CD.

2. Dominating: I am not thinking of bad luck, I'm thinking about how I can overcome this queue and get that CD!

3. Obliging: No bad luck! I know you really want the CD, and you should take it...

4. Avoiding: Maybe you are lucky in love! And then tonight we're going out!

5. Compromising: Let's decide quickly how to share this CD, it is the only thing to do.

Virtual agent responses:

1. Search for a solution? And what if we went to the bar? After a snack we'll see!

2. Instead of jumping the queue we are skipping the snack. Let's go to the bar and let's think about it afterwards!

3. Let's rather defer the decision until after the snack! Shall we go to the bar?

4. You are right! Shall we go to the bar? Sweets put me in good spirits ...

5. Instead of sharing the CD let's share a snack. Let's go to the bar and let's think about it afterwards!

I.2.0.3 Third interaction

Possible user choices:

1. Integrating: How about for now we focus our efforts to get through this queue and find the best solution?
2. Dominating: I want the CD immediately, you never know! We can have a snack some other time.
3. Obliging: No need to distract me, don't worry! I would be happy if you take it the CD!
4. Avoiding: Indeed, it would be good to have a break, and then decide calmly. And what shall we do about the queue?
5. Compromising: You could go to get a snack from the vending machine, so we don't lose our place in the queue.

Virtual agent responses:

1. Now I am too disappointed, let's leave the queue and think about it later after the bar.
2. The idea of not having a CD all to myself makes me feel bad, let's think about how to share it later!
3. It is more to distract me from this disappointment! Never mind the queue, shall we go to the bar?
4. I am too disappointed to queue, let's think about it calmly!
5. I am too disappointed to talk about it, let's leave the queue and decide later!

I.2.0.4 Fourth interaction

Possible user choices:

1. Integrating: I have an idea! We could buy the CD. I love autographs, I would take the cover and you could have the CD!
2. Dominating: You go to the bar if you want to. I will buy the CD myself and I will keep it!
3. Obliging: You can take the CD, I still have a new one to listen to!
4. Avoiding: Ok, let's not think about it now. After the bar our minds will be freer.
5. Compromising: Let's split the payment and buy the CD now! We can leave it in the house so anyone who wants to can listen to it.

Virtual agent responses:

1. I don't know ... I also need a fresh mind to decide ...
2. Yes, go ahead and buy it ... right now only a snack will make me feel better!
3. Thank You! Now we can think of our snack!
4. Ok, well, let's go to the bar!
5. I don't know ... I also need a fresh mind to understand how to share it...

I.2.0.5 Final interaction

Possible user choices:

1. Integrating: I will discuss the possibility of sharing the CD on the basis of our interests with [VirtualAgent].
2. Dominating (if last choice was Dominating): As agreed with [VirtualAgent], I will take the CD.
3. Dominating (if last choice was not Dominating): I will try to take the CD.
4. Obliging (if last choice was Obliging): As agreed with [VirtualAgent], he will take the CD.

5. Obliging (if last choice was not Obliging): I will let [VirtualAgent] take the CD.

6. Avoiding (if last choice was Avoiding): As agreed with [VirtualAgent], the decision will be taken after a break at the bar.

7. Avoiding (if last choice was not Avoiding): I will try to postpone the decision until after the break at the bar.

8. Compromising: I will discuss with [VirtualAgent] of the possibility of leaving the CD in the house, and use it as required.

Appendix J

Scenario 5: Dominating

J.1 Introductions

THIS scenario deals about two partners of the opposite gender who have to decide what to watch on TV. One prefers the show Masterchef, while the other would like to watch Report. The scenario options and details are reported as if the user gender was male.

J.1.0.1 Introduction to the scenario if Integrating style is the most appropriate

Situation description: You live with your wife Anna in a home where there is only one TV in the living room. You must decide what to watch that evening.

User description: You are Luca, a post office employee. You love watching Report because it deals with topics close to your own experiences and you find it extremely interesting.

Virtual Agent description: Anna: your wife, loves Masterchef. She just got back from work, and right now she feels particularly unhappy because of a bad phone call with a manager at work.

Additional details: Remember: - Anna is very unhappy because of work. - You would like neither of you to give up your favourite program.

J.1.0.2 Introduction to the scenario if Dominating style is the most appropriate

Situation description: You live with your wife Anna in a home where there is only one TV in the living room. You must decide what to watch that evening.

User description: You are Luca, a post office employee. You love watching Report because it deals with topics close to your own experiences and you find it extremely interesting. That night Report will have a very important report on a case you are following at work. You know that Masterchef is rebroadcast the following day.

Virtual Agent description: Anna: your wife, loves Masterchef. She just got back from work, and right now she feels particularly unhappy because of a bad phone call with a manager at work.

Additional details: Remember: - Anna is very unhappy because of work. - Masterchef will be rebroadcast. - Report will cover extremely important issues for you that evening.

J.1.0.3 Introduction to the scenario if Obliging style is the most appropriate

Situation description: You live with your wife Anna in a home where there is only one TV in the living room. You must decide what to watch that evening.

User description: You are Luca, a post office employee. You love watching Report because it deals with topics close to your own experiences and you find it extremely interesting.

Virtual Agent description: Anna: your wife, loves Masterchef. She just got back from work, and right now she feels particularly unhappy because of a bad phone call with a manager at work. For her, after dinner TV is an important and relaxing moment, and she particularly likes sharing it with you.

Additional details: Remember: - Anna is very unhappy because of work. - For you your wife's relaxation is just as important as yours.

J.1.0.4 Introduction to the scenario if Avoiding style is the most appropriate

Situation description: You live with your wife Anna in a home where there is only one TV in the living room. You must decide what to watch that evening.

User description: You are Luca, a post office employee. You love watching Report because it deals with topics close to your own experiences and you find it extremely

interesting. You need for the TV to be a moment of relaxation without any discussion.

Virtual Agent description: Anna: your wife, loves Masterchef. She just got back from work, and right now she feels particularly unhappy because of a bad phone call with a manager at work.

Additional details: Remember: - Anna is very unhappy because of work. - You are really stressed and you need to relax.

J.1.0.5 Introduction to the scenario if Compromising style is the most appropriate

Situation description: You live with your wife Anna in a home where there is only one TV in the living room. You must decide what to watch that evening. Both your favourite programs will be rerun the following day.

User description: You are Luca, a post office employee. You love watching Report because it deals with topics close to your own experiences and you find it extremely interesting.

Virtual Agent description: Anna: your wife, loves Masterchef. She just got back from work, and right now she feels particularly unhappy because of a bad phone call with a manager at work.

Additional details: Remember: - Anna is very unhappy because of work. - Both Masterchef and Report will be rebroadcast.

J.2 Interaction

J.2.0.1 First interaction

Virtual agent: What bad luck! Tonight both Masterchef and Report are on TV!

User: Which program shall we watch?

Virtual agent: Simple, we watch Masterchef, no discussion!

Possible user choices:

1. Integrating: What do you think of finding a solution that makes us both happy instead?
2. Dominating: Let's not even talk about it. For me Report is important tonight.
3. Obliging: I really love watching Report, but I believe that you are happier watching Masterchef...
4. Avoiding: You seem stressed has something happened at work?
5. Compromising: Let's watch a bit of each program, so we do not argue?

Virtual agent responses:

1. You will also be happy to watch Masterchef!
2. For you, Report is important, for me, Masterchef is important!
3. I don't want to hear another word about Report!
4. Always the same arguments! I will check what time Masterchef starts; I really want to watch it.
5. And how? No way! Don't even go there.

J.2.0.2 Second interaction

Possible user choices:

1. Integrating: Shall we try to watch other programs together that we could both like?
2. Dominating: I want to watch Report; it will deal with issues that are important for my work.
3. Obliging: I understand that it means a lot to you, and we can start watching Masterchef as long as you want, then when the advertising comes on we'll change the channel for a few minutes!

4. Avoiding: But why did they schedule the two programs at the same time?

5. Compromising: We could watch Masterchef, but every now and then change the channel and see if they are covering issues that are of interest to me.

Virtual agent responses:

1. If it is not Masterchef I don't want to watch it! And anyway, Report will be rebroadcast!

2. Tonight I am stressed and I want to watch Masterchef to relax. Report is on again tomorrow!

3. You are starting to use your brain! But we won't change channel, anyway Report is on again tomorrow!

4. There is only a problem because you don't want to watch the rerun of Report without arguing tonight!

5. How stressful! We would leave something half done!

J.2.0.3 Third interaction

Possible user choices:

1. Integrating: I would like to see some topics on Report... Was there a particular reason why you are interested in Masterchef tonight?

2. Dominating: They are rebroadcasting Masterchef as well, and I need to watch Report tonight.

3. Obliging: If Masterchef is so important for you, I could watch the rerun of Report ...

4. Avoiding: Blame these broadcasters that make us argue!

5. Compromising: Is there a way to stay in the same room and watch both programs?

Virtual agent responses:

1. It was a tough day and Masterchef helps me relax!
2. You want to argue huh? Tell the truth! I had a hard day and don't want to argue anymore!
3. You guessed it! Tonight I am watching Masterchef!
4. But I don't want to argue, I want to watch Masterchef!
5. Watch Masterchef with me then! We can put off Report until at least tomorrow when it is rebroadcast!

J.2.0.4 Fourth interaction

Possible user choices:

1. Integrating: If some parts of Masterchef are of particular interest to you, we can see if they fit with the Report topics that I'd like to follow.
2. Dominating: I want to watch Report, the more nervous you get, the less I will relent! We can argue as long as you like!
3. Obliging: I can't stand seeing you so stressed, Masterchef tonight! You are my priority.
4. Avoiding: Right now I need a couple of hours to take care of some documents, let's talk about it later!
5. Compromising: It seems to me that the only compromise is to watch both programs, for half the time each.

Virtual agent responses:

1. It is an overly complicated solution, tonight I don't want to stress out!
2. So we have to argue! You wanted it this way!
3. Thank You! I'm glad you understood!
4. It would be better if you are in another room, so maybe I can relax a little.
5. You are thinking of overly complicated solutions, tonight I don't want to stress out!

J.2.0.5 Final interaction

Possible user choices:

1. Integrating: I will try again to propose to see the most interesting parts of both programs to [VirtualAgent].
2. Dominating (if last choice was Dominating): I will be staying strong on my position: I want to watch Report.
3. Dominating (if last choice was not Dominating): I will try to watch Report.
4. Obliging (if last choice was Obliging): As agreed with [VirtualAgent], we'll watch Masterchef.
5. Obliging (if last choice was not Obliging): I will let [VirtualAgent] decide: we'll watch Masterchef.
6. Avoiding (if last choice was Avoiding): As agreed with [VirtualAgent], the decision will be taken after I've finished taking care of the documents.
7. Avoiding (if last choice was not Avoiding): I will try to postpone the decision until after I've finished taking care of the documents.
8. Compromising: I will try again to propose watching both programs, each half of the time, to [VirtualAgent].

Appendix K

Motivation and pleasantness questionnaire

Thank you for your participation.

1. Complete the following statements by choosing a single alternative from those present. You have just completed the tests, playing with Enact and responding to Roci II. By comparing the two tools, you believe that

	Enact*	Roci II*	I don't know
You took more time to finish	◯	◯	◯
You found more interesting	◯	◯	◯
You were more keen on completing	◯	◯	◯
You paid more attention completing	◯	◯	◯

* We remind you that the ROCI-II is the multiple choice test, while Enact is the virtual game.

2. Complete the following statements by marking a single alternative among the 5 present, expressing the degree of judgment that is closest to your perception of the two tests.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
I think it took me too much time to finish the Enact game, compared to the Roci II	○	○	○	○	○
I felt bored completing the Roci II test compared to the Enact game	○	○	○	○	○
I felt more pleased completing Roci II questions than those of the Enact game	○	○	○	○	○
I got distracted several times while completing the Roci II compared to the Enact game.	○	○	○	○	○

* We remind you that the ROCI-II is the multiple choice test, while Enact is the virtual game.

3. Complete the following statements by marking a single alternative among the 5 present, expressing the degree of judgment that is closest to your perception of the two tests.

	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
While completing the Enact game, I gained more knowledge concerning my ability to handle conflicts	◯	◯	◯	◯	◯
I am motivated to further deepen this new knowledge in Enact (concerning my ability to manage conflicts)	◯	◯	◯	◯	◯
While completing the Roci II test, I learned more information concerning my ability to handle conflicts	◯	◯	◯	◯	◯
I am motivated to further deepen this new knowledge in the ROCI-II (concerning my ability to manage conflicts)	◯	◯	◯	◯	◯

4. How long you think it took you to complete:

	Less than 10 minutes	Between 10 and 15 minutes	More than 15 minutes
ENACT*	◯	◯	◯
Roci II*	◯	◯	◯

* We remind you that the ROCI-II is the multiple choice test, while ENACT is the virtual game.

Demographic information

To conclude, answer the following questions by choosing one and only one of the options provided for each question, indicating the one that is closest to the description of your person. We remind you that the questionnaire is anonymous.

1. Year of birth _____

2. Employment

- o Student
- o Unemployed
- o Working student
- o Non-working student
- o other (specify) _____

3. If you work, what kind of work are you currently doing?

- o I work in contact with people
- o I work in a team or group
- o I work without requiring to have contact with people

4. Qualification

- o Middle school diploma
- o High school diploma
- o Other post-diploma qualification
- o Bachelor's degree _____
- o Master's degree _____
- o Other (specify) _____

5. Are you enrolled in a graduation course?

- o Yes, a Bachelor (specify which course) _____
- o Yes, a Masters (specify which course) _____
- o No
- o Other (specify) _____

6. Marital Status

- ◯ Unmarried
- ◯ Married
- ◯ Other

7. Current accommodation

- ◯ I live with my family
- ◯ I live alone
- ◯ I live with my partner
- ◯ I live with friends/students
- ◯ Other (specify) _____

8. Accommodation

- ◯ Rented room or apartment
- ◯ Own property
- ◯ Other (specify) _____

9. Nationality

_____

Appendix L

Research conducted within and outside of the ENACT Project

In detail, the research which has been conducted in cooperation with the project partners is:

- The Training Need Analysis conducted prior to the development of the software to investigate the needs of our target group;
- The research conducted upon the first version of the scenarios depicting real life situations as part of the Assessment environment of the Enact platform. In particular the aim was to create scenarios that would reflect situations that can be experienced at any age, and that would trigger motivation to continue the negotiation;
- The research conducted upon the implementation of the backend and database infrastructure to contain the user profiling data, the scenarios and the experiments information of the Enact platform, with a focus on making the database accessible and easy to interact with across multiple platforms;
- The implementation of the interface and user environment with adaptive and gamification design principles for the human-computer interaction;
- The research conducted to create 3D bot gestures and animations that could accurately reflect each negotiation style involved in the theory, including the variables about gender, ethnicity and age of the virtual characters;

- The first validation of the Enact Assessment, whereas each of the project partners recruited a specific sample of participants;
- The second validation of the Enact Assessment, whereas each of the project partners recruited a specific sample of participants.

What has been investigated exclusively within the scope of this work, and independently from the ENACT framework is:

- The testing, evaluation, rewriting and improvement of the scenarios of the Enact platform for the second release;
- The research upon design of the Training environment and tutor feedback of the Enact platform;
- The development of the Intelligent Tutoring System included within the Enact platform;
- The research about the attractiveness, user-friendliness, pleasantness and usability of the first interface of the platform;
- The research about the attractiveness, user-friendliness, pleasantness and usability of the second interface of the platform;
- The study about the evaluation of the implicit pleasantness, measured through the perception of time, of Enact compared to standard psychological negotiation tests;
- The study about the evaluation of the explicit pleasantness of Enact compared to standard psychological negotiation tests;
- The data collection of natural sentences during the interactions between users and virtual agents;
- The natural language analysis of the utterances of the users across different negotiation styles;

- A machine learning system for the automatic classification of negotiation styles using user interactions in natural language.

Nomenclature

ENACT The Project that embraces the present work, Enhancing Negotiation skills through on-line Assessment of Competencies and interactive mobile Training

Enact The name of the game platform developed within the scope of the ENACT Project

ICT Information-Communication Technology

ITS Intelligent Tutoring System

MOOC Massive Open Online Course

ROCI – II Rahim Organizational Conflict Inventory

SME Small-Medium Enterprises

TKI Thomas-Killman Inventory

TNA Training Need Analysis

Index