CLOSING THE GAP

DESIGNING COMPUTER GAMES TO REDUCE THE DIFFERENCE IN HOW PLAYERS TREAT THEIR HUMAN AND THEIR AI TEAM-MATES

CHRISTOPHER ONG ENG HWA

NATIONAL UNIVERSITY OF SINGAPORE

2014

CLOSING THE GAP

DESIGNING COMPUTER GAMES TO REDUCE THE DIFFERENCE IN HOW PLAYERS TREAT THEIR HUMAN AND THEIR AI TEAM-MATES

CHRISTOPHER ONG ENG HWA

(B.Social Sci. (Hons), NUS, 2009)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF ARTS

COMMUNICATIONS & NEW MEDIA

NATIONAL UNIVERSITY OF SINGAPORE

2014

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety.

I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Christopher Ong Eng Hwa
June 3, 2014

Acknowledgment

While a thesis such as this one usually is credited to one individual, the reality is that without the efforts and inputs of many others, such an undertaking would be much more difficult than it already is. As such, I would like to express my gratitude and appreciation to the following individuals and groups (in no particular order):

Associate Professor Kevin McGee – for being the best possible supervisor I could have had. Without your guidance (and your pushing me to work on one experimental study per semester), I would not have been able to have those conference paper publications to my name, let alone complete this thesis sufficiently. Using "The Funnel" as a means of organizing the structure of my papers will be something I will always remember, as well as the lab meetings of the **ParTech Lab**.

My lab mates **Assistant Professor Alex Mitchell, Prashanth, Joshua Wong, You Yingdan, Jing Chiang, Evelyn Chew, Tiffany Neo, Cheong Kakit, Chuah Teong Leong and Assistant Professor Tim Merritt** for sitting through endless revisions of the ideas behind this thesis, and providing much valued feedback and insights that helped refine it (despite listening to it ad nauseum). Special mention must go to **Teong Leong** who helped develop the game Defend the Pass and **Tim**, whose work on the Cooperative Attribution Framework provided the foundation for my own.

My colleagues from the **Department of Communications of New Media, NUS**, for their support provided and the opportunity to do this thesis. I am especially grateful to **Ms Retna** for all the assistance and advice she has given in helping me get this thesis completed (especially with all the forms and instructions).

The **lecturers and students of NM2102 and NM2103** over the various semesters – **Associate Professor Lim Sun Sun, Assistant Professors Julian Lin, Zhang Weiyu and Leanne Chang** for entrusting your students to me as participants for the studies that have appeared in this thesis. To the **students from NM2102 and NM2103** (many of whom are graduating with me this year), thank you for taking part in my studies, and helping me add to the body of knowledge in this field of research I was exploring. Without you, there would be no data (and no thesis).

Nabil, Anna Fillipova and Stephanie Lee, my "Holy Trinity of

Statistics Gods". Thank you for all the help you have given me in helping understand and interpret the results of my studies, and putting up with my strange questions about statistical methods at the most unusual of hours of the day (and night).

The other friends of mine like my cell group mates **Sheryl Ang & Ong Chin Siang, Paul Yew, Jesher Tan, Calvin Fong, Sherilyn Tan, Glenn Wong and Gabriel Xu**, the various team members of the different batches of **Random Blends**, **CNM majors from the Class of 2009 to Class of 2014**, and members of **NUSPsyche** (there's too many of you to mention individually), thank you for the support as you've given me over the course of this thesis. Though you may feel that you may not have done much, the words of encouragement and asking me how things were going are much appreciated in helping me keep at this process.

My family – **Jon, Ter, Dan, Bea** for their support and especially **Mum** for the constant reminders and nagging about getting the thesis finished. I know you probably still don't quite understand what the topic is about, but the nagging and prodding (much as I was annoyed by it) served to remind me to stay the course and get this finished.

Lynette, who I love dearly and has been my pillar of support throughout this process (especially when things got more tedious and annoying). For the emotional support, being around for me to bounce ideas (especially about the calculations) off, and the proof-reading (all in one evening too) helped make the writing of this thesis much easier than it could have been. You are possibly one of the best things that has happened to me, and I am forever grateful for that.

Finally (and most importantly) I would like to thank **God**, to whom I am grateful for the chance of being able to make it this far in my life, which has been possible through what you have made possible for me. This completed thesis is perhaps a testament to the fact that (small) miracles are possible – that an epileptic with brain damage (and thought to have a less-than 1% chance of being what I am now) could do a Masters' thesis centered around computer games.

To the reader of this thesis, I hope you find the material here useful in furthering your own quest for knowledge to push the field further, and always remember that it is with the support of those around you that makes that much easier to accomplish.

June 3, 2014

Contents

Name	:	Christopher Ong Eng Hwa
Degree	:	Master of Arts
Supervisor(s)	:	Associate Professor Kevin McGee
Department	:	Communications & New Media
Thesis Title	:	CLOSING THE GAP Designing Computer Games to Reduce the Difference in How Players Treat Their Human and Their AI Team-Mates

Abstract

Research has shown that there are significant differences between the way humans behave towards (or in the presence of) and feel about human team-mates – compared to the way humans behave towards (or in the presence of) and feel about computer team-mates. This difference has important consequences for the design of multiplayer games. What can be done to reduce the difference in how people respond towards human and computer team-mates in multiplayer games?

There has been research examining how the behavior towards and perception of either human or AI team-mates can be moderated by such manipulations as the use of of team mates-stereotypes, representations of the systems used to interact, and the anthromorphism of the team-mate's appearance.

However, few studies that look at moderating the difference in how people respond to human and AI team-mates. This thesis addresses the following research question: will differences in the display of *scoring* information and differences in player *interdependence* moderate behavioral and emotional differences in in real-time cooperative games with human and AI team-mates?

Two studies were conducted to see whether specific design changes would moderate behavioral and emotional differences for the two types of team-mates. Each study used a two-by-two configuration (two versions of game and two types of team-mates), with a minimum of 60 participants playing a team-mate game; each participant played two rounds of several games, half the time playing with a human team-mate and half the time playing with an AI team-mate. Data gathering included game logs and self-reported answers to questionnaires.

One of the studies involved changes to the display of *scoring* information in a game Defend the Pass that required players to choose whether or not to sacrifice their team-mate. The other study involved changes to team-mate *interdependence* in a game Return the Ball; in one version of the game participants chose whether they or their team-mate got to play – in the other version, participants chose between two positions/roles which allowed both team-mates interdependently contribute to the outcome.

The experiments' results showed that manipulations of scoring information and manipulation of interdependence between team members were able to moderate behavior towards human and AI team-mates (through the frequency of choices made). However, for the difficulty in making those choices (i.e. emotional difficulty), manipulations of scoring information were unable to moderate emotional difficulty between human and AI team-mates, while manipulation of interdependence between team members was able to. This is because the manipulation of interdependence between team members eliminated the trade-off situation in the Return the Ball study, while the manipulation of scoring in the study Defend the Pass study did not eliminate the trade-off situation in that game. The results of these studies provide designers of cooperative games with three insights if they want to ensure that players will have the same type of experience with human and AI team-mates –(1) both behavior and emotional difficulty must be addressed; (2) trade-offs must be addressed and (3) caution is necessary when manipulating interdependence between team members.

Keywords : Team-mate Games

List of Tables

List of Figures

Chapter 1

Introduction

> This chapter establishes the context of the thesis – people treat human team-mates differently from AI team-mates in real-time cooperative games – and poses the question of what can be done to reduce this difference. It ends with an outline summary of the chapters of the thesis document.

1.1 Background & Motivation

Games such as *Defense of the Ancients (DotA) 2* and *Team Fortress 2* often involve multiple players cooperating as part of a team in order to achieve a common goal. These games are designed with this multiplayer cooperative element as one of the main features, and require groups of players playing over a local network, or remotely via the Internet. However, certain circumstances such as network problems may result in an insufficient number of human players to form teams. Erik Johson, project lead for *DotA 2* acknowledges this problem, saying the the goal of his project team for the AI bots in *DotA 2* is as follows [1]:

> Our goal with the AI is just that their [the player's] experience isn't destroyed just because one person couldn't finish the game.

It is clear from this example that game developers see the use of AI team-mates controlled by the computer as a solution to the problem of missing human team-mates. While the developers' goal is to ensure that the player's experience does not abruptly end just because not all players were able to complete the game, it is not necessarily true that playing with the AI team-mate / opponent which substituted for an unavailable human one would give players the same quality of experience. In order to understand whether the quality of player's experience is maintained before and after such a substitution occurs, it is necessary to understand how people respond towards human and computer team-mates, of which there are two main approaches.

The Social Responses to Communication Technology (SRCT) approach [18] examines how people respond towards computers and other media. They term the interaction between humans using computers as computer-mediated communication (CMC), and the interaction between humans and computers as human-computer interaction (HCI). This approach encompasses the Media Equation [25] and the Computers Are Social Actors (CASA) paradigm [19], and makes the following claim [18]:

> A summary of the SRCT viewpoint is "human-human interaction equals human-computer interaction". The word equals has a weak meaning – the same general patterns between HCI and CMC – and a strong meaning – identical results between HCI and CMC.

The "strong" meaning of equality according to SRCT is:

> To examine equality in the strongest possible sense, one must have an experiment in which the situation, the procedures, and the measures are identical for those participants who believe they are interacting with a computer and those participants who believe they are interacting with a person (i.e., for experimental purposes, human-human interaction must be construed as computer-mediated communication between two humans).

Studies that have used the SRCT approach to examine CMC versus HCI with team-mates has found that in terms of the SRCT's strong meaning of "identical results between HCI and CMC", people actually respond more favorably towards humans than computers both behaviorally and socially. SRCT researchers have attempted to explain this using the "Black Sheep Effect", as suggested in Johnson & Gardner [6]:

> ...[people] treat the computer like a person to the extent that they perceive the computer as a member of the ingroup, albeit a member of the group that does not contribute positively to their group identification.

An alternative means of explaining this difference in responses towards human and AI team-mates is provided by the Cooperative Attribution Framework [17]. This framework seeks to explain why people respond more favorably towards human team-mates than AI team-mates.

Examples of behavioral and social responses where people respond more favorably towards human team-mates than AI team-mates encompass attributes such as the sense of flow and enjoyment [14, 24, 31], risk-taking [15], assigning credit and blame [16], and sacrificing team-mates [17]. In fact, comparative studies in the SRCT approach comparing people's responses to social science phenomena in CMC and HCI have found results [6, 18] that are consistent with this alternative approach.

These results suggest that if game developers were to use AI team-mates as substitutes for unavailable human team-mates, a quality of a player's game experience would be compromised, as shown by Rajava et al [24] and Weibel et al[31]. It thus becomes clear that AI team-mates are not perfect substitutes for missing human team-mates, and that despite using the AI team-mates as a means of ensuring the player's experience is not destroyed, the quality of experience will compromised, due to the difference in how people respond towards human and computer team-mates.

It is clear that without addressing this second problem – the difference in how people respond towards human and computers, the solution of substituting AI team-mates for missing human ones does not completely solve the original problem. This then presents an interesting opportunity to address the following question in this thesis: What can be done to reduce the difference in behavioral and social responses towards human and AI team-mates in multiplayer games?

1.2 Thesis outline

The remainder of this thesis document is structured as follows:

Related Work A review of previous work that examined what can be done to change people's behavior towards and perceptions of AI interactants or team-mates in a variety of computer-mediated environments and contexts.

Research Problem An articulation of the research gap, the main research problem, and the original contribution of the thesis work.

Methodology A brief summary of the methodology and study design of the two studies in this thesis.

Theoretical Framework An explanation of theoretical framework(s) used as a reference point to analyze participant behavior towards human and AI team-mates in cooperative games.

Study 01: Changes to Score Display A study examining how the modification of available game-state information influences behavior towards and perception of human and AI team-mates.

Study 02: Changes to Interdependence A study examining how a modification of the level of interdependence between team-mates influences perception of and behavior towards human and AI team-mates.

Discussion An overall discussion comparing the findings of the two studies, and the limitations of the studies.

Conclusion A summary the findings, as well as possible directions for future research.

Chapter 2

Related Work

This chapter provides a review of previous work on moderating the differences in responses people have towards others in CMC and HCI scenarios – and moderating responses towards others in either CMC or HCI contexts. The results of this research show that manipulations such as visibility of scoring information, level of behavioral realism of the AI agent, whether a participants' views are made public or private, and whether humor is utilized by other interactants in the task can moderate differences in responses between CMC and HCI. Also, when examining either CMC or HCI scenarios, alignment of team-mate ability and task appropriateness, representation of interface elements or level of anthropomorphism are manipulations that can moderate the difference in responses towards other interactants.

2.1 Moderating differences in responses between CMC and HCI

Studies have shown that differences in how responses towards human and AI partners can be moderated by elements such as the visibility of scoring information, level of behavioral realism of the AI agent, whether a participants' views are made public or private, and whether humor is utilized by other interactants in the task.

A study was conducted by Ong et al [22] that examined how the use of different representations of game state information moderated the difference in how people responded towards human and AI team-mates in a real-time cooperative game. When game state information such as the team's cumulative total score was visible, participants chose to protect the human team-mate more than the AI team-mate. However, when this information was not made visible, differences in behavior towards the human and AI team-mates were not statistically significant, suggesting both types of team-mates were treated similarly when scoring information was not made visible. However, this variation in visibility of information did not moderate the difference in self-reported emotional difficulty felt when choosing one's position with a presumed human team-mate compared to the AI team-mate. In other words, in both versions (scoring information visible and not visible) participants reported having more emotional difficulty in choosing their position with their presumed human team-mate than their AI team-mate.

Blascovich's Threshold Model of Social Influence [2] seeks to provide a method on to how to moderate the difference in social responses towards computers such that it is more similar to social responses towards humans. In the Threshold Model, in order for computer agents to elicit social responses from users the same way that other humans (represented by avatars) do, a threshold of social verification (i.e co-presence or social presence) must be crossed. The model states that the social verification is a function of agency (whether the other interactant the user is exposed to is a human or a computer) and behavioral realism (how believable the behavior is). The model assumes that if the other interactant is human, then agency will be high, but if it is a computer, it will always be low. Thus, to moderate the level of social verification that a user has with a computer interactant to cross the threshold, the computer interactant's level of behavioral realism must be increased to compensate for low agency.

Lee and Nass [10] conducted two experiments where users interacted with either other humans in computer-mediated communication (CMC) or with agents in human-computer interaction (HCI). They investigated how group size (one interactant vs four interactants) and visual representation (varying levels of anthromorphism) of the interactants influenced participants' public compliance and private conformity to a group decision in a social dilemma task. It was found that whether the participants' views were made public or kept private moderated the difference between participants' own views and those of others. Participants would comply more with group opinion in the CMC condition than the HCI condition when their views are made public, but compliance will not be significantly different between the CMC and HCI conditions when their views are kept private.

In Morkes et al [18], two experiments were conducted to investigate the influence of humor of participants' social responses towards their fellow interactants in a cooperative task. The first experiment examined this in the context of CMC interaction while the second experiment examined this in the context of HCI interaction. Participants reported that the other interactants were more likeable, made more jokes and were more sociable in the humor condition than the one where humor was not utilized, in both the CMC and HCI experiments. However, it was found that in terms of self-reported similarity between themselves and their team-mate, participants reported that they felt significantly more similar to the other interactant in the CMC condition when humor was used, but there was no significant difference in the level of similarity felt with the other interactant in the HCI condition.

2.2 Moderating responses towards others in CMC or HCI

There is also research has examined manipulations that can moderate the difference in participant responses towards interactants that could be either avatars (human) or agents (AI). Such manipulations include alignment of team-mate ability and task appropriateness, representation of interface elements or level of anthropomorphism (especially in Embodied Conversational Agents research).

Research by Plaks & Higgins [23] found that participants' performance in a cognitive task with a team-mate would be moderated by how well the information about the team-mate's demographic information fit the stereotypes associated with the nature of the task they were required to perform. Results of the study showed that if participants felt that their team-mates would be able to perform well in the task because their stereotypes were a good fit with task requirements, they would perform poorly. However, if participants felt that team-mates would not be able to perform well in the task because their stereotypes were a poor fit with task requirements, they would perform well. Furthermore, participants in the poor-fit condition performed significantly better than those in the good-fit condition. In other words, the extent to which a team-mate's stereotype fits the task requirements would moderate participants' willingness to put in effort and perform in their share of the task.

Haley & Fessler [5] found that the amount of generosity exhibited by allocators in a Dictator Game could be moderated by the perception of whether they were being observed or not. In the version where the interface of the computer-based Dictator Game had a pair of stylized eyes integrated as part of the interface, participants serving as allocators were found to be more generous towards recipients than those in the condition where there was text in place of the stylized eyes in the interface.

In the field of Embodied Conversational Agent (ECA) research, there have been many studies that examined the effect of the level of anthromorphism on reponses towards computer agents. Studies similar to the ones conducted by Kiesler et al [9], Lee & Nass [10], Gong [4] and Nowak & Biocca [20] have all found that the use of different levels of anthromorphism is able to moderate participants' behavior and social responses towards computer agents that they interact with. However, as these studies have shown, it is not necessarily true that computer agents that are the most anthromorphic would rate the highest in terms of social responses from the participants, as in the case of Lee & Nass [10], Kiesler [9] and Nowak & Biocca [20].

2.3 Summary

This chapter covered the two categories of related work:

- Work that examined what manipulations could moderate the difference in responses to others in CMC situations compared to HCI situations. These manipulations involved elements such as the visibility of scoring information, level of behavioral realism of the AI agent, whether a participants' views are made public or private, and whether humor is utilized by other interactants in the task.
- Work that examined what manipulations could moderate people's responses towards others in either CMC or HCI situations. Such manipulations include alignment of team-mate ability and task appropriateness, representation of interface elements or level of anthropomorphism (especially in Embodied Conversational Agents research).

Chapter 3

Research Problem

This chapter identifies the research gap that currently exists: the use of an approach other than anthropomorphism of the AI agent as a means of moderating the difference in responses towards human and AI team-mates. The research question to address this gap raises the question asking what can be done to moderate the difference in responses towards human and AI team-mates in real-time cooperative games. It concludes with the original contributions that this thesis will add to existing knowledge: identifying and analysing what type of manipulations can moderate these differences, using concepts from Game Theory and Decision-Making research.

3.1 Research Gap

There is work that examines moderating behavioral and social responses towards others in CMC and HCI. Though some studies that are motivated by the SRCT approach do compare people's responses towards others in CMC and HCI scenarios such as work done by Johnson & Gardner [6], Lee & Nass [10], Morkes et al [18]. They mainly compare the effects that their manipulations (such as the private or public nature of a participant's viewpoints) affect the responses towards others in CMC and HCI scenarios. Apart from the work by Ong et al [22], there have been very few studies that focus specifically on how the difference in responses towards others in CMC and HCI scenarios can be moderated.

Much of the work that investigates the difference in responses towards others in CMC and HCI scenarios places the focus on varying the level of anthropomorphism of the avatar (in CMC) or the agent (in HCI), which can be in terms of the visual representation of the agent or avatar, or the behavioral realism of the agent. Varying visual representation involves using different variations of avatar or agent appearance that more closely resembles actual humans (through animated 3D models or photo-realistic images), while varying behavioral realism of the agent involves designing it such that it is able to exhibit behaviors and non-verbal cues (such as smiling or nodding in response to a participant). Moderating the difference in social responses towards avatars and agents through varying the level of behavioral realism is explained by the Threshold Model of Social Influence [2].

There have been problems found with this approach as observed in von der Pütten et al [29], as the participants' perception of whether the behavioral realism is sufficiently similar to a human's is subjective. Also, the level of behavioral realism that an agent can display is bound by limitations of technology available.

Much of the existing work has focused on different aspects of agent anthromorphism – varying the appearance and behavior of the agent to resemble humans – as the primary method of moderating the difference in how people respond towards human interactant compared to a computer agent. The gap in the research that this thesis could help fill is to examine other factors that may help moderate this difference in responses, which are independent of the characteristics of the agent. Examples of two such factors could be the relationship between a participant and the avatar or agent that they are interacting with, while another would be the presence of indicators of performance of a task.

3.2 Research Problem

The experience that a player has while playing a real-time cooperative game such as *DotA 2* or *Team Fortress 2* would be influenced by the choices they need to make during the gameplay, and how difficult it was to make that choice. The choices made relate to the players' behavioral responses, while the difficulty in making the choices relate to players' emotional responses.

This makes real-time cooperative games a suitable context to examine how variations in factors independent of an agent's characteristics – such as the relationship between team members or indicators of the team's – would moderate the difference in how players respond towards human and AI team-mates. Thus the research question of this thesis would be the following: What can be done to moderate the difference in behavioral and emotional responses towards human and AI team-mates in cooperative games?

3.3 Original Contributions

This thesis contributes to the existing work on behavioral and social responses towards others in CMC and HCI scenarios. It identifies and analyses the type of manipulations to that can be made to reduce the difference in behavioral and emotional difficulty responses, in the context of real-time cooperative games played with human and AI team-mates.

Concepts from the fields of Game Theory and Decision-Making – Interdependence Theory [8], Trade-offs and Multi-Attribute Utility Theory (MAUT) [7] in particular – are used to interpret the data gathered from 2 separate experiments, and along with statistical analyses of the recorded data, are used to demonstrate how the relationship between team-mates in a real-time cooperative game can be manipulated to moderate the difference in how people respond towards human and AI team-mates in terms of their behavior and emotional difficulty experienced while making these decisions. It also will explore why some manipulations may change both the behavior and the emotional difficulty involved in that behavior, while others may only change behavior.

3.4 Summary

This chapter has done the following:

- Highlight the existing Research Gap that exists regarding what methods – apart from anthropomorphism of the AI agent – that can be used to moderate the differences in behavior towards others in CMC and HCI situations.
- Establish the Research Problem, asking what can be done to moderate the difference in how players' behavioral and emotional responses towards human and AI team-mates in real-time cooperative games.
- Highlight the original contributions of this thesis to the existing body of knowledge – its use of concepts from Game Theory and Decision-Making to address the Research Problem, and the identification of reasons that suggest why not all manipulations will affect the differences in behavior (and emotional difficulty associated with that behavior) when players play with human team-mates compared to AI team-mates.

Chapter 4

Method

This chapter provides an overview of the methodology used in both studies conducted in this thesis. It provides brief information about the participants involved, the types of materials used to conduct the studies, the study protocol, a justification of the use of a presumed human (PH) team-mate instead of an actual human confederate, and the means of collecting and analyzing data from the studies. More detailed information about each study will be provided in the chapters devoted to them.

4.1 Overview of Studies

The studies used in this thesis were aimed at moderating the differences in behavior towards and feelings about human and AI team-mates in real-time cooperative games. They were designed with the intention to demonstrate that a manipulation of certain design elements in the games would help moderate the difference in responses such that it would be reduced (i.e. participants would respond more similarly towards human and AI team-mates). The studies involved multiple researchers, who helped with design and refinement of the cooperative games and data-gathering questionnaire. However, the author of this thesis was the one who was primarily responsible for designing the cooperative games and data-gathering questionnaire, recruiting participants and running of the experiments, as well as collecting, analyzing and interpreting the data.

4.2 Participants

Participants in the two studies were undergraduates from the National University of Singapore, who participated in the studies as a requirement for course credit. The researchers were not involved in the course from which participants were recruited.

4.3 Materials: Cooperative Games

The researchers designed the two cooperative games specifically to investigate how changes to the relationship between team-mates can moderate the difference in behavior and emotional difficulty involved with that behavior towards human and AI team-mates. The game Defend the Pass (DTP) was designed to investigate the effect of changing the visibility of scoring information on the difference in behavioral and social responses towards human and AI team-mates. The game Return the Ball (RTB) was designed to investigate the effect of changing the level of interdependence between team-mates on the difference in behavioral and social responses towards human and AI team-mates. A common feature of both games was the gameplay feature which required participants to make a choice about their and / or their team-mate's position (which had a direct impact on themselves and / or their team-mate) before proceeding to actually play the game with the positional choices enforced.

4.4 Protocol

Since each study was a 2x2 experiment, half the participants in each study played one version of the respective cooperative games, while the other half played a different version. In the first study, the two versions of DTP were the SCORE version where information about the score and performance was shown; and the NO-SCORE version where this information was not shown. In the second study, the two versions of RTB are the one are the INDEPENDENT version where the player in the front who returns the ball does not depend on the player in the back to successfully return the ball; and the INTERDEPENDENT version where the player in the front depends on the player in the back to successfully return the ball. The details of both games will be covered more comprehensively in their respective chapters.

Each participant played the respective cooperative games for two rounds each, one round with a human team-mate and the other with an AI team-mate. Counterbalancing to rule out order effects was done by making half the participants play with the AI team-mate first and the other half play with the human team-mate first. Each round consisted of the respective game being played multiple times with the same team-mate, much like iterated Prisoners' Dilemma games are played. In DTP, a round consisted of five games. In RTB, a round consisted of seven games.

At the end of each round of games played with the respective team-mates, participants were asked to provide feedback via questionnaire. At the end of both rounds, they were asked to provide feedback in another questionnaire that compared their experiences with the PH and AI team-mates.

4.5 Deception about PH team-mate identity

Though participants were told that the human team-mate they were playing with was joining the game from a remote location over the network, much like a typical Computer-Mediated Communication (CMC) situation, the reality was that they were playing with the exact same AI team-mate in both rounds. In other words, both rounds were similar to a typical Human-Computer Interaction (HCI) situation. This then makes the team-mate in the CMC situation more of a "presumed human" team-mate (PH), and for the remainder of the thesis, the human team-mate will be referenced as such.

The reason for this deception – that the CMC situation was actually a HCI situation – was to ensure consistency of the team-mate's performance during gameplay, and thus any differences in behavioral or social responses towards the different team-mates would be due to the perceived identity of the team-mate (i.e. AI or PH) rather than any other factors.

In fact, Morkes et al [18] suggest that there are benefits in using a PH team-mate if one is to examine the SRCT definition of equality in the strong sense:

> To examine equality in the strongest possible sense, one must have an experiment in which the situation, the procedures, and the measures are identical for those participants who believe they are interacting with a computer and those participants who believe they are interacting with a person (i.e., for experimental purposes, human-human interaction must be construed

as computer-mediated communication between two humans). The comparison of text-based (perceived) CMC and HCI, as performed in this study, is an example of this approach. It has the added advantage that one can eliminate human qualities of appearance, gesture, speech characteristics, and so forth, all of which would be confounding aspects of the human-computer difference.

4.6 Data Gathering and Analysis

Participant behavior in both studies was gathered via game logs which recorded details of all games played by each participant. The main data that was collected from the game logs centered around how often participants selected a particular strategy in each round of games with the PH and AI team-mates, in the different versions of the respective cooperative games. The resulting data, along with the feedback gathered from the questionnaire in the form of likert scale and ranking data were analyzed via statistical methods using SPSS software to determine if any statistically significant differences in the behavioral and social responses of participants existed between the PH and AI team-mates.

4.7 Summary

This chapter has done the following:

- Provide general details about the participants who were involved in the two studies conducted for this thesis
- Gave a brief description of the two games used in the respective studies and what they were designed to investigate
- Described the general study protocol used, with a description of the entire process that each participant went through
- Justified why participants were deceived into thinking they were playing with a human conferate when they were actually playing with the same AI team-mate
- Described how data in the studies was gathered and analyzed.

Chapter 5

Theoretical Framework

This chapter establishes the theoretical framework that is used to contextualize the results of the two studies conducted, demonstrating how trade-offs in decision-making lead to differences in players' responses towards human and AI team-mates, and how increasing interdependence between team members can eliminate this trade-off. Using the example of the game *Capture the Gunner*, this chapter illustrates trade-offs (and how they are represented) in the decision-making process, along with the concept of Interdependence Theory as a means of representing and calculating the level of interdependence between team members.

5.1 Trade-offs in Cooperative Games: Theoretical Framework

In examining six real-time cooperative games, Ong & McGee [21] found in three of them had trade-off situations. In these games that had trade-off situations, players would select the non-optimal choice out of concern for their team-mate's experience of the game. The presence of trade-offs were identified as being linked to three characteristics – (1) level of interdependence between team members, (2) threat of damage or death and (3) symmetry of roles. Of the three, the level of interdependence between team-mates was deemed to be the most important in determining the presence of trade-off situations. These trade-off situations arose because players have many different attributes they considered when they made decisions while playing, and some of them were in conflict.

Examples of the multiple attributes that could factor into a player's decision-making can include the following: the player's own enjoyment of the game or ability to participate, the player's desire to be a good team-mate, the desire to ensure the team wins or performs well, among others.

Ong & McGee [21] suggested that players in cooperative games with other human team-mates have a desire to be a good team-mate because of the player's concern for their team-mate's experience of the game. When there is a trade-off between being a good team-mate and the team's performance, players will opt for the non-optimal choice that is in favour of being a good team-mate over the dominant strategy that guarantees better team performance.

The analysis of six real-time cooperative games in this research also found that this trade-off between being a good team-mate and team performance was not present in all games. The presence of this trade-off in players' decision-making was due to three design factors in the real-time cooperative games analysed. These included level of interdependence between team-mates, symmetry of roles and the threat of damage or death (to one or both team members). Of these three design factors, it was suggested that level of interdependence was most important in determining the presence of a trade-off.

Making the choice out of the desire to be a good team-mate will occur more frequently when playing with a human team-mate. This choice would also be more difficult to make – more "emotionally difficult" [12] – because of the inner conflict / tension a player may feel when deciding whether to choose between being a good team-mate and having better team performance.

Making choices out of the desire to be a good team-mate would occur less frequently with an AI team-mate (and be less "emotionally difficult") due to reasons highlighted by Merritt as part of the Cooperative Attribution Framework [17]. A comparison of these behavioral (and emotional) responses where team-mates are human vs team-mates are AI, there may be consistent with existing work that demonstrates that there is a difference in how people respond towards human and AI team-mates.

The observations in Ong & McGee [21] therefore suggest that one could try increasing the (perceived) level of interdependence between team-mates as a means to reduce or remove the trade-off, since the games with a higher level of (perceived) interdependence between team-mates did not appear to have a trade-off situation present. If increasing the level of interdependence managed to successfully remove trade-offs between being a good team-mate and team performance, then the difference in how players would respond towards human and AI team-mates would perhaps be reduced.

In order to ensure that any manipulations have successfully managed to influence the presence of trade-offs and level of interdependence between team-mates, it is necessary to use some concrete means of expressing and quantifying these concepts.

5.2 Measuring the Dependence on a Team-mate

Among the different concepts defining and measuring Interdependence, Interdependence Theory [8] is suggested as a means of representing the dependence of one team member on another through the use of an Index of Dependence (ID). The calculation of this Index of Dependence is demonstrated using the game *Capture the Gunner* (CTG).

The concept of interdependence is complex and multi-faceted, with different researchers each having different definitions for what they define to be interdependence. Wageman [30] highlighted that at a broad level, there are two types of interdependence – what is structured in versus how people actually behave. These are referred to as structural and behavioral interdependence respectively. Since structural interdependence involves elements external to the individual(s) and their behavior, it is possible to manipulate this kind of interdependence through the design of the activity's features of the task, definition of goals, distribution of rewards etc. As such, structural interdependence can be differentiated further according to inputs and outcomes – known as task and outcome interdependence respectively.

Examining task interdependence in work teams has often used Thompson's work[26] as a basis for analysis. Thompson's work specified the existence of three types of task interdependence – pooled, sequential and reciprocal – where the required complexity of coordination between members increased from one type to the next. In other words, a situation where members of a group are sequentially-interdependent has more interdependence than a situation where members of a group are pooled-interdependent, and that a sequentially-interdependent group contains all elements of a pooled-interdependent group.

Victor and Blackburn [28] suggest that because Thompson's work classifies interdependence using a Guttman-type scale, it is not able to quantify the extent to which they vary. For example, it is not possible to determine whether four "pooled interdependence" is more than two "reciprocal interdependence". They also identify Interdependence Theory [8] as one framework that is able to address this issue.

Interdependence Theory uses game theory payoff matrices as a way to represent the relationship between two members of a 2-person group (or dyad), in terms of their own outcomes from interacting with the other party. In the context of games, the outcomes can be in the form of points scored.

Interdependence Theory payoff matrices are meant to show how the choice of each member performing one of two possible actions affects the outcome of a particular team member. It illustrates the extent to which a team member's outcomes are dependent on their own actions as well as the actions of the team-mate. This can be expressed in three types of control that exist in an outcome matrix:

Actor Control (AC) the control that an individual A has over his or her own outcomes independent of the actions of the partner, B.

Partner Control (PC) the control that the partner B has over the outcomes of individual A, independent of the actions of A.

Joint Control (JC) the control that the actions of A and B together have on A's outcomes.

The Index of Dependence (ID) for A can thus be calculated from the AC, PC and JC values, using the formula in 5.1:

$$ID = \frac{JC^2 + PC^2}{JC^2 + PC^2 + AC^2}$$

Figure 5.1: Formula for Index of Dependence

The ID values range from 0 (completely independent) to 1 (completely dependent), and any value that is within that range indicates that there is some degree of interdependence of outcome between team members.

The game *Capture the Gunner* (CTG) [13] will be used as an example to illustrate the outcome matrix and calculation of the Index of Dependence. CTG is a real-time cooperative game that can be played between two members on the same team. It involves the two team members (black and grey circles) having to both successfully touch the gunner (yellow circle) consecutively at specific points, in order to succeed and progress to the next level.

While attempting to do this, they need to avoid bullets that the gunner fires, because if either team member is hit by a bullet, the team will lose

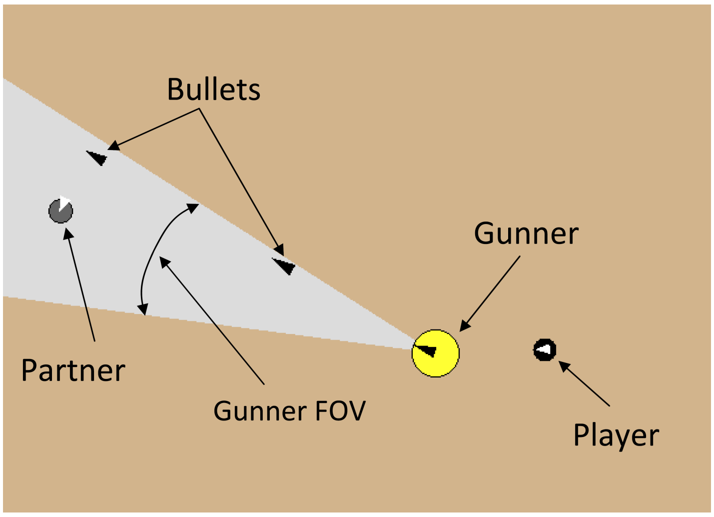


Figure 5.2: *Capture the Gunner* (CTG)

and the game will end. In order for a player to make it easier for their team member to successfully touch the gunner, the player can hit a button on the keyboard to signal to the gunner, drawing its attention towards itself (and away from the team-mate). This increases the chance of the gunner targeting the player by 50%. Thus a player who signals the gunner this way creates extra risk for himself, since the majority of the gunner's attention will be focused on him. The relationship between the two team members in CTG can be represented in the outcome matrix in Table 5.1:

Player in CTG who has already made his touch on the gunner and helping team-mate by signaling for gunner's attention		
	Player manages to evade gunner's bullets	Player gets hit by gunner's bullets
Team-mate manages to touch gunner without getting hit	1	0
Team-mate gets hit by gunner's bullets	0	0

Table 5.1: 2x2 Matrix for CTG

Based on Table 5.1, the team will score 1 point if: (1) the player has already made his touch; (2) was helping the team-mate draw the gunner's attention while evading the gunner's bullets (3) and the team-mate managed to touch the gunner without getting hit. If either team member gets hit by a bullet, then they score 0 points (because the game ends). Based on this, the AC, PC and JC values for the player can be calculated.

Actor Control (AC) depends only on the actions of the player, and this value is derived by comparing the average degree to which the outcomes as a result of the player evading the gunner's bullets is greater than the outcomes as a result of the player getting hit by the gunner's bullets. Comparing the column for "Participant manages to evade gunner's bullets" (1 + 0) against the column for "Participant gets hit by gunner's bullets" (0 + 0), the average of these two is 0.5, and is the value of AC.

Partner Control (PC) depends only on the actions of the team-mate, and this value is derived by comparing the average degree to which outcomes as a result of the team-mate touching the gunner without getting hit is greater than the outcomes as a result of the team-mate getting hit by the gunner's bullets. Comparing the row for "Team-mate touches gunner without getting hit" (1 + 0) against "Team-mate gets hit by gunner's bullets" (0 + 0), the average for these two is 0.5, and this is the value of PC.

Joint Control (JC) depends on the actions of the player and the team-mate, and this value is derived by comparing the average degree to which the outcomes as a result of both team members evading the bullets or both getting hit is greater than if either team member had gotten hit. Comparing the cells where the player and team-mate both successfully evade or both are hit (1 + 0) against when only the player or the team-mate gets hit (0 + 0), the average for these two is 0.5, and this is the value of JC.

The Index of Dependence (ID) for player A in this scenario is shown in Table 5.3:

$$ID = \frac{0.5^2 + 0.5^2}{0.5^2 + 0.5^2 + 0.5^2} = 0.67$$

Figure 5.3: Formula for Index of Dependence

An ID value of 0.67 suggests that to a certain extent, player A's outcomes depend on the actions of his team-mate. This indicates that there is some interdependence between both team members.

5.3 Trade-offs and Emotional Difficulty

Findings from decision-making research have demonstrated that maximizing utility for oneself is not the only attribute that influences people's decisions. In fact, there are multiple attributes, each with their own level of importance(weightage) that influence a decision. When these attributes are in conflict, they create situations where each option in a choice has a trade-off between the gains and losses in each of these attributes.

These trade-offs can be represented in a format suggested by Chatterjee & Heath [3], which can be then used to determine how difficult it is to choose between the options available. This provides an indication of a person's emotional difficulty [12] in choosing between the two options, and is demonstrated with an example using the game *Capture the Gunner* (CTG).

When examining behavior and emotional difficulty involved in making decisions, it is necessary to have some means of quantifying and calculating what is gained from choosing a particular option. Using the example of CTG, the utility gained from choosing a particular option is calculated to help explain behavior and the difficulty in choosing between options is calculated to help explain emotional difficulty.

Research into the process of decision-making has identified that some decisions are difficult to make while others are easy. Multi-Attribute Utility Theory (MAUT) [7] recognizes that there are different motivations people have for picking one option over another when faced with a decision, and has attempted to decompose these choices into component attributes that factor into a decision. An example of multiple attributes considered in a decision could include the desire to maximise a player's self-interest and the desire to be a good team-mate.

When different motivations that a player has (e.g. maximizing self-interest and being a good team-mate) are in direct conflict, a trade-off situation is the result. In such instances, a player must make choices that help fulfill one motivation and the expense of the other. Depending on what each motivation is and how much a player prioritizes each of them, the decision between the options to fulfill the conflicting motivations can be more emotionally difficult or less emotionally difficult.

In the case of CTG, after player A has successfully touched the gunner, he can decide to take risks to help the team-mate (by signaling) or avoid the gunner to ensure his survival. The places him a situation where he has two interests that are in conflict: increase the risk to himself (to help the team-mate), versus minimize risk to himself (and make it harder for the team-mate to complete the level). This is in addition to the other interest of wanting the team to perform well. Using hypothetical values, assuming player A has been asked to rate the following questions on a 5-point Likert scale, the values for these two attributes are shown in Table 5.2.

From the Table, it is clear that player A values his own survival more than helping his team-mate, and thus this information can be used to assign weightages to the different attributes used in decision-making. Therefore,

Statement	Ratings
It is important that I help my team-mate	2
It is important that I ensure my own survival	4

Table 5.2: Ratings of Attributes in CTG

assuming that three attributes that player A cared about when playing CTG were the team's performance, his own survival and helping his team-mate, the weightages could be distributed in a manner shown in Table 5.3.

Attributes	Weightage
Weightage-Performance	0.5
Weightage-Survive	0.3
Weightage-Help	0.2

Table 5.3: Weightages of Attributes in CTG

Thus, using these assigned weights for the attributes of team performance, concerns about player A's own survival and concerns about helping the team-mate, player A has two possible options to choose between – helping the team-mate make a successful touch (by taking a risk) or not helping the team-mate. Since signaling the gunner increases the risk of being targeted by 50%, it will be assumed that this reduces player A's chances of survival by 50%, while simultaneously giving the team-mate a 50% higher chance of successfully making a touch on the gunner. In both cases whether or not player A helps the team-mate, the assumption is that the team will still just gain one point.

The gains and losses from these two choices can be represented in Table 5.4, in a manner similar to that found in Chatterjee & Heath [3]. In their calculations, they used the average attribute levels as a common reference point for any losses and gains in an attribute when choosing one option over another. This was used to then determine how much was gained or lost (relative to the average value) by choosing an option. The use of proportions (percentages) to represent the gains and losses for each attribute was to provide a common metric for measurement since different attributes may have different types of measurements.

The utility gains or losses calculated using MAUT can be used to help explain why there is a tendency to choose a particular option. The weightages for attributes can be represented by Wt_1, the percentage gains (when x is more than zero) are represented by X_G and the percentage losses (when x is less than zero) are represented by X_L.

Decision	Wt Performance	Wt Survive	Wt Help	Score	Risk to Player A	Team-mate chances of success
Help	0.5	0.3	0.2	1 (0.00%)	-50 (-100%)	50 (+100%)
Don't Help	0.5	0.3	0.2	1 (0.00%)	0 (+100%)	0 (-100%)
Average				1	-25	25

Table 5.4: Gains and Losses in CTG

Thus, utility gained from choosing engage in a particular behavior (choosing one option over the other) with two weighted attributes that are considered in decision-making can be expressed in the formula shown in Figure 5.4, assuming there is a gain for Attribute 1 and a loss for Attribute 2.

$$Utility = Wt_1 * X_G + Wt_2 * X_L$$

Figure 5.4: Formula for Utility of choosing option

In order to examine how much utility player A believes they will get from helping the team-mate, it can be calculated using the formula in Figure 5.4, and this calculation is shown below in Figure 5.5, using the percentage gains and losses from choosing to help the team-mate.

$$UtilityofHelping = 0.5 * 0.0\% + 0.3 * -100.0\% + 0.2 * 100\% = -10.0$$

Figure 5.5: Formula for Utility of choosing option

According to the calculations, the overall utility gained from choosing to help the team-mate is -10, which suggests that it is not in player A's interest to help the team-mate by taking a risk and drawing the gunner's fire.

To see how close the two options are in terms of utility (and how difficult it would be to choose between them), this difficulty of choice can be calculated using the formula listed in Figure 5.6. In this formula, The rationale for the formula used by Chatterjee & Heath (based on Tversky & Kahneman [27]) is as explained below:

> ...can then be subjected to a standard value curve assessment where losses are assumed to carry about 2.5 times more

value than gains, and where diminishing marginal sensitivity is represented by taking the proportions to the power of 0.88 (Tversky and Kahneman, 1992).

Using the values from Table 5.4, the degree to which the two options (help or don't help) are close in terms of utility gained or lost is shown in Figure 5.7.

$$Gain(X) = (Wt_1) * (X_G)^{0.88}$$

$$Loss(X) = -2.5 * (Wt_1) * (X_L)^{0.88}$$

$$f(X) = Gain(X) - Loss(x)$$

Figure 5.6: Formula for Size of Trade-offs

$$Gain(Help) = 0.2 * 100.00\%^{0.88} = 0.2$$

$$Loss(Help) = 2.5 * 0.3 * 100.00\%^{0.88} = -0.75$$

$$f(Help) = Gain(Help) + Loss(Help) = -0.55$$

$$Gain(NoHelp) = 0.3 * 100.00\%^{0.88} = 0.3$$

$$Loss(NoHelp) = 2.5 * 0.2 * 100.00\%^{0.88} = -0.5$$

$$f(NoHelp) = Gain(NoHelp) + Loss(NoHelp) = -0.2$$

$$Difficulty = f(Help) - f(NoHelp) = -0.35$$

Figure 5.7: Calculating Size of Trade-offs in CTG

As seen from Figure 5.7, the difference between the overall utility of player A helping the team-mate versus not helping the team-mate is a negative value, suggesting that the decision is inclined towards not helping the team-mate. The closer the value is to 0, the more difficult it is to make the decision.

5.4 Rationale for calculating the Index of Dependence and Trade-offs

Calculating the values of trade-offs and the values of the Index of Dependence is relevant to this thesis because it provides an empirical means of validating the effect of the manipulations carried out in the two studies.

The research problem of this thesis is concerned with examining how the difference in responses towards human and AI team-mates in real-time cooperative games can be moderated. The findings of Ong & McGee [21] suggest three things in relation to this:

1. People care about their team-mate's experience when faced with trade-offs, and thus they will be willing to choose the non-optimal option that will allow their team-mate to have a better experience.
2. Since people respond towards human and AI team-mates differently, as seen in the work by Merritt et al [17, 16, 14, 15], they would be less likely to make these non-optimal choices with an AI team-mate compared to a human team-mate.
3. If the interdependence between team members is one characteristic that determines the presence of a trade-off situation, then it would be necessary to see how the manipulations in the two studies influence both trade-offs and interdependence in their attempt to reduce the difference in how people respond towards human and AI team-mates.

5.5 Summary

This chapter has done the following:

- Provided the theoretical framework as the basis to analyse the results of the two studies. This framework highlighted how trade-offs lead to people making non-optimal decisions out of concern for their team-mate's experience of the game and how interdependence could help remove the presence of these trade-offs.
- Highlighted Interdependence Theory as a means of representing the relationship between team members, and the calculation of the Index of Dependence using the game *Capture the Gunner* with hypothetical values to determine how much one team member depends on the other

for their outcomes.

- Highlighted Multi-attribute Utility Theory (MAUT) and how choices and trade-offs are represented in decision-making research. Using the game *Capture the Gunner* with hypothetical values, demonstrated how the utility of choosing a particular option could be calculated, as well as demonstrating how the difficulty of choosing between two options could be calculated.
- Justified why it was necessary within the scope of this thesis to calculate the Index of Dependence, Utility of choosing a particular option, and the difficulty in choosing between two options.

Chapter 6

Study 01: Changes to Score Display

This chapter provides details about a study to determine whether changes to the display of game scores would moderate the difference in team-mate behavior and experience. The results of the study – the manipulation of the visibility of scoring information moderated differences in behavior but not emotional difficulty – are presented and analysed, using calculated values of the Index of Dependence, utility gained from a choice and the difficulty of choosing between the two options as further evidence to support the results.

6.1 Study Details

A 2x2 study was conducted with 73 participants who played two versions of a real-time cooperative game *Defend the Pass* (DTP). They played two rounds of five games each – one round with a "presumed human" (PH) team-mate and the other round with an artificial intelligence (AI) team-mate. One version of the game had scoring and performance information shown (SCORE), while the other version had this information hidden (NO-SCORE). Data about the participants' selection of strategy to protect or sacrifice the team-mate and responses from a set of questionnaires was collected.

6.1.1 Participants

There were 73 participants involved in this experiment (16 males, 57 females). They were undergraduate students between the ages of 20 and

24 years old, with an average age of 21.1 years. They filled in a short questionnaire that gathered demographic information such as age, gender, experience with computer games, whether they had played cooperative games with human team-mates, and whether they had played cooperative games with computer team-mates. This data for the 2 versions is shown in Table 6.1:

	Score	**No Score**
Males	10	6
Females	27	30
Mean Age (years)	21.3	20.9
Experience with computer games:		
Very Little	6	11
Little	6	6
Some	11	11
Much	9	9
Coop games (human)	13	14
Coop games (AI)	4	2

Table 6.1: *Defend the Pass* demographics by condition

6.1.2 Game: *Defend the Pass* (DTP)

Defend the Pass (DTP) is a real-time cooperative game where the participant can play with an AI team-mate or a PH team-mate. The objective of the game is for both team members to cooperate and kill an army of 30 monsters that is attempting to escape through a path (the "Pass") in the middle of the screen (see Figure 6.1). The formation of the 30-monster army is procedurally generated at the start of each game. The monsters move downwards from the top of the screen, towards the bottom of the screen. If they manage to exit from the bottom of the screen, they are deemed to have escaped.

Each team member controls their own avatars, which are placed near the bottom of the game area. The participant's avatar (in green) is placed on the right side of the Pass, while the team-mate's avatar (blue) is placed on the left side of the Pass. The participant's avatar is labelled "ME", while the team-mate's avatar label depends what type of team-mate the participant is playing with. If the team-mate is AI, then the avatar is labelled "AI" followed by a number (e.g. AI-04). If the team-mate is PH, then the avatar is labelled with a name that is similar to that of the participant's

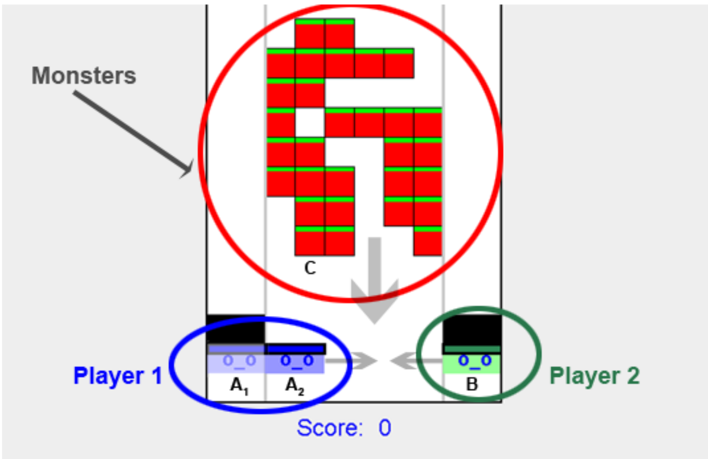


Figure 6.1: DTP game screen

gender (e.g. “James” or “Jessica”). Each team member has their own health bars, with a horizontal one above their avatars, and a vertical one on their respective sides of the Pass.

To stop the monsters from escaping, both team members can shoot bullets horizontally along the line that they are on. The team-mate shoots automatically, while the participant can shoot by pressing the up arrow key. Ammunition for both team members is limited to 100 bullets (represented by a bar with a gun below it), and it does not refill once expended. Monsters that are in line with the team members’ positions when they are shooting will get hit and take damage. Each monster requires four hits to kill, and this is represented by a green health bar at the top of every monster. Once a monster dies, they will disappear from the game space. The game then ends when no more monsters remain on screen because they have been killed or managed to escape.

There are two strategies that participants choose between at the start of each game of DTP, which involves selecting the position of the team-mate using the mouse before the start of each game. The “Protect” strategy places the team-mate in Pos 1 (See Figure 6.2), under the black block on the left side of the Pass. When the team-mate is placed in this position, it is “protected”, and able to shoot at monsters that are aligned horizontally with its position. However, placing the team-mate in the protected position makes it more difficult to kill the monsters that are moving through the Pass, resulting in more of them escaping. The “Sacrifice” strategy places the team-mate in Pos 2, which is on the Pass, making that section of the Pass narrower. Placing this team-mate in the “sacrificed” position makes it easier to kill more monsters and prevent them from escaping. However, each monster that touches the team-mate will reduce its health, and eventually the team-mate will die. The team-mate dying is a guaranteed consequence of choosing the "Sacrifice" strategy because the total number of hits from monsters required to kill the team-mate is 80% of number of monsters in the leftmost column of the army, meaning the total number of monsters that damage the team-mate will always be more than the team-mate's maximum health.

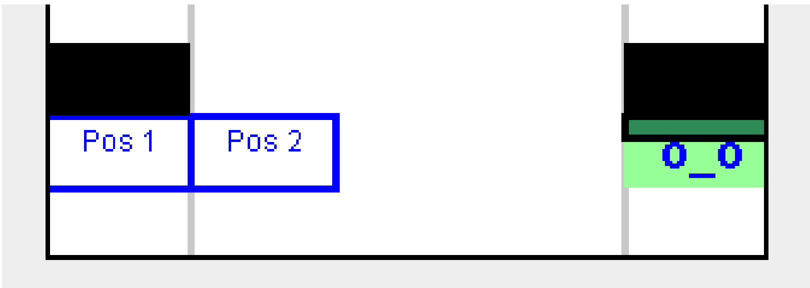


Figure 6.2: Positioning the team-mate

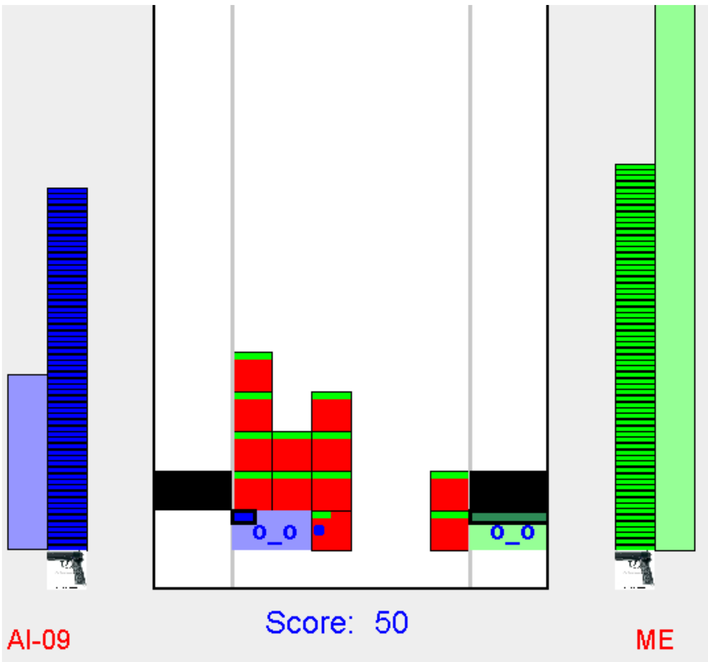


Figure 6.3: Monsters in the pass

6.1.3 Experimental Manipulation

The two different versions of DTP played by the participants were the SCORE and No-SCORE versions. The SCORE version (played by 37 participants) displayed information related to performance of the team cumulatively over the round of five games. The information displayed included the current cumulative running total score achieved by the team, which was updated real-time with every monster that was killed or that escaped. Each successful kill awarded the team 10 points, while each monster that managed to escape deducted 55 points from the team's score. At the end of each game in the round, participants were presented the current cumulative score, as well as information about which strategy was chosen, how many monsters escaped and how many monsters had been killed. Participants had to click on a link in the dialogue box in order to progress to the next game, thus ensuring they had to see this information before proceeding.

To help better inform their decision between sacrificing and protecting the team-mate, they had information available about the number of monsters killed, the team's cumulative score and the status of the team-mate during the game.

The No-SCORE version (played by 36 participants) did not display any of this information to participants at all. There was no running total score shown during gameplay, and the dialogue box presented to participants at the end of each game was blank except for the link they needed to click to progress.

In this case, the information available to players included the status of the team-mate of the game, no information about score, and no accurate information about number of monsters killed – players could still have a rough gauge, but they would not know for sure.

6.1.4 Study Session Protocol

Participants arrived at private testing room, and did not meet other participants. They were assured that their comments would be kept anonymous and confidential. At the start of the study, participants filled out the questionnaire that gathered the demographic information shown in Table 6.1. They then proceeded to view a series of slides that explained how the experiment would be run, as well as the game mechanics behind the game, followed by playing a tutorial round to gain a better understanding of the game, and clarify any uncertainties that participants had about the game or the experiment. The tutorial round consisted of four games played with an AI team-mate called AI-Tutorial, where participants tried the sacrifice and protect strategies twice each to get an understanding of the consequences of each strategy. The version of the game presented to participants in the slides and tutorial games corresponded with the version they were assigned to play during the actual experiment.

Once participants acknowledged that they understood the game, the actual experiment started. Each participant was assigned to play two rounds of five games each – one round was played with the AI team-mate while the other was played with the PH team-mate. Counterbalancing via alternating the order of team-mates that participants played with was done to rule out order effects. The researcher informed participants that during the experiment they were not being assessed on their performance, and that information about their performance and choices was not being logged. Also, the researcher was not present in the room while participants were playing the game. These measures were to ensure that there would be no pressure to perform well or concerns about the researcher's opinion of the participants based on their choices having an influence on the strategy they chose.

After each round of five games, the researcher re-entered the room, and got participants to fill in a questionnaire that sought to gather information about their experience of the recently completed round of games.

There were no indications during any of the sessions that participants doubted the identity of the PH team-mate. No additional checks were conducted about this as pilot studies and previous studies have shown that participants make it quite clear if they doubt the identity of the team-mate.

6.1.5 Measures

The study gathered demographic data, game log data, and self-reported quantitative data (5-point and 10-point Likert scales) from participants.

The game logs captured the activity of each participant for every game played with both types of team-mates. For each game played, the logs detailed which position the team-mate was placed in, the number of monsters that escaped (separated into instances when protecting and sacrificing), the amount of ammunition used by the participant, the amount of ammunition used by the team-mate, the amount of time taken by participants to choose their team-mate's position, and the time at which the team-mate died (if applicable).

After each round of 5 games, participants filled in a questionnaire, which had a series of statements to rate on a Likert scale relating to the round of games they had just completed. These questions included:

- I considered the survival of my team-mate when deciding on my team-mate's position (5-point Likert scale)
- I considered the goals of the team when deciding on my team-mate's position (5-point Likert scale)

After both rounds of games, participants were asked questions in which they were required to compare their human and computer team-mates. These questions were only asked at the end of both rounds, to minimize the possibility of participants' responses from the first play session influencing their behavior in the second.

Also, since the questions seek to make a comparison between both types of teammates, it was necessary for them to play both sessions in order to make a fair comparison. These questions included:

- How emotionally difficult was it to decide where to position your human team-mate? (10-point scale, 1-not difficult at all to 10-very difficult)
- How emotionally difficult was it to decide where to position your computer team-mate? (10-point scale, 1-not difficult at all to 10-very difficult)

The questions used in the questionnaire were not based on previously-used measures (it is not clear that such pre-existing measure exist), but the questions used were refined for validity via feedback received during pilot testing.

6.2 Results

There are two main results of this study: (1) the manipulation of the visibility of scoring information reduces the difference in how often participants sacrifice their AI team-mate compared to their PH team-mate; (2) however this manipulation does not reduce the difference in emotional difficulty participants have in choosing whether to sacrifice their team-mate when playing with their AI team-mate compared to their PH team-mate. Other results include those relating to the performance of the teams in each of the possible scenarios, as well as results relating to participants' experience of playing DTP.

6.2.1 Behavioral Results

In the SCORE version, participants sacrificed their AI team-mate (71.35% of the time) more often than their PH team-mate (55.14% of the time), as seen in Figure 6.4. A paired-samples t-test found this difference to be statistically significant ($p < 0.01$), $t(36)=3.60$.

In the NO-SCORE, participants sacrificed their AI team-mate (51.11% of the time) almost as often as their PH team-mate (44.44% of the time),

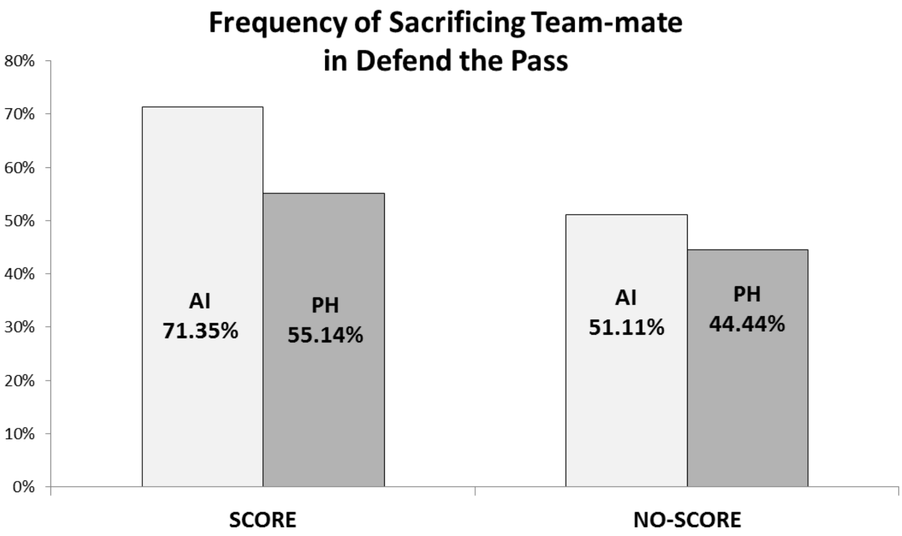


Figure 6.4: Comparison: percentage of games when team-mate is sacrificed

as seen in Figure 6.4. A paired-samples t-test found the difference to be not statistically significant ($p > 0.05$), t(35)=1.87.

When playing with the AI team-mate, participants playing the SCORE version of DTP sacrificed their team-mate more often (71.35% of the time) than those who played the NO-SCORE (55.11% of the time) version. An independent-samples t-test found this difference to be statistically significant ($p < 0.01$), t(71)=3.33.

When playing with the PH team-mate, participants playing the SCORE version sacrificed their team-mate almost as often (55.14% of the time) as those who played the NO-SCORE version (44.44% of the time). An independent-samples t-test found this difference to be not statistically significant ($p > 0.05$), t(71)=1.55.

6.2.2 Emotional Difficulty Results

In the SCORE version of DTP, participants reported having more emotional difficulty in choosing their position with the PH team-mate (3.81 out of 10) than with the AI team-mate (1.84 out of 10). A paired-samples t-test found this difference to be statistically significant ($p < 0.01$), t(36)=-1.973.

In the NO-SCORE version, participants reported having more emotional difficulty in choosing their position with the PH team-mate (4.00 out of 10) than the AI team-mate (1.92 out of 10). A paired-samples t-test found this difference to be statistically significant ($p < 0.01$), t(35)=-0.286.

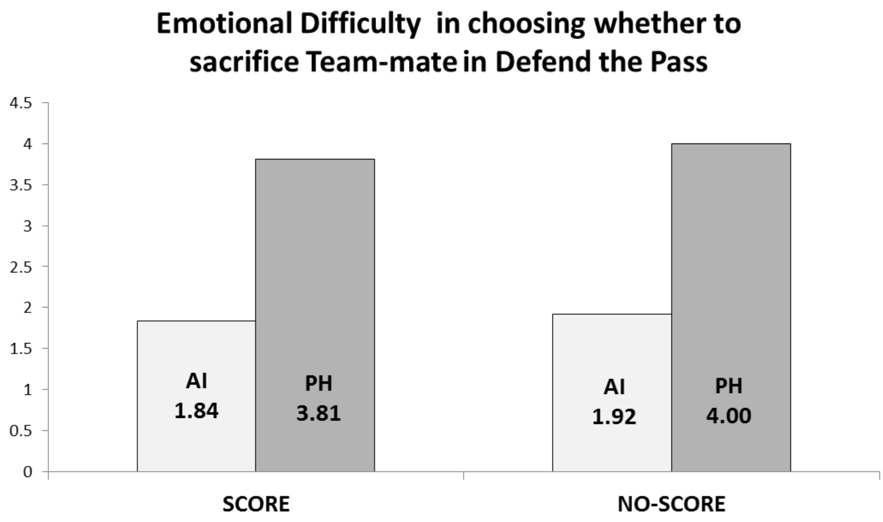


Figure 6.5: Comparison: Emotional Difficulty ratings of players

6.2.3 Performance-related Results

The average number of monsters killed by the team and the average score received when placing the both types of team-mate in either the sacrificed or protected position in each version of the game is shown in Table 6.2. However, these scores range from negative to positive values and thus need to be normalized for consistency. This can be done by adding 1650 to all the scores, since the lowest possible score a player can obtain is -1650 (-55 points for 30 monsters escaping), and this addition would set -1650 to be the "absolute zero" for the points scored.

	SCORE		NO-SCORE	
	AI	PH	AI	PH
Sacrifice				
Monsters Killed	26.29	27.14	25.88	26.7
Average Score	95.95	142.7	N.A.	N.A.
Normalized Score	1745.95	1792.7	N.A.	N.A.
Protect				
Monsters Killed	20.46	19.63	20.4	20.18
Average Score	-224.7	-270.35	N.A.	N.A.
Normalized Score	1425.3	1378.65	N.A.	N.A.

Table 6.2: Team performance in DTP

6.2.4 Experience-related Results

Participants were asked a variety of questions in the self-reported questionnaires at the end of each round of games. Two of the questions are related to the motivations that contribute to the trade-off in DTP, and their results are shown in Table 6.3.

Statement	SCORE		NO-SCORE	
	AI	PH	AI	PH
I considered the survival of my team-mate when deciding on my team-mate's position	2.32	2.86	2.58	3.52
I considered the goal of the team when deciding on my team-mate's position	4.18	3.89	3.75	3.67

Table 6.3: Self-reported ratings on participants' motivations for choices

The results demonstrate that participants consider the survival of the PH team-mate more than that of the AI team-mate in both versions of DTP. It also demonstrates that they consider the goal of the team more with the AI team-mate than with the PH team-mate in both versions.

6.3 Discussion

This section discusses the results of the study, with the use of the trade-off calculations and Index of Dependence to provide support in explaining the behavior and emotional difficulty results.

The Index of Dependence for both versions of DTP show that the manipulation has not resulted in a significant change in the interdependence between team members. The calculations of the trade-off values show that the manipulation of scoring information has reduced the difference in utility that participants gain when choosing to sacrifice the team-mate. This mirrors the behavioral results of the study.

The calculation of the difficulty in choosing between the two options illustrates that the manipulation of scoring information has not reduced the difference in difficulty of choosing with the AI and PH team-mates in the NO-SCORE version compared to the SCORE version. This mirrors the emotional difficulty results of the study.

6.3.1 Calculating the Index of Dependence (ID)

Calculations of the Index of Dependence (ID) for DTP indicate that between the SCORE and NO-SCORE versions of the game, the values of the ID are about the same. This suggests that the manipulation of the visibility of scoring information did not affect the level of interdependence between team members.

Victor & Blackburn [28] used the concepts from Interdependence Theory to determine the degree to which one team member depends on another. Using the information about team performance in DTP (see Table 6.2), it is possible to represent the relationship between team members in a 2x2 matrix, using the team performance as an indicator of their output.

An example of this 2x2 matrix for number of monsters killed is shown in Table 6.4 for the case of a participant playing the NO-SCORE version with an AI team-mate, when deciding to protect the team-mate. A full list of all the matrices can be found in Appendix A.

Protected AI Team-mate (NO-SCORE version)		
	Team-mate successful in killing Monsters	Team-mate unsuccessful in killing Monsters
Participant successful in killing Monsters	30	20.4
Participan unsuccessful in killing Monsters	20.4	0

Table 6.4: Representation of relationship when participant protects AI team-mate in NO-SCORE version

In Table 6.4,the assumption is that if both team members are able to perform perfectly, all 30 monsters will be killed, hence that value is assigned to that particular cell. If both fail to kill any monsters at all, then the value assigned to the cell would be 0. It is also assumed that if a team performs imperfectly, the value of those cells would correspond to the average number of monsters killed in that particular scenario.

The matrices can then be used to calculate the three components that make up the Index of Dependence – Actor Control (AC), Partner Control (PC) and Joint Control (JC), using the formula described in the Theoretical Framework chapter of this thesis. Table 6.5 shows the calculated values of AC, PC, JC and ID for each of the eight possible scenarios.

	SCORE		**NO-SCORE**	
	AI	**PH**	**AI**	**PH**
Sacrifice				
Actor Control	15	15	15	15
Partner Control	15	15	15	15
Joint Control	-11.29	-12.14	-10.88	-5.18
Protect				
Actor Control	15	15	15	15
Partner Control	15	15	15	15
Joint Control	-5.46	-4.63	-5.4	-11.7
Index of Dependence (ID)				
ID (Sacrificing)	0.610	0.623	0.604	0.616
ID (Protecting	0.531	0.522	0.530	0.528

Table 6.5: Actor Control, Partner Control, Joint Control and Index of Dependence for 8 possible scenarios

The spectrum for the Index of Dependence (ID) ranges from 0 (not dependent on team-mate at all) to 1 (completely dependent on team-mate).

As seen from the table, the participant is somewhat dependent on the team-mate whether they decide to protect or sacrifice the team-mate. Furthermore, the dependence on the team-mate is higher when the team-mate is sacrificed as opposed to when the team-mate is protected. This is not surprising because a sacrificed team-mate not only helps to kill monsters, but also narrows the path they can move along, making them easier for the participant to kill.

When comparing the participants' ID in the SCORE and NO-SCORE versions of DTP, the values are about the same. This suggests that the the manipulation of the visibility of scoring information does not significantly affect the level of interdependence between team members.

6.3.2 Determining Weightages of Attributes

In DTP, participants are faced with the trade-off between two attributes they may deem important – the survival of their team-mate and the performance of the team (represented by the number of monsters that escape and the score obtained). A calculation of the weightages for these two attributes was done using the experience-related results of the questionnaire. The three attributes assigned weightages were the the score, the number of monsters killed and whether the team-mate survived.

As seen from Table 6.3, participants report considering the survival of the team-mate more in the NO-SCORE version of DTP compared to the SCORE version, and this suggests that regardless of team-mate identity, team-mate survival is considered more important when scoring information is hidden as opposed to when it is shown.

Also, participants report considering the goal of the game when choosing their team-mates' position in the SCORE version of DTP compared to the NO-SCORE version. This suggests that regardless of team-mate identity, the goal of the team is more important when scoring information is shown as opposed to when it is hidden. Furthermore, when comparing between PH and AI team-mates, considering the survival of the team-mate was rated higher for the PH team-mate than the AI team-mate. Also, considering the goal of the team was rated higher for the AI team-mate than the PH team-mate for both conditions. These results suggest a few things:

- Participants are concerned more with the survival of the PH team-mate than the AI team-mate.
- Participants are concerned more with the survival of the team-mate in the NO-SCORE version of DTP compared to the SCORE version.

- Participants are concerned less with the goal of the team with the PH team-mate than the AI team-mate.
- Participants are concerned with the goal of the team less in the NO-SCORE version of DTP compared to the SCORE version.

Therefore, if we were to assign weights to these two attributes of team-mate survival and goal of the game, they could be assigned as in Table 6.6 to reflect the importance that is placed by participants on the different attributes in the two different versions of DTP.

	SCORE		NO-SCORE	
	AI	PH	AI	PH
Weightage-Kills	0.5	0.5	0.5	0.5
Weightage-Score	0.4	0.3	0.4	0.3
Weightage-Survival	0.1	0.2	0.1	0.2

Table 6.6: Weightages for the two attributes in DTP

6.3.3 Choosing between the two options in DTP

The different options that participants get to choose between can be represented in the format by [3] for each version of the game played with the different types of team-mates, including the weightage of each attribute considered. This is shown in Table 6.7. The percentage gains and losses for choosing to sacrifice rather than protect (and vice versa) are included in parenthesis.

Version	Team-mate	Decision	Wt Kill	Wt Score	Wt survive	Monsters Killed	Score	Team-mate Survival
SCORE	AI	Sacrifice	0.5	0.4	0.1	26.29 (+12.47%)	1071 (+10.11%)	0 (-100%)
		Protect	0.5	0.4	0.1	20.46 (-12.47%)	750.3 (-10.11%)	1 (+100%)
	PH	Sacrifice	0.5	0.3	0.2	27.14 (+16.06%)	1118 (+13.02%)	0 (-100%)
		Protect	0.5	0.3	0.2	19.63 (-16.06%)	704.7 (-13.02%)	1 (+100%)
NO-SCORE	AI	Sacrifice	0.5	0.4	0.1	25.88 (+11.84%)	N.A.	0 (-100%)
		Protect	0.5	0.4	0.1	20.4 (-11.84%)	N.A.	1 (+100%)
	PH	Sacrifice	0.5	0.3	0.2	26.7 (+13.91%)	N.A.	0 (-100%)
		Protect	0.5	0.3	0.2	20.18 (-13.91%)	N.A.	1 (+100%)

Table 6.7: Gains and Losses for each choice in DTP

6.3.4 Explaining Behavioral Results

According to the behavioral results, participants playing the SCORE version sacrificed their AI team-mates more frequently than their PH team-mates. In the NO-SCORE version, the frequency at which participants sacrificed their PH and AI team-mates was about the same. This suggests that the manipulation of the visibility of scoring information was able to moderate the difference in how frequently participants sacrificed their AI team-mates compared to their PH team-mates.

The rationale for such behavior can be examined in terms of the gains or losses that participants feel the decision to sacrifice brings them. Using the formula proposed in Multi-Attribute Utility Theory it is possible to determine the utility a participant feels they will gain or lose by choosing to sacrifice their team-mate. The overall gains and losses from sacrificing the respective team-mates in both versions of DTP is shown in Table 6.8.

	SCORE		NO-SCORE	
	AI *Weight	**PH*Weight**	**AI*Weight**	**PH*Weight**
Percentage Gain / Loss for Kills	12.47*0.5	16.06*0.3	11.84*0.5	13.91*0.3
Percentage Gain / Loss for Score	10.11*0.4	13.02*0.3	0.00*0.4	0.00*0.3
Percentage Gain / Loss for Team-mate survival	-100.00*0.1	-100.00*0.4	-100.00*0.1	-100.00*0.4
Overall Gain / Loss for Sacrificing	**0.28**	**-31.28**	**5.82**	**3.77**
Difference between PH and AI team-mate	**31.56**		**2.05**	

Table 6.8: Gains and Losses for each choice in DTP

In the SCORE version, data from the table shows that participants felt they would gain much more utility from choosing to sacrifice an AI team-mate compared to a PH team-mate. The large difference in utility gained mirrors the significant difference in frequency of sacrificing the AI team-mate compared to the PH team-mate in this version of DTP. This is logical because if participants feel they gain more from sacrificing the AI team-mate compared to the PH team-mate, this would be reflected in their actual behavior when making decisions.

In the NO-SCORE version, data from the table shows that participants felt they would gain about the same amount of utility from choosing to sacrifice an AI team-mate as they would for sacrificing a PH team-mate. This small difference mirrors participants' frequency of sacrificing their team-mates in the NO-SCORE version – the frequency of sacrificing both types of team-mates was about the same. This is logical because if participants feel that they gain around the same amount of utility when playing with AI and PH team-mates, it would be unlikely that they choose to sacrifice one team-mate more frequently than the other.

6.3.5 Explaining Emotional Difficulty Results

According to the Emotional Difficulty results, participants playing the SCORE version had significantly more Emotional Difficulty in choosing their position with the PH team-mate compared to the AI team-mate. In the NO-SCORE version, participants also reported having significantly more Emotional Difficulty in choosing their team-mate's position when playing with the PH team-mate compared to the AI team-mate. These results suggest that the manipulation of the visibility of scoring information was not able to moderate the difference in how much Emotional Difficulty participants had in choosing their position when playing with a PH team-mate compared to an AI team-mate.

The rationale for these Emotional Difficulty results can be examined in terms of how close the two options were to each other in terms of overall gains and losses in utility. This can be done using the formula proposed by Chatterjee and Heath [3], as detailed in the Theoretical Framework chapter of this thesis. The data illustrating this for both versions of DTP with the respective team-mates is shown in Table 6.9. A positive value for difficulty of choice indicates that the choice was tending towards protecting the team-mate.

	SCORE		NO-SCORE	
	AI	**PH**	**AI**	**PH**
Gain from Protecting	0.1	0.2	0.1	0.2
Loss from Protecting	0.33	0.37	0.19	0.22
Gain from Sacrificing	0.13	0.17	0.07	0.09
Loss from Sacrificing	0.25	0.50	0.25	0.50
Overall Utility from Protecting (Gain - Loss)	-0.23	-0.17	-0.09	-0.02
Overall Utility from Sacrificing (Gain – Loss)	-0.12	-0.33	-0.18	-0.41
Difficulty of choice: Utility of Protecting – Utility of Sacrificing	**-0.11**	0.16	**0.09**	**0.39**
Difference in difficulty of choice for AI and PH team-mates (PH – AI)	**0.27**		**0.30**	

Table 6.9: Difficulty in choosing between options in DTP

Based on the data in Table 6.9, it suggests that in the SCORE version it was harder for participants to choose between the two options when playing with the PH team-mate than the AI team-mate. The difference between the difficulty in choosing between options with the AI team-mate compared to the PH team-mate mirrors the significant difference in Emotional Difficulty results in this version of the game.

Data from the table shows that in terms of the difference in difficulty for choosing the position with the PH and AI team-mates, this difference was about the same in the SCORE and NO-SCORE conditions, which mirrors the significant difference in Emotional Difficulty results in this version of the game.

This shows that between the two versions, the difference when playing with AI team-mates compared to PH team-mates is not reduced. This suggests that the manipulation of the visibility of scoring information did not moderate this difference.

6.4 Summary

This chapter did the following:

- Provided a detailed description of the study design regarding the participants, study protocol and data measured
- Described the game used for the study (*Defend the Pass*) and the experimental manipulations used to examine the research question
- Presented and analysed the behavioral results: the removal of scoring information reduced the difference in how often participants chose to sacrifice their AI team-mate compared to their PH team-mate when comparing the SCORE and NO-SCORE versions.
- Presented and analysed the emotional difficulty results: the removal of scoring information did not reduce the difference in emotional difficulty that participants felt in choosing to sacrifice their PH team-mate compared to their AI team-mate when comparing the SCORE and NO-SCORE versions.

Chapter 7

Study 02: Changes to Interdependence

This chapter provides details about a study to determine whether changes to interdependency would moderate the difference in team-mate behavior and experience. The results of the study – the manipulation of the interdependence between team members moderated differences in behavior and emotional difficulty – are presented and analysed, using calculated values of the Index of Dependence, utility gained from a choice and the difficulty of choosing between the two options as further evidence to support the results.

7.1 Study Details

A 2x2 study was conducted with 61 participants who played two versions of a real-time cooperative game *Return the Ball* (RTB). They played two rounds of seven games each – one round with a "presumed human" (PH) team-mate and the other round with an artificial intelligence (AI) team-mate. In one version of the game, participants were told that there was no impact of the actions of the back paddle on the front paddle's actions (INDEPENDENT), while in the other version the actions of the back paddle determined where the ball would next come down for the front to hit (INTERDEPENDENT). Data about the participants' selection of strategy to play using the front or back paddles, and responses from a set of questionnaires was collected.

7.1.1 Participants

There were 61 participants involved in this experiment (17 males, 44 females). They were undergraduate students between the ages of 19 and 27 years old, with an average age of 20.9 years. They filled in a short questionnaire that gathered demographic information such as age, gender, how often they played computer games what genre of games they played, whether they had played cooperative games with human team-mates, and whether they had played cooperative games with computer team-mates. This data for the two versions is shown in the table below:

	Independent	**Interdependent**
Males	7	10
Females	24	20
Mean Age (years)	20.6	21.3
Play computer games:		
Very Rarely	6	11
Rarely	6	6
Sometimes	9	9
Often	6	2
Very Often	4	2
Coop games (human)	23	21
Coop games (AI)	19	18

Table 7.1: Demographics by condition

7.1.2 Game: *Return the Ball* (RTB)

Return the Ball (RTB) is a real-time cooperative game where the participant can play with an AI team-mate or a PH team-mate. The objective of the game is to prevent the ball dropping from the top of the screen to exit through the bottom. Each team member controls one paddle, either the one in the front (closer to the top of the screen) or the one in the back (closer to the bottom of the screen). This depends on which paddle the participant decides to pick for themselves, which is done before the start of each game, using the mouse (see Figure 7.1).

A participant picking the front paddle automatically assigns the team-mate the back paddle, and vice versa. The participants' paddle (green) is labelled "Myself", while the team-mate's avatar (blue) label depends what type of team-mate the participant is playing with. If the team-mate is AI,

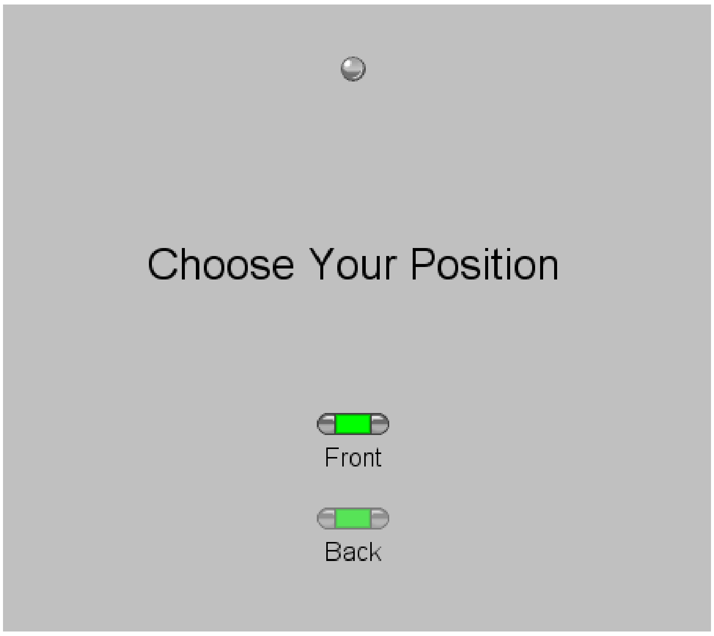


Figure 7.1: Choice between front and back paddles

then the avatar is labelled “Team-mate (COM)”. If the team-mate is PH, then the avatar is labelled “Team-mate (P2)”.

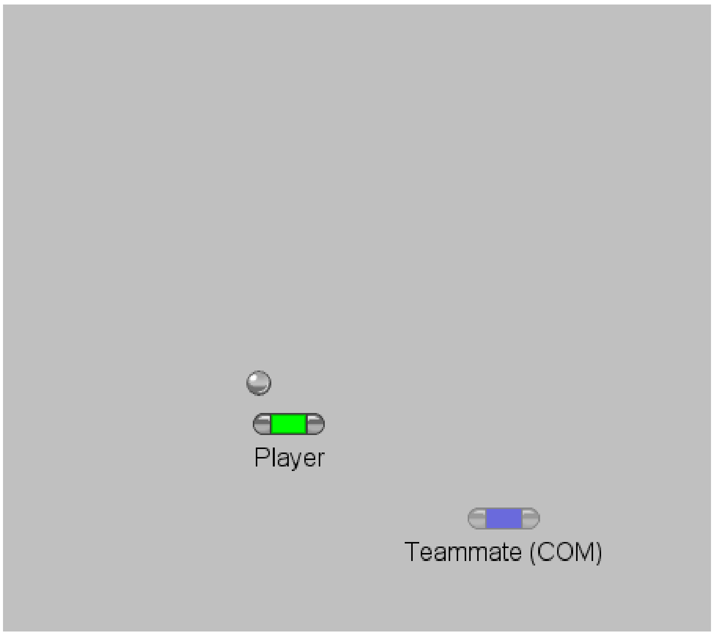


Figure 7.2: *Return the Ball* gameplay screen

The front and back paddles are of similar size and shape, but differ in the following way: the front paddle is more opaque while the back paddle is more transparent (see Figure 7.2), and this corresponds to the ability of the paddles to return the ball successfully. A ball falling from the top of the screen will bounce off the front paddle back towards the top. Whenever the ball has bounced off the front paddle, the front paddle will become transparent and unable to move, until the ball has spawned in the new position and proceeded to start falling again. With every ball that is successfully returned by the front paddle, one point was added to the team's cumulative total score that was shown to them at the end of each game played. Also, each successful return of the ball increases the speed of the ball as it drops after spawning to be 10% faster than the previous falling speed.

The back paddle is more transparent than the front, and is able to move all the time. However, it is unable to prevent the ball from passing through towards the bottom even though the player controlling the back paddle tries to position it below the falling ball. While it may appear that the transparency of the back paddle would make it unable to contribute towards fulfilling the objective of the game, in actual fact it is able to determine where the ball will fall from next. This is because the spawn position for where the ball will next drop from is determined by the following conditions – if the difference in horizontal distance between the front and back paddles is more than 100 pixels, then the ball will spawn and drop in the midpoint of the x-coordinates of the front and back paddles. Otherwise, the ball will spawn 150 pixels to either the left or the right of the front paddle's x-coordinate (at the time of spawning).

In order to move the paddles, participants can use the left and right arrow keys to move their own paddles horizontally along the lines they have been placed. The team-mate's paddle moves automatically, based on the following situations:

- **Team-mate controlling front paddle:** The team-mate will move towards the ball in order to return it, with a random time delay before it starts moving.
- **Team-mate controlling back paddle (Random movement):** The team-mate will randomly choose between 3 courses of action – move towards x-coordinate 200, x-coordinate 400 or remain stationary.
- **Team-mate controlling back paddle (Setting up spawn):** The team-mate will move to 150 pixels to either the right or left of the participant-controlled front paddle.

7.1.3 Experimental Manipulation

The two different versions of RTB played by the participants were the INDEPENDENT and INTERDEPENDENT versions. In the INDEPENDENT version (played by 31 participants), participants were not told that the back paddle's position relative to the front paddle could determine where the ball would spawn next. This was to give them a particular understanding of RTB's mechanics – that there was no relationship between the actions of the front and back paddles, i.e. they were independent of one another. The algorithm for the team-mate controlling the back paddle was "Random movement", to ensure the team-mate would behave in a way that was consistent with the participants' understanding of this version of RTB.

To help inform their decision about whether to place themselves or their team-mate in the front paddle, participants had information about the performance of the team member in the front paddle (in the form of the score) and an idea of who has been sidelined by being placed in the back paddle (from the belief that only the front paddle is able to contribute).

In the Interdependent version (played by 30 participants), they were told about how the position of the back paddle relative to the front paddle could determine where the ball would spawn next, though they were not given the specific details of the spawning position, instead being told that the following: that the ball would fall somewhere between the front and back paddles if they were not too close. If not, the ball would fall from a random position. The algorithm for the team-mate controlling the back paddle was "Setting up spawn", to ensure the team-mate would behave in a way that was consistent with the participants' understanding of this version of RTB.

In this case, participants had information about the performance of the front paddle and also who has been been placed in the back paddle. The main difference between the INDEPENDENT and INTERDEPENDENT versions is that the interdependence between team members means the team member using the back paddle is no longer sidelined since they are able to contribute.

7.1.4 Study Session Protocol

Participants arrived at private testing room, and did not meet other participants. They were assured that their comments would be kept anonymous and confidential. At the start of the study, participants filled out the questionnaire that gathered the demographic information shown in Table 7.1. The researcher then proceeded to explain to them how the experiment would be run, as well as the game mechanics behind the game according to the respective versions they were assigned to play.

This was followed by playing a tutorial round to gain a better understanding of the game, and clarify any uncertainties that participants had about the game or the experiment. The tutorial round consisted of four games played with an AI team-mate called "Team-mate (AI)", which participants were told was the same AI team-mate as the one they would be playing with during the actual experiment. In these four games, participants played two of them using the front paddle and the other two using the back paddle, to get an understanding of what it was like to play in both positions.

In order for participants in the INTERDEPENDENT version to better understand the relationship between the front and back paddles, the tutorial for the INTERDEPENDENT version had a colored rectangle overlaid for the region that was between the two paddles (see Figure 7.3). This colored region grew and shrank in size according to the distance between the two paddles, and changed from yellow to red if the minimum distance between paddles had been violated (resulting in the random spawn location of the ball). This colored rectangle was not shown during the actual version of RTB that participants played, in order to to give them an unfair advantage over those in the INDEPENDENT condition.

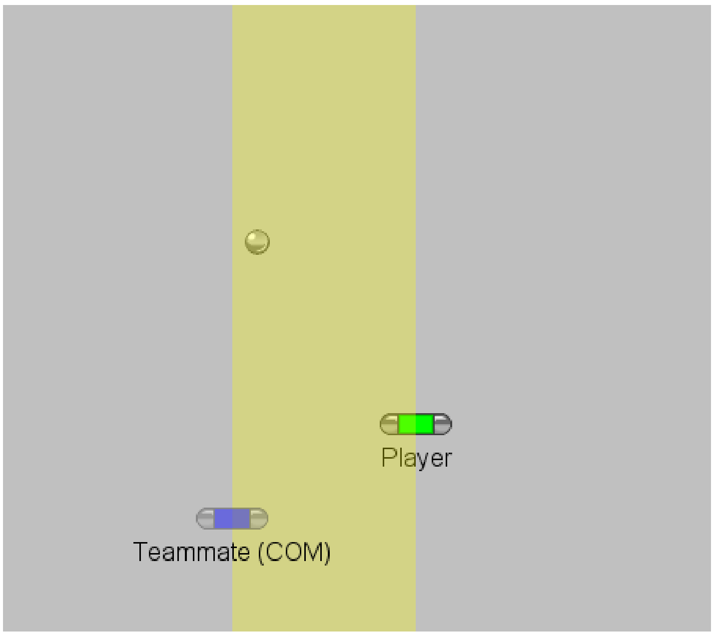


Figure 7.3: Tutorial for Interdependent condition

Once participants acknowledged that they understood the game, the actual experiment started. Each participant was assigned to play two rounds of seven games each – one round was played with the AI team-mate while the other was played with the PH team-mate. Counterbalancing via alternating the order of team-mates that participants played with was done to rule out order effects. The researcher informed participants that during the experiment they were not being assessed on their performance, and that information about their performance and choices was not being logged. These measures were to ensure that there would be no pressure to perform well would not influence which positions they decided to place themselves.

After each round of seven games, the researcher got participants to fill in a questionnaire that sought to gather information about their experience of the recently completed round of games.

There were no indications during any of the sessions that participants doubted the identity of the PH team-mate. No additional checks were conducted about this as pilot studies and previous studies have shown that participants make it quite clear if they doubt the identity of the team-mate.

7.1.5 Measures

The study gathered demographic data, game log data, self-reported quantitative data (5-point Likert scales and ranking questions)

The game logs captured the activity of each participant for every game played with both types of team-mates. For each game played, the logs detailed which team member was in the FRONT position, the number of hits that each team member was able to make, and total amount of time a team member had spent in the front position up till that point.

After each round of seven games, participants filled in a questionnaire about the round of games they had just completed.

In order to highlight any possible differences in perception of interdependence, participants were asked the following Likert-scale questions:

- I consider both of us as members of the same team
- I watched what the other player was doing when I was in the back position
- I watched what the other player was doing when I was in the front position
- What I do in the back position depends on what the other player does in the front position
- What I do in the front position depends on what the other player does in the back position

In order to gain better insight to any possible differences in position-choice, participants were asked them to pick the top 5 reasons (and rank them from 1 to 5) for how they chose their position in the 7 games with the other player. The reasons included:

- My ability to return the ball
- The other player's ability to return the ball
- My desire to try out both positions
- The position I feel I'm better at
- The position I feel the other player is better at
- Whatever helps the team do better
- No particular reason
- My desire to make sure both players get to play
- The position I prefer to play
- My desire to make sure I have fun
- My desire to make sure the other player has fun
- Randomly chose

In order to make a comparison between both types of team-mates, after completing both rounds of games, participants were asked to compare their PH and AI team-mates. To minimize the possibility of responses from the first round influencing their game-play choices in the second round, these questions were only asked at the end of both rounds, These questions included:

- How emotionally difficult was it for you to choose your position when playing with the human player?
- How emotionally difficult was it to choose your position when playing with the computer player?

The questions used in the questionnaire were not based on previously-used measures (it is not clear that such pre-existing measures exist), but the questions were refined for validity via feedback received during pilot testing.

7.2 Results

There are two main results of this study: (1) the manipulation of interdependence between team members reduces the difference in how often participants choose to use the front paddle with their PH compared to with their AI team-mate; (2) this manipulation also reduces the difference in emotional difficulty participants have in choosing whether to use the front paddle when playing with their PH team-mate compared to their AI team-mate.

7.2.1 Behavioral Results

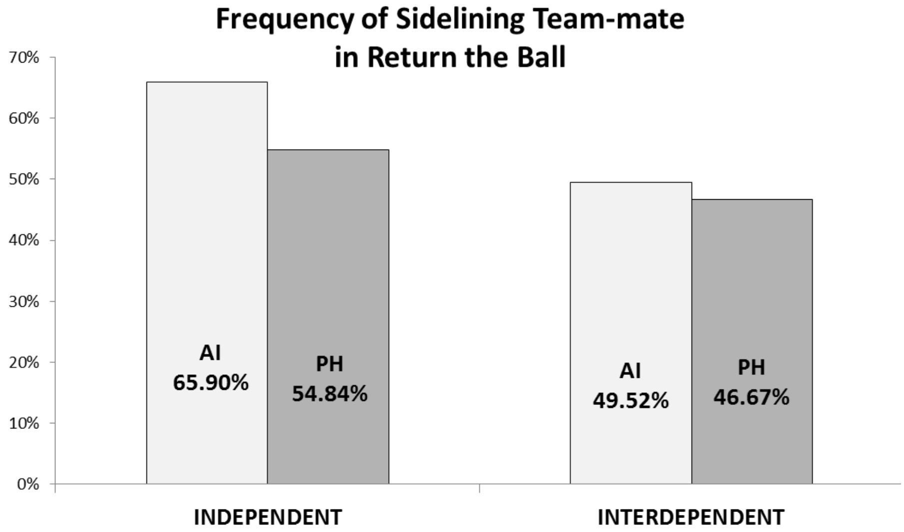


Figure 7.4: Comparison: percentage of games when team-mate is sidelined

In the INDEPENDENT version, participants chose to sideline (by choosing the front paddle) their AI team-mate (65.90% of the time) more often than their PH team-mate (54.84% of the time), as seen in Figure 7.4. A paired-samples t-test found this difference to be statistically significant ($p < 0.01$), $t(29)=4.08$.

In the INTERDEPENDENT version, participants sidelined their AI team-mate (49.52% of the time) almost as often as their PH team-mate (46.67% of the time), as seen in Figure 7.4. A paired-samples t-test found the difference to be not statistically significant ($p > 0.05$), $t(29)=0.95$.

When playing with the AI team-mate, participants playing the INDEPENDENT version of RTB sidelined their team-mate more often (65.90% of the time) than those who played the INTERDEPENDENT (49.52% of the time) version. An independent-samples t-test found this difference to be statistically significant ($p < 0.01$), $t(29)=3.62$.

When playing with the PH team-mate, participants playing the INDEPENDENT version sidelined their team-mate more often (54.84% of the time) as those who played the INTERDEPENDENT version (46.67% of the time). An independent-samples t-test found this difference to be statistically significant ($p < 0.05$), $t(39)=2.18$.

7.2.2 Emotional Difficulty Results

In the INDEPENDENT version of RTB, participants reported having more emotional difficulty in choosing their position with the PH team-mate (3.74

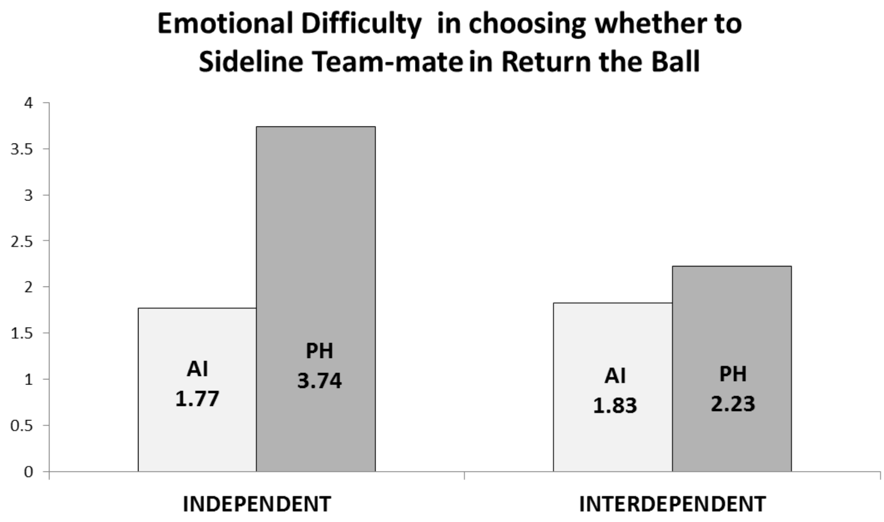


Figure 7.5: Comparison: Emotional Difficulty ratings of players

out of 10) than with the AI team-mate (1.77 out of 10). A paired-samples t-test found this difference to be statistically significant ($p < 0.01$), t(30)=-3.93.

In the INTERDEPENDENT version, participants reported having more emotional difficulty in choosing their position with the PH team-mate (2.23 out of 10) than the AI team-mate (1.83 out of 10). A paired-samples t-test found this difference to be not statistically significant ($p > 0.05$), t(29)=-1.84.

7.2.3 Performance-related results

The average number of balls returned by the team at the end of each game in when the participant either sidelines the team-mate (by choosing the front paddle) or not sidelining the team-mate (by choosing the back paddle) in each version of the game is shown in Table 7.2.

	INDEPENDENT		INTERDEPENDENT	
	AI	PH	AI	PH
Average Score (Front)	3.5	3.6	3.9	3.9
Average Score (Back)	4.5	4.0	4.8	4.9

Table 7.2: Team performance in RTB

7.2.4 Experience-related Results

Participants were asked a variety of questions in the self-reported questionnaires at the end of each round of games relating to their experience while playing each round of games.

Participants were asked a variety of questions in the self-reported questionnaires at the end of each round of games relating to their experience while playing each round of games.

To investigate whether participants were paying attention to the team-mate and what they were doing while using the back paddle, they were asked to (1) rate on a Likert scale to what extent they were watching their AI and PH team-mates while playing; (2) pick three statements that represented what they were doing while playing using the back paddle. Results of these two questions are are shown in Table 7.3.

Question	**INDEPENDENT**		**INTERDEPENDENT**	
	AI	**PH**	**AI**	**PH**
I watched what the other player was doing when I was in the back position	3.81	4.10	4.37	4.27
I watched what the other player was doing when I was in the front position	2.32	2.77	3.37	3.43
When I was playing in the back position with the other player, I:				
Was observing where the ball spawned	**20** (21.5%)	**20** (21.5%)	**20** (22.2%)	**17** (18.9%)
Not Applicable because I did not play in the back	**3** (3.2%)	**3** (3.2%)	**3** (3.4%)	**3** (3.4%)
Did not care what was going on in the front	**11** (11.8%)	**4** (4.3%)	**4** (4.4%)	**1** (1.1%)
Was observing what the other player was doing	**22** (23.7%)	**27** (29%)	**20** (22.2%)	**23** (25.6%)
Was observing how well the other player managed to return the ball	**16** (17.2%)	**18** (19.4%)	**15** (16.7%)	**17** (18.9%)
Was observing where the other player was	**20** (21.5%)	**19** (20.4%)	**28** (31.1%)	**29** (32.3%)
Other reasons	**1** (1.1%)	**2** (2.2%)	**0** (0%)	**0** (0%)
Total	**93**	**93**	**90**	**90**

Table 7.3: What participants were paying attention to while playing RTB

Participants in the INTERDEPENDENT condition reported that they paid more attention to what both types of team-mates were doing than those in the INDEPENDENT condition.

When they were playing using the back paddle, the four most frequently chosen reasons participants cited to describe what they were doing were observing the spawn position of the ball, observing what the other player was doing, observing how well the other player managed to return the ball, and observing where the other player was. This was true for games played with PH and AI team-mates in both versions of RTB.

7.2.5 Choice Rationale Results

In order to understand how participants made their choices, they were asked to select five reasons that they used to make their decisions out of a list of 13. Table 7.4 shows the frequency and percentage of how often a reason was chosen out of the total possible times that it could be picked

(155 in the INDEPENDENT version and 150 in the INTERDEPENDENT version).

Reason	INDEPENDENT AI	INDEPENDENT PH	INTERDEPENDENT AI	INTERDEPENDENT PH
Performance-Related				
My ability to return the ball	18	14	19	19
The other player's ability to return the ball	11	8	13	12
The position I feel I'm better at	13	9	24	23
The position I feel the other player is better at	5	9	10	11
Whatever helps the team do better	10	12	23	15
Participants' own interests				
The position I prefer to play	17	15	19	21
My desire to make sure I have fun	22	13	11	7
Team-mate's interest				
My desire to make sure both players get to play	9	19	2	7
My desire to make sure the other player has fun	2	8	1	2
Miscellaneous Reasons				
Randomly chose	12	12	8	8
No particular reason	14	15	5	7
My desire to try out both positions	22	21	15	18
Totals for each category of reasons				
Performance-Related reasons	**57** (36.78%)	**52** (33.55%)	**89** (59.34%)	**80** (53.33 %)
Participant own interest reasons	**39** (25.17%)	**28** (18.05%)	**30** (20.00%)	**28** (18.67%)
Team-mate's interest reasons	**11** (7.09%)	**27** (17.14%)	**3** (2.00%)	**9** (6.00%)
Miscellaneous reasons	**48** (30.96%)	**48** (30.96%)	**28** (18.06%)	**33** (22.00%)

Table 7.4: Reasons for choosing to play using front paddle

The most frequently chosen reasons are those related to performance of the team. These type of reasons are chosen more frequently in the INTERDEPENDENT version of the game compared to the INDEPENDENT version, with a difference in frequency of about 20% for both the PH and AI team-mates.

Miscellaneous reasons, which include experimenting with both positions ("My desire to try out both positions") and picking without any intention or purpose (e.g. randomly or with no particular reason) are the next most common type of reasons chosen. These type of reasons are chosen less frequently in the INTERDEPENDENT version of RTB compared to the INDEPENDENT version, and this is true with both PH and AI team-mates.

The third-most frequently chosen reasons are those that pertain to the participants' own interests such as preferred positions and making sure they get to play. These reasons are chosen almost as frequently in the INDEPENDENT and INTERDEPENDENT versions of the game.

The least frequently chosen reasons are those pertaining to the team-mates' interests. The frequency of choosing these reasons is lower in the INTERDEPENDENT version compared to the INDEPENDENT version. Furthermore the decrease in frequency of choosing these reasons is larger in games played with the PH team-mate (11 % decrease) than the AI team-mate (5 % decrease).

7.2.6 Perception-related Results

Participants were asked to answer a series of questions in the questionnaire that pertained to their understanding of RTB as a game, and their perceptions of how the game worked. The results for these questions are shown in Table 7.5.

Question	INDEPENDENT		INTERDEPENDENT	
	AI	PH	AI	PH
What I do when in the back position is depends on what the other player does in the front position	2.97	2.97	4.17	4.37
What I do when in the front position is depends on what the other player does in the back position	1.42	1.48	2.90	3.07
When comparing between the Front and Back positions				
Only the player in the Front is able to influence the outcome of the game	**31** (100.0%)		**1** (3.3%)	
Only the player in the Back is able to influence the outcome of the game	**0** (0.0%)		**3** (10%)	
Both Front and Back player are able to influence the outcome of the game	**0** (0.0%)		**26** (86.7%)	
Total	**31**		**30**	
Between the Front and Back positions				
The Front position is more important	**29** (93.5%)		**4** (13.3%)	
The Back position is more important	**0** (0.0%)		**7** (23.4%)	
Both positions are equally important	**2** (6.5%)		**19** (63.3%)	
Total	**31**		**30**	

Table 7.5: Participants' perceptions of the relationship between paddles in RTB

As seen from Table 7.5, the extent to which participants agree with the statement suggesting that their actions while using the back paddle depend on those of the front paddle are higher in the INTERDEPENDENT version than the INDEPENDENT version of RTB.

In terms of the ability of the different paddles to influence the outcome of the game, all participants playing the INDEPENDENT version reported that the front paddle is more important. The majority of the participants also reported that the front paddle is much more important in determining the outcome of the game than the back paddle.

In the INTERDEPENDENT version, the majority of participants reported that both front and back paddles are able to influence the outcome of the game. The majority of participants playing the INTERDEPENDENT version also reported that both paddles were equally important in determining the outcome of the game.

Table 7.6 indicates participants' idea of whether the respective versions of RTB are considered team-mate games, and which type of game participants think they resemble.

Question	INDEPENDENT	INTERDEPENDENT
I consider myself and the other player members of the same team	3.37	4.25
Return the Ball is the most similar to:		
Members of a bowling team who each take turns bowling individual games and then add their individual scores to a total team score. All members play, but only one at a time.	**6** (19.4%)	**1** (3.3%)
Members of 4x100m baton relay team who pass the baton from one member to the next during a race. All team members participate, but only during the segment when they are carrying the baton.	**1** (3.2%)	**2** (6.7%)
Members of a tennis doubles team who each cover for each other on different parts of the court. Both members play at all times, but only one returns the ball each time.	**8** (25.8%)	**16** (53.3%)
Two members of a volleyball team who set the ball up for one another to hit the ball towards their opponents. Each member participates, but they do different things at different times.	**0** (0.0%)	11 (36.7%)
Members of a cricket team (or a baseball team) when it is their team's turn to bat. Whether a team-member gets to bat depends on where they are on the list of batters -- and whether too many of their team-mates are "struck out" before they get to bat.	**6** (19.4%)	**0** (0.0%)
The captain and his substitute on a football team. Only one of the team members plays at one time and it is the decision of the captain which one gets to play -- but either/both player's goals contribute to the team's overall score.	**10** (32.2%)	**0** (0.0%)
Total	**31**	**30**

Table 7.6: Participants' impressions of RTB as a team-based game

In both the INDEPENDENT and INTERDEPENDENT versions, participants reported that they considered themselves and the other player (their team-mate) to be members of the same team. Participants considered themselves and the other player to be more of a team in the INTERDEPENDENT version than the INDEPENDENT version.

In the INDEPENDENT version of RTB, the most frequently-chosen scenario that RTB is similar to is the football scenario, followed by the tennis doubles scenario. In the INTERDEPENDENT version, the most frequently-chosen scenario that RTB is similar to is the tennis scenario, followed by the volleyball scenario.

7.3 Discussion

This section discusses the results of the study, with the use of the trade-off calculations and Index of Dependence to provide support in explaining the behavior and emotional difficulty results.

The Index of Dependence for both versions of RTB show that the manipulation has not resulted in a significant change in the interdependence between team members. However, participants do feel that the team members are more interdependent with one another, and this is also demonstrated in how there is increased Joint Control in determining the game's outcome in the INTERDEPENDENT version of RTB compared to the INDEPENDENT version.

The calculations of the trade-off values show that the manipulation of scoring information has reduced the difference in utility that participants gain when choosing to sacrifice the team-mate. This mirrors the behavioral results of the study.

The calculation of the difficulty in choosing between the two options illustrates that the manipulation of interdependence between team members has reduced the difference in difficulty of choosing with the AI and PH team-mates, and has also eliminated the presence of a trade-off situation since the option to play using the back paddle brings much more benefits to the participant. This result helps to explain why participants have more similar levels of emotional difficulty in choosing the position in the INTERDEPENDENT version of RTB compared to the INDEPENDENT version.

7.3.1 Perceptions of RTB as a team-mate game

The INDEPENDENT and INTERDEPENDENT versions of RTB may seem drastically different as computer games, since participants have very different understandings of how each version of the game works. In fact, it could even be argued that the INDEPENDENT version of RTB (because of the lack of involvement of the player using the back paddle) cannot be considered to be a team-mate game at all, since only team member can play at any point of time. While this may appear to be the case if one considers only each standalone game of RTB in the INDEPENDENT version, it should be noted that participants are told that a complete round of RTB consists of a series of 7 games played consecutively with one team-mate. This is very much similar to decision-making research involving iterated Prisoners' Dilemma games where each participant plays multiple games with the same player consecutively.

Also, the results in Table 7.6 suggest that participants do actually consider the INDEPENDENT version of RTB to be a team-mate game, providing a rating of 3.37 out of 5 for the Likert scale statement I consider myself and the other player members of the same team. Furthermore, to further validate their idea of RTB as a team-mate game, the results of the question asking them which type of team sports scenario that the INDEPENDENT version of RTB resembles, participants indicated that the football, tennis, bowling and baseball / cricket examples were most representative. From this, it can be inferred that to a degree, participants do consider the INDEPENDENT version of RTB to be a team-mate game.

7.3.2 Calculating the Index of Dependence (ID)

Calculations of the Index of Dependence (ID) for DTP indicate that between the INDEPENDENT and INTERDEPENDENT versions of the game, the values of the ID are about the same. While this suggests that the level of interdependence between team members was not affected by the manipulation, perception-based results and the values of Joint Control suggest that in fact that the manipulation affected the interdependence between team members in the INTERDEPENDENT version of RTB compared to the INDEPENDENT version.

Using the 2x2 outcome matrix in Interdependence Theory, it is possible to represent the relationship between team members in RTB. The number of balls returned successfully by the team at the end of each game is used to represent the outcomes as a result of the team members playing together. Since the study involving RTB intends to manipulate the level of Interdependence between team members, the outcome matrices for one example for each version of the game will be shown. A full list of all the matrices can be found in Appendix B.

For the INDEPENDENT version where the participant plays using the front paddle with an AI team-mate, the relationship between team members can represented as shown in Table 7.7.

Participant uses Front paddle with AI team-mate (INDEPENDENT version)		
	Participant successful in returning ball	Participant unsuccessful in returning ball
Team-mate successful in coordinating with participant	3.5	0
Team-mate unsuccessful in coordinating with participant	3.5	0

Table 7.7: Representation of relationship when participant uses front paddle with AI team-mate in INDEPENDENT version)

Since participants are led to believe that the back paddle does not have any impact on the outcome of the game, then whether or not the team-mate is successful or unsuccessful in coordinating with the participant, the score will be dependent solely on whether the participant successfully returns the ball. If the participant does, then the output will be 3.5, and if the participant fails, the output will be 0.

In the case of the INTERDEPENDENT version, the 2x2 matrix is slightly different. Assuming the same situation of the participant playing using the front paddle with the AI team-mate, the relationship between the participant and team-mate can be represented as shown in Table 7.8.

Participant uses Front paddle with AI team-mate (INTERDEPENDENT version)		
	Participant successful in returning ball	Participant unsuccessful in returning ball
Team-mate successful in coordinating with participant	3.9	0
Team-mate unsuccessful in coordinating with participant	3.5	0

Table 7.8: Representation of relationship when participant uses front paddle with AI team-mate in INTERDEPENDENT version)

In the INTERDEPENDENT version, since participants are made aware that the back paddle is able to influence the position of the ball's spawning, it would be more effective if participants and their team-mates cooperate to ensure the best possible outcome. Therefore if the team-mate (back) is successful in coordinating with the participant (front), who manages to successfully return the ball, then the team will have an outcome of 3.9.

However, if the team-mate (back) is unsuccessful in coordinating with the participant (front), the team will have an outcome of 3.5, similar to that of the INDEPENDENT version. The reason for this is that failure to coordinate in the INTERDEPENDENT version of RTB is similar to playing the INDEPENDENT version of RTB (where participants do not believe there is a need for them to coordinate anyway).

Although the front and back paddles need to coordinate with one another in influencing the spawn location of the ball, it is still up the front paddle to hit the ball to continue the game. This is why regardless of whether the team-mate coordinates successfully with the participant, as long as the participant fails to successfully return the ball, the team's outcome will be 0.

With this data, the Actor Control (AC), Partner Control (PC), Joint Control (JC) and resultant Index of Dependence (ID) can be calculated for all 8 scenarios for RTB, as seen in Table 7.9.

As seen from Table 7.9, comparing the IDs in the INDEPENDENT and INTERDEPENDENT versions suggest that any differences in the IDs as a result of the manipulations are not significant. At first glance, this would suggest that the manipulation did not affect the level of interdependence

	INDEPENDENT		INTERDEPENDENT	
	AI	PH	AI	PH
Participant uses Front paddle				
Actor Control	3.50	3.60	3.70	3.75
Partner Control	0	0	0.20	0.15
Joint Control	0	0	0.20	0.15
Participant uses Back paddle				
Actor Control	0	0	0.15	0.45
Partner Control	4.50	4.00	4.65	4.45
Joint Control	0	0	0.15	0.45
Index of Dependence (ID)				
ID (Front)	0	0	0.01	0.02
ID (Back)	1	1	0.99	0.98

Table 7.9: Actor Control, Partner Control, Joint Control and Index of Dependence for 8 possible scenarios in RTB

between team members. In both versions of the game, the ID values for the participant using the front paddle suggest that the outcome is dependent largely (if not entirely) on the participant, since the value is 0 or close to 0.

Similarly, the ID values for the participant using the back paddle suggest that the outcome is dependent largely (if not entirely) on the team-mate (who is controlling the front paddle). These results are logical because in both versions of the game, the design is such that only the front paddle can return the ball, and hence the team's outcomes depend primarily on the front paddle's ability to return the ball successfully.

Upon closer examination however, examining the values for Joint Control suggest that there has been some effect of the manipulation on interdependence. As Joint Control (JC) represents how the actions of both team members affect the outcome, the JC values illustrate that in the INDEPENDENT version, there is no JC between team members, while in the INTERDEPENDENT version, there is some degree of Joint Control. This is seen in the situation with both AI and PH team-mates, and also when participants are playing using either the front or back paddle.

Additionally, when looking at the Actor Control (AC) values when participants play using the back paddle, in the INDEPENDENT version of the game, the AC values for participants is 0. In the INTERDEPENDENT version however, it is clear that participants playing using the back paddle have some degree of AC though it is small.

Though the ID values for the INTERDEPENDENT version of RTB do not suggest that the manipulation has actually managed to modify the level of interdependence between team-mates for each option, results in Table 7.5 suggest otherwise.

The results in Table 7.5 give an indication of how participants perceive the relationship between the two paddles in the respective versions of RTB. As seen from the table, when asked about whether what they do when playing in the back position depends on what is done in the front position, participants reported that they agreed more with this statement in the INTERDEPENDENT version than the INDEPENDENT version.

Also, when asked about their perceptions of about the importance of both positions and the influence of each position on the game, it is clear that in the INDEPENDENT version, participants feel that it is the front paddle is clearly more important and able to influence the outcome more than the back paddle. However in the INTERDEPENDENT version, the majority of the participants believe that both positions are equally important and have equal influence in determining the outcome of the game, contrary to what is suggested by the IDs for each paddle that the participant uses in the INTERDEPENDENT version. These ID values are still valid because ultimately according to the design of RTB, it is the front paddle that determines whether the game continues or ends.

One possible reason why the IDs do not accurately reflect the change in the level of interdependence between team members between the two versions of RTB is what these IDs measure with regards to the design of RTB as a game. It is clear in RTB that only the front paddle is able to return the ball, and so in terms of output for the team in terms of points scored a participant will depend almost entirely on whoever is using the front paddle. However, as raised earlier, while the overall IDs remain similar between both versions, the Joint Control differs.

This is consistent with the overall idea in RTB's game design – though both team members are able to contribute towards achieving the goal of the team, only the front paddle is able to keep the game going by successfully returning the ball. This is possibly why some participants in the INTERDEPENDENT condition still consider the front paddle to be more important, since failure of the front paddle to return the ball means the end of the game.

7.3.3 Weightages of Attributes

In RTB, participants are faced with a trade-off between attributes they may deem important – team performance and the sidelining of a team member.

As seen from Table 7.4 when ignoring miscellaneous reasons for choosing a particular position in RTB, the most commonly-chosen reasons are those relating to performance, followed by those relating to participant's own interests, and finally those relating to the team-mate's interests. These results suggest the following:

- Participants' consider the team's performance to be the most important category of reasons when choosing which position to assign.
- Participants' consider their own interests more important than their team-mate's interests when choosing which position to assign.

Therefore, the weightages for the different attributes related to the trade-off could be assigned in a manner shown in Table 7.10, to reflect the importance placed on the different attributes.

	SCORE		NO-SCORE	
	AI	**PH**	**AI**	**PH**
Weightage-Performance	0.5	0.5	0.5	0.5
Weightage-Self	0.4	0.3	0.4	0.3
Weightage-Team-mate	0.1	0.2	0.1	0.2

Table 7.10: Weightages for RTB

7.3.4 Choosing between options in RTB

The different options that participants get to choose between in each version of RTB with the respective team-mate (see Table 7.11) can represented in the format [3], with the assigned weightages for each attribute involved. The percentages gains and losses in each attribute as a result of choosing one option over the other are included in parenthesis.

Version	Team-mate	Decision	Wt Team	Wt Self	Wt Team-mate	Score	Participant Participate	Team-mate Participate
IND	AI	Front	0.5	0.4	0.1	3.5 (-12.50%)	1 (+100%)	0 (-100%)
		Back	0.5	0.4	0.1	4.5 (12.50%)	0 (-100%)	1 (+100%)
	PH	Front	0.5	0.3	0.2	3.6 (-5.26%)	1 (+100%)	0 (-100%)
		Back	0.5	0.3	0.2	4 (5.26%)	0 (-100%)	1 (+100%)
INT	AI	Front	0.5	0.4	0.1	3.9 (-10.34%)	1 (+100%)	0 (-100%)
		Back	0.5	0.4	0.1	4.8 (10.34%)	0 (-100%)	1 (+100%)
	PH	Front	0.5	0.3	0.2	3.9 (-11.36%)	1 (+100%)	0 (-100%)
		Back	0.5	0.3	0.2	4.9 (11.36%)	0 (-100%)	1 (+100%)

Table 7.11: Gains and Losses for each choice in RTB

7.3.5 Explaining Behavioral Results

According to the behavioral results, participants playing in the INDEPENDENT version of RTB played using the front paddle more frequently with their AI team-mates than their PH team-mates. In the INTERDEPENDENT version, the frequency at which they played using the front paddle with their AI and PH team-mates was about the same. This suggests that the manipulation of the level of interdependence between team-mates was able to moderate the difference in how frequently participants chose the front paddle with their AI team-mates compared to with their PH team-mates.

The rationale for such behavior can be examined in terms of gains or losses participants feel that the decision of choosing the front or back paddle will bring them. Using the formula proposed in MAUT, the percentage gains / losses in participants' utility from choosing to either play using the front paddle is shown in Table 7.12.

	INDEPENDENT		INTERDEPENDENT	
	AI *Weight	PH*Weight	AI*Weight	PH*Weight
Percentage Gain / Loss for Team Performance	-12.5*0.5	-5.26*0.5	-10.34*0.5	-11.36*0.5
Percentage Gain / Loss for Self	100.00*0.4	100.0*0.3	0.00*0.4	0.00*0.3
Percentage Gain / Loss for Team-mate	-100.00*0.1	-100.00*0.2	0.00*0.1	0.00*0.2
Overall Gain / Loss for using Front paddle	**23.75**	**3.75**	**-5.17**	**-5.68**
Difference between PH and AI team-mate	**20.00**		**0.51**	

Table 7.12: Gains and Losses for choosing to play in front in RTB

In the INDEPENDENT version of RTB, Table 7.12 shows that participants felt that they would gain much more utility from playing using the front paddle with the AI team-mate compared to the PH team-mate. The large difference in utility gained for using the front paddle with the AI team-mate compared to the PH team-mate mirrors the significant difference in frequency of choosing the front paddle in the INDEPENDENT version of RTB. This helps explain the behavioral results because logically, if a participant feels they are likely to have larger gains when playing with an AI team-mate than a PH team-mate, then they would most probably opt to play using the front paddle more with the AI team-mate than the PH team-mate.

In the INTERDEPENDENT version of RTB, the table shows that participants felt they would gain only slightly more utility when playing with the AI team-mate than the PH team-mate if they chose to play using the front paddle. Surprisingly, participants felt that they would actually lose utility if they were to play using the front paddle instead of playing using the back paddle.

These results – the small difference in utility lost when choosing to play using the front paddle and the overall losses when choosing to play using the front paddle – mirror those of the behavioral results in the game logs. Firstly, how often participants opt to play using the front paddle with AI team-mates and PH team-mates is about the same. Secondly, the game logs show that overall, participants chose to play using the front paddle less than half the time (49.52% of the time with AI and 46.67% of the time with PH, as seen in Table 7.4). Considering that participants feel that overall they lose when using the front paddle in the INDEPENDENT version, it would thus be logical that they opted to play using the front paddle less frequently.

7.3.6 Explaining Emotional Difficulty Results

The Emotional Difficulty results in this study show that participants playing the INDEPENDENT version had reported having significantly more Emotional Difficulty in choosing their position with their PH team-mate compared to when playing with their AI team-mate. In the INTERDEPENDENT condition however, this difference was not significant – participants reported having about the same amount of Emotional Difficulty in choosing their position with both types of team-mate. These results suggest that the manipulation of level of interdependence between team members was able to moderate the difference in Emotional difficulty participants had in choosing their position with PH and AI team-mates.

The Emotional Difficulty results can be analysed in the context of the overall gains and losses in utility participants would have in choosing each option, and how close the two options were in terms of utility – which determines how difficult it is to choose between them. The calculations of the gains and losses in utility, using Chatterjee and Heath's formula [3] is shown in Table 7.13. A positive value for the difficulty of choice indicates that the choice was tending towards playing using the front paddle.

Based on the data in Table 7.13, it suggests that the difference in difficulty of the choice participants had to make with AI and PH team-mates was larger in the INDEPENDENT version than the INTERDEPENDENT version. This mirrors the results of the difference in Emotional Difficulty ratings for this version of the game. This suggests that the manipulation of interdependence between team members was able to moderate this differ-

	INDEPENDENT		INTERDEPENDENT	
	AI	**PH**	**AI**	**PH**
Gain from using Front paddle	0.4	0.3	0	0
Loss from usinig Front paddle	0.45	0.59	0.17	0.18
Gain from using Back paddle	0.18	0.24	0.07	0.07
Loss from using Back paddle	1.00	0.75	0	0
Overall Utility from using Front paddle (Gain - Loss)	-0.05	-0.29	-0.17	-0.18
Overall Utility from using Back paddle (Gain – Loss)	-0.82	-0.51	0.07	0.07
Difficulty of choice: Utility of Back – Utility of Front	**0.77**	**0.22**	**-0.24**	**-0.25**
Difference in difficulty of choice for AI and PH team-mates (PH – AI)	**0.55**		**0.01**	

Table 7.13: Difficulty in choosing between options in RTB

ence in Emotional Difficulty participants had when choosing their positions with AI and PH team-mates.

7.3.7 Elimination of Trade-offs

The results in Table 7.13 also highlight something else about the effect of increasing the level of interdependence between team members – the trade-off in choosing between the two positions has been eliminated. When looking at the gains and losses for choosing the front and back paddles in the INTERDEPENDENT version of RTB, it is clear that opting to use the front paddle gives participants only losses in utility (but no gains). Opting to use the back paddle in this version of RTB gives participants only gains in utility (but no losses).

As such, there does not seem to be a trade-off between attributes that exists in the choices available to participants in the INTERDEPENDENT version of RTB – it makes more sense to opt to play using the back paddle more than it does to use the front, because if a participant is thinking in terms of utility gained, using the front paddle is disadvantageous.

It can thus be inferred that in this case, the manipulation of the level of interdependence between team members in RTB has been able to eliminate the trade-off that exists between the attributes of team performance and sidelining a team member. This is because the increased level of interdependence between team members has managed to remove the losses that would arise from sidelining a team member.

7.4 Summary

This chapter did the following:

- Provided a detailed description of the study design regarding the participants, study protocol and data measured

- Described the game used for the study (*Return the Ball*) and the experimental manipulations used to examine the research question
- Presented and analysed the behavioral results: the increase in interdependence between team members reduced the difference in how often participants chose to play using the front paddle with their AI team-mate compared to their PH team-mate when comparing the INDEPENDENT and INTERDEPENDENT versions.
- Presented and analysed the emotional difficulty results: the increase in interdependence between team members reduced the difference in emotional difficulty that participants felt in choosing to play using the front paddle compared to their AI team-mate when comparing the INDEPENDENT and INTERDEPENDENT versions.

Chapter 8

Comparative Analysis

This chapter analyses the results of both studies, comparing the behavioral and emotional difficulty results of Study 01 (*Defend the Pass*) against that of Study 02 (*Return the Ball*). It explains why the manipulation in *Defend the Pass* only moderates the behavioral differences while the manipulation in *Return the Ball* moderates behavioral and emotional difficulty differences. This is because the manipulation in *Return the Ball* eliminates the trade-off situation but the manipulation in *Defend the Pass* does not. The analysis provides designers with three insights that can help in the design of cooperative games– designers should (1) Address both behavioral and emotional difficulty differences; (2) Address the trade-off situation and (3) Exercise caution when manipulating interdependence. This chapter also addresses any potential objections that could be raised regarding the studies conducted for this thesis.

8.1 Analysis of Games – DTP vs RTB

This section examines the behavioral and emotional difficulty results of the two studies and compares them with one another. The effect of the manipulations on behavioral results of each study are because they have an impact on information that participants use in their decision-making process.

The manipulation in RTB is able to reduce the difference in emotional difficulty because it changes the nature of the choice that needs to be made, by eliminating an attribute (sidelining the team-mate) that contributes to the trade-off being present. The manipulation in DTP does not reduce the difference in emotional difficulty because it does not affect an attribute that determines the nature of the choice made – participants playing the SCORE and NO-SCORE versions are reminded of their decision to sacrifice their team-mate.

8.1.1 Behavioral Responses

In the two studies, the respective manipulations were able to moderate participants' behavior. In Study 01, the manipulation of the visibility of scoring information in DTP reduced the behavioral difference in how often the team-mate was sacrificed when comparing rounds played with the PH and AI team-mates (see Figure 8.1). In Study 02, the manipulation of the interdependence between team-mates in RTB reduced the behavioral difference in how often the participant chose to play using the front paddle when comparing rounds played with the PH and AI team-mates (see Figure 8.2).

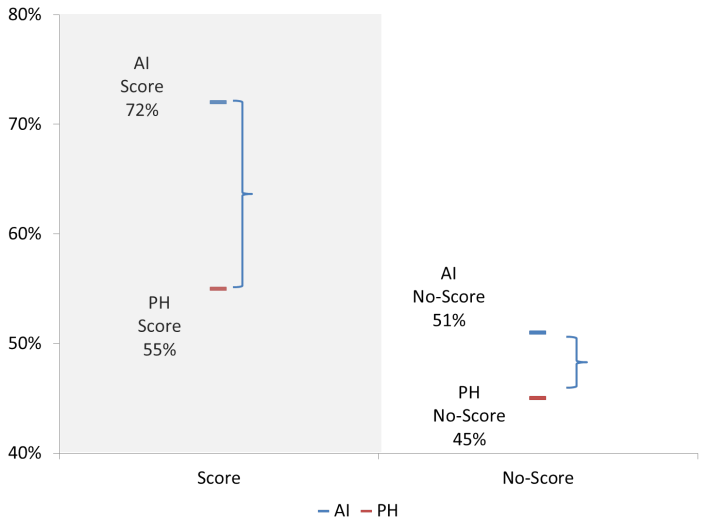


Figure 8.1: Behavioral responses in Score and No-Score versions of DTP

In both studies, it is also clear that the manipulations only have an effect on the behavioral responses towards the AI team-mate, which was confirmed by conducting Independent samples t-tests. These tests showed that there were statistically significant differences in how participants responded towards the AI team-mate in the DTP's SCORE vs NO-SCORE and RTB's INDEPENDENT vs INTERDEPENDENT versions of the games. Such differences were not observed with the PH team-mate in both studies.

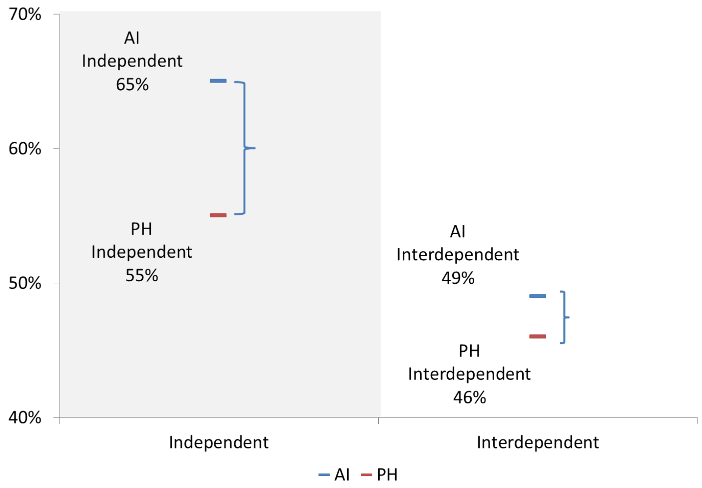


Figure 8.2: Behavioral responses in Independent and Interdependent versions of RTB

8.1.2 Emotional Difficulty

In the two studies, only the manipulation of interdependence between team-mates in RTB reduced the emotional difficulty difference when comparing rounds played with the PH and AI team-mates (see Figure 8.4). The manipulation of visibility of scoring information in DTP was not able to reduce the emotional difficulty difference when comparing rounds played with the PH and AI team-mates (see Figure 8.3).

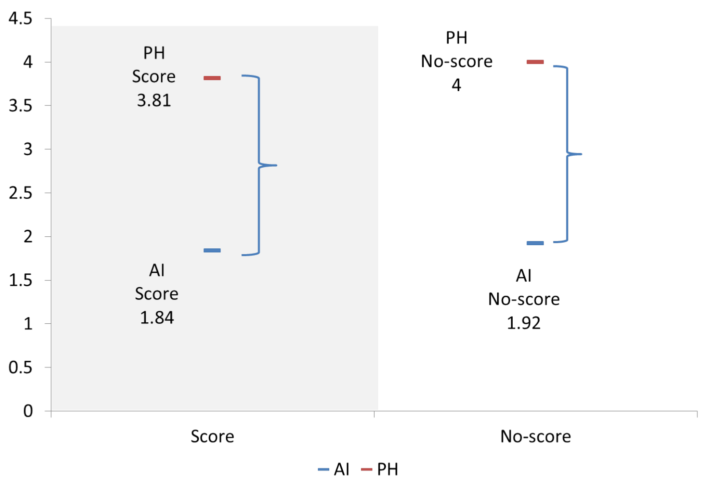


Figure 8.3: Emotional Difficulty in Score and No-Score versions of DTP

In RTB, it is clear that the manipulation of interdependence between team-mates only has an effect on the emotional difficulty when playing with the PH team-mate, which was confirmed with the by conducting an Interdependent Samples t-test. This test showed that there was a statistically significant difference in participants' emotional difficulty with the

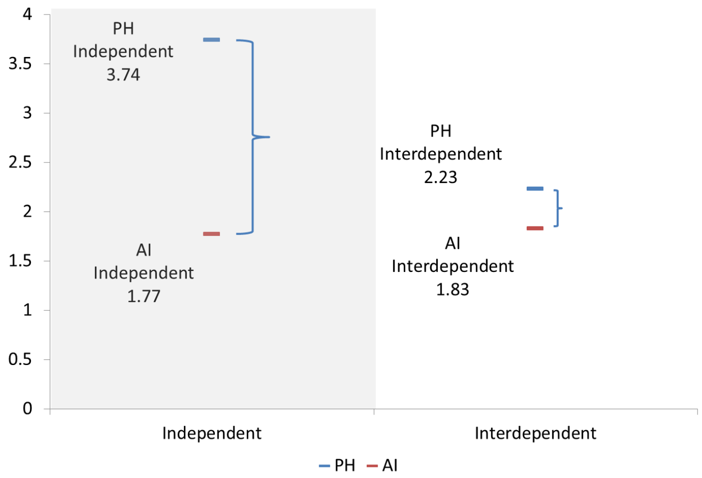


Figure 8.4: Emotional Difficulty in Independent and Interdependent versions of RTB

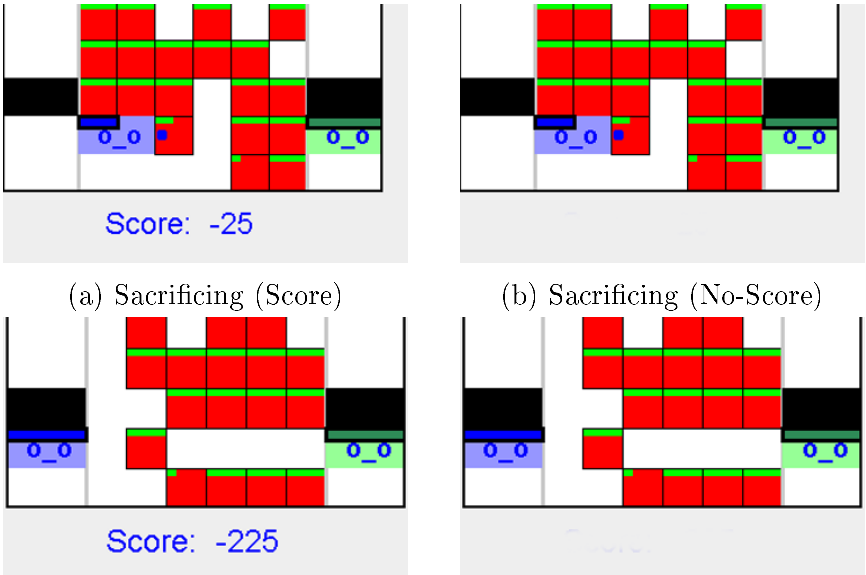
PH team-mate in the INDEPENDENT vs INTERDEPENDENT versions of the games.

8.1.3 DTP: Explaining the Results

The manipulation of the visibility of scoring information is able to moderate the difference in behavior with PH and AI team-mates in the NO-SCORE version compared to the SCORE version because participants have one less attribute that they can consider in the decision-making process. The manipulation is unable to moderate the differences in emotional difficulty in the NO-SCORE version compared to the SCORE version because participants are reminded about whether the team-mate will survive or be killed regardless of whether scoring information is visible or not.

In DTP's SCORE version, the two factors that influence participants' decisions about whether to sacrifice their team-mate are succeeding at the game and the survival of the team-mate. These factors are in conflict with one another, and in each option, a participant can choose between (sacrifice or protect), with a trade-off between these two factors. In order to choose between sacrificing and protecting the team-mate, they use some available information – performance information (in terms of score and estimation of monsters killed) and status of the team-mate (sacrificed or protected).

In the NO-SCORE version, the removal of scoring information reduces the behavioral differences when comparing between PH and AI team-mates. This is because by removing the score, participants only have the estimation of the number of monsters killed to give an indication of team performance. This reduction in visual information impacts their ability gauge the conse-



(a) Sacrificing (Score) (b) Sacrificing (No-Score)

(c) Protecting (Score) (d) Protecting (No-Score)

Figure 8.5: Comparison of decisions in SCORE and NO-SCORE version of DTP

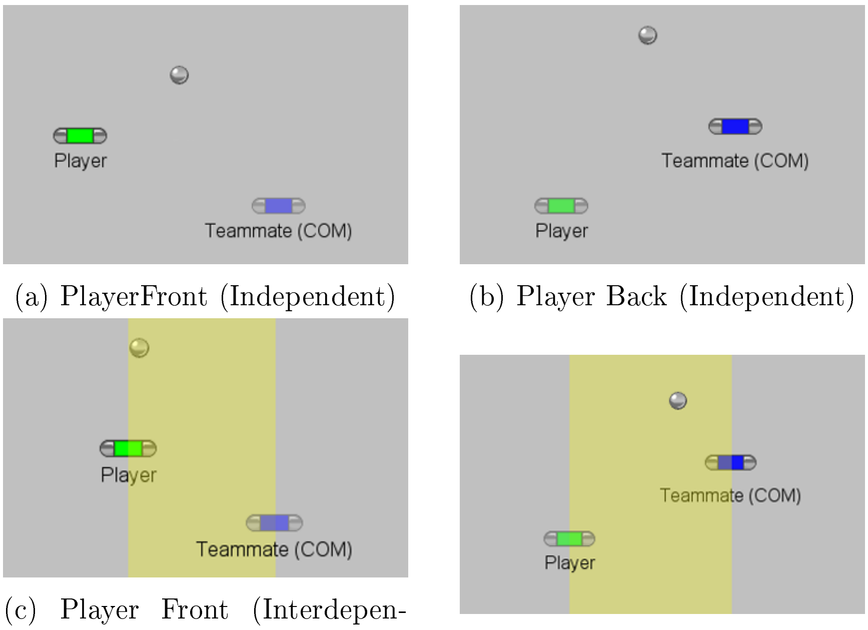
quences of their choices on performance, and would affect how often they decided to sacrifice their team-mate.

The manipulation of visibility of scoring information does not reduce emotional difficulty differences when comparing between PH and AI team-mates in the NO-SCORE version of DTP. This is because even though the information about the score has been removed, participants still have a clear reminder about the status of the team-mate as a consequence of their choice. As a result, the trade-off present in the SCORE version of DTP is also present in the NO-SCORE version, explaining why the difference in emotional difficulty is not reduced.

8.1.4 RTB: Explaining the Results

The manipulation of the level of interdependence between team members is able to moderate the difference in behavior with PH and AI team-mates in the INTERDEPENDENT version compared to the INDEPENDENT version because the understanding about the roles of both paddles changes – both paddles can now contribute to the team's success rather than having to only depend on the front paddle. The manipulation is able to moderate the emotional difficulty differences in the INTERDEPENDENT version of RTB because by increasing the level of interdependence between team members, it eliminates the possibility of a team member getting sidelined when they are assigned to then back paddle. This removes the trade-off situation, and no longer makes it more emotionally difficult to choose which paddle to use when playing with the PH team-mate compared to the AI team-mate in the INTERDEPENDENT version.

In RTB's INDEPENDENT version, the two factors that influence participants' decision about whether they put themselves or their team-mate in front paddle are succeeding at the game and whether a team member is being sidelined. These two factors are in conflict with one another, and each option a participant can choose between (own use of front paddle or team-mate use of front paddle), there is a trade-off between these two factors. In order to choose between taking the front paddle themselves or letting the team-mate take the front paddle, they use some available information – performance information (in terms of number of balls successfully returned) and which team member is being sidelined by using the back paddle.



(a) PlayerFront (Independent)

(b) Player Back (Independent)

(c) Player Front (Interdependent)

(d) Player(Interdependent)

Figure 8.6: Comparison of decisions in INDEPENDENT and INTERDEPENDENT version of RTB

In the INTERDEPENDENT version of RTB, the interdependence between team members reduces behavioral differences when comparing between rounds played with PH and AI team-mates. This is because by making the team members interdependent, it suggests although only the front paddle is able to hit the ball, the back paddle is able to make the front paddle's task easier. This suggests to players that the paddles embody different roles that both contribute towards overall collective performance. The change in their understanding of the nature of the two paddles would then affect how often they chose to use the front paddle.

The manipulation of interdependence between team members in the INTERDEPENDENT version of RTB reduces the difference in emotional difficulty when comparing rounds played with PH and AI team-mates. This is because by making team members interdependent, both paddles are now able to contribute towards overall collective performance, without either of the team members getting sidelined. The information available to participants about performance and the team member in the back paddle remains the same in both versions of RTB. However, the information about which team member is being sidelined by using the back paddle is removed – making the two paddles interdependent results in neither team member being sidelined. Furthermore, since the back paddle is able to make it easier for the front paddle to return the ball successfully, the likelihood of succeeding in the game is higher. As a result, the trade-off between succeeding at the game and a team member being sidelined is eliminated, explaining why the difference in emotional difficulty is reduced.

8.2 3 Insights for Designing Real-time Cooperative Games

The findings and analyses of the study results found that in order for game designers to reduce the difference in how players respond towards human and AI team-mates, they need to consider both the behavioral and emotional difficulty aspects of decisions they present players with.

The analyses also show that addressing the trade-off situation is necessary if a designer wants to moderate this difference in responses. Lastly, while moderating the level of interdependence between team members is a good means of removing trade-off situations, designers must be aware that even seemingly simple changes to the interdependence between team members can change the nature of the relationship between them in a variety of ways.

8.2.1 Address Behavioral and Emotional Difficulty Differences

Designs that seek to ensure that the difference in players' experiences of a game with human and AI team-mates must not only address behavioral differences but emotional difficulty differences as well. This is because in situations where players have to make decisions involving their team-mates, it is necessary to examine not only what choice was made, but also how difficult it was to make the choice (i.e. the emotional difficulty involved in decision-making).

The research does not suggest that designers should not include choices that give players emotional difficulty, but instead suggest that the emotional difficulty experienced when playing with with human team-mates compared to when playing with AI team-mates should be more similar *relative to each other*. This means that designers can use this information and present players with choices that are emotionally difficult, but are as emotionally difficult with an AI team-mate as they would be with a human team-mate.

8.2.2 Address the Trade-off situation

When designers want to ensure that player's have similar types of experiences with AI and human team-mates in terms of behavior and emotional difficulty involved in the choices, they must examine the attributes that contribute towards the decision-making process for these choices. In other words, they must address the trade-off situation.

As established by Ong & McGee [21], a general trade-off situation stems from players wanting the team to perform well vs wanting to be a "good team-mate". The manipulation in DTP (removing scoring information) only makes it less accurate for a player to determine performance, but does not affect the desire to be a "good team-mate" because choosing to sacrifice their team-mate has a negative impact on the team-mate's experience of the game (i.e. the team-mate will die).

The manipulation in RTB (increasing the interdependence between team members) changes the relationship between the player and the team-mate – both are now able to contribute towards team performance and neither will be left out as a result of the decision. This manipulation affects the desire to be a "good team-mate" because choosing to let the team-mate use the back paddle no longer has a negative impact on the team-mate's experience of the game (i.e. the team-mate will be able to contribute and not be side-lined).

The research does not suggest that games should not have trade-off situations at all, since giving players hard choices can help add to the richness of the experience. What it does suggest is that if the trade-off involves an attribute that is linked to the player's desire to be a "good team-mate", then the emotional difficulty involved in the decision-making is likely to be significantly higher with the human team-mate than the AI team-mate.

8.2.3 Be Cautious When Manipulating Interdependence

In this thesis, manipulating the interdependence between team members is shown to be a good way to help reduce the difference in behavior and emotional difficulty players have when making decisions. However, as highlighted by Wageman [30], interdependence is a complex concept with many different interpretations and definitions.

It has been suggested that the Index of Dependence (ID) can serve as an empirical means of calculating the dependence of one team member on another. Though it can serve as an indicator of whether team members are really more dependent on one another as a result of a particular manipulation, it only measures this in terms of the team's outcomes.

As seen from the study involving RTB, the Index of Dependence values for INTERDEPENDENT version of RTB do not seem significantly different from those of the INDEPENDENT version. This is a result of the unique nature of *Return the Ball*, where only the front paddle is able to return the ball.Despite this, participants still report that there is a difference in how much the team members depend on one another, based on the perception-related results of the study. On closer examination of the relationship between team members in the INTERDEPENDENT condition, team members depend on each other in a variety of ways, with examples of sports providing an analogy of the circumstances of each.

Both team members depend on one another because the performance of the team is dependent on the number of balls returned by the team member controlling the front paddle during each of the seven games, since only the front paddle can return the ball. However, the results show that on average, each team member does get to spend some time using the front paddle, and thus the participant needs to depend on the team-mate's performance to a certain extent. An example from sports that has a similar type of situation is team bowling, where each member of the bowling team adds to the combined score of each bowler, though only one member can bowl at any point of time.

Also, team members are dependent on each other not only do they need to pay attention to the ball, but also to the position of their team-mate in order to succeed. This is because the spawn position of the ball is determined by the positions of the front and back paddles relative to each other. This is similar to the situation in doubles tennis – both players on the court have to pay attention to each other's position (to make sure maximum coverage of the court), as well as where the ball is coming from. Though only one team member is able to hit the ball, both must adjust their positions in relation to each other to maximize the chance of successfully returning the ball.

Finally, team members depend on one another because of the nature of the roles and abilities of each paddle. The front paddle is the one that is able to return the ball, and helps ensure the success of the team in the process. The back paddle, though unable to return the ball, is able to make it easier for the front paddle to return the ball by influencing the spawn position of the ball such that it is closer to the front paddle. This is similar to the situation in volleyball, where one team member can help the other "set up" the ball for a return.

As can be seen from these three examples, the simple manipulation of interdependence between team members (by revealing the relationship between the front and back paddles) has resulted in quite a few changes in terms of how team members depend on one another. Also, while there has been no change in the fact that only the front paddle is able to return the ball, but it is clear that the team members are more interdependent.

8.3 Possible Objections

This section highlights and addresses possible objections to different aspects of the research described in this thesis. Possible objections raised about the experimental design employed include the use of these real-time cooperative games instead of typical prisoners' dilemma games, the duration of each game played within a round, the use of a between-subjects design and the slight deception of using an AI team-mate to play as the PH team-mate. Possible objections raised about the measurement and analysis of data include the measurement of the concept of Emotional Difficulty, and the choice of statistical analyses used to interpret the results.

8.3.1 Experimental Design

There may be some readers who feel that the nature of the research problem being examined does not warrant an experimental design that varies from conventional game theory experimental setups, since the researcher is investigating manipulations affecting choices made by participants.

While the use of typical Game Theory experimental setups can also measure the emotional difficulty and behavioral differences in the decision-making process with human and AI team-mates, their nature of just making a decision without the additional feature of playing the game with the consequences of the earlier decision can be considered to be somewhat simplistic. Through the inclusion of the gameplay phase where there is real-time coordination and interaction with the team-mate, this is more representative of the actual context of computer games.

8.3.2 Game Duration

Another possible objection is that the duration of the game sessions were relatively short and that the overall behavior or self-reported experience of participants may have been different if they had a longer period of exposure to their teammates. However, the amount of exposure participants had is similar to the duration in other games studies [11] and there was no indication that playing longer would have made a difference. In fact, prolonging the game sessions by adding more games to a round may result in the unintentionally influencing participants' decisions due fatigue or boredom. Participants in the DTP study played 14 games (four tutorial, four AI team-mate, four PH team-mate) while those in the RTB study played 18 games (four tutorial, seven AI team-mate, seven PH team-mate).

8.3.3 Between-Subjects Design

Counterbalancing was done in both experiments to ensure that order effects were ruled out, but there might still be objections about the use of a between-subjects design in the studies. The use of the between-subjects design in this 2x2 experiment meant that participants in the respective experiments were exposed to only one version of the cooperative game, rather than being shown both versions. The objection that could be raised is that there might be significant variability between the two groups – or significant observer (researcher) expectancy bias.

As noted earlier regarding both studies, the two groups were equivalent in terms of age and gender distribution. Statistical analyses found that across the other demographic data categories collected such as experience playing games, cooperative games, or cooperative games with AI team-mates etc, any differences were not statistically significant.

In terms of expectancy bias, the study was not designed or conducted with the goal of confirming a particular outcome. Rather, it was designed to see whether there was a particular correlation. Beyond that, although the same researcher conducted the sessions and ran the quantitative analysis on the results, the substantive results were quantitative (game-play logs and Likert-like responses) and thus did not involve coding or interpretation. It is therefore unlikely that the substantive results of the paper were the result of group variance or research expectancy bias.

8.3.4 Human confederate as team-mate

Another potential objection to be raised is that having an actual human team-mate playing with the participants (rather than an AI team-mate) may have led to different results from those obtained. The studies were structured such that when playing with the PH team-mate, participants were told they were playing with a remotely-located team-mate, who joined the game via the network. Thus, even if the team-mate was an actual human confederate instead of the PH team-mate, the structure of the sessions would remain as originally designed – participants would not see or interact directly with their team-mate outside of the game.

In addition, both games are simple enough to leave little or no room in terms of variation in playing style (as evidenced by the number of participants who used similar strategies in the game). It is therefore unlikely that the use of an actual human team-mate rather than an AI would have made a noticeable difference to the participants in terms of gameplay.

8.3.5 Measurement of Emotional Difficulty

One issue that may be considered problematic when examining the results of this thesis is the measurement of emotional difficulty involved in the decision-making process in each study. As seen from the questionnaire, participants are asked about how difficult it is to *choose their position* when playing with the different team-mates. It could be argued that this question does not directly address the conflict between the two attributes they use in the decision-making process (e.g. protecting the team-mate vs team performance in DTP).

Though the questionnaire does not directly address the conflict during decision-making in the two studies, it still requires participants to gauge how difficult it was to make the choice of position. Since the results show there is higher emotional difficulty in the decision-making with the PH team-mate than the AI team-mate in some versions of the respective games, it indicates that participants do feel that some degree of conflict in the decision-making process exists for them.

Another objection that could follow-on from this is the fact that the types of emotional difficulty experienced by participants in each of the studies may not necessarily be the same, which would make it difficult to generalize the results. In other words, the emotional difficulty participants experienced in DTP could be very different from that in RTB, due to the nature of the respective games and the trade-offs involved in each. While this concern is valid, it should be noted that what this thesis aimed to do was examine how the *difference* in emotional difficulty in choices made with human and AI team-mates could be moderated, and does not include the types or categories of emotional difficulty as part of its scope.

8.3.6 Choice of Statistical Analyses

The 2x2 nature of the experimental design used in both studies would lead some to suggest that it would be more appropriate to use an ANOVA rather than independent-samples and paired-samples t-tests when analyzing the results.

The aim of these studies are not to determine if there is any interaction that occurs involving the between-subjects and within-subjects factors. Instead the studies focus on how the respective manipulations moderate the comparative difference in how participants respond to PH and AI team-mates. An ANOVA would be more appropriate in an experimental set up where there is a comparison being made between multiple pairs. This would help reduce type-I bias that can occur when there are three or more within-subjects conditions being compared. Since the between-subjects conditions have only one pair each for this study, the use of t-tests would be sufficient.

8.4 Summary

This chapter did the following:

- Compared the behavior and emotional difficulty results of the studies using *Defend the Pass* (DTP) and *Return the Ball* (RTB)
- Explained the results for the study using DTP: The manipulation of visibility of scoring information does not reduce emotional difficulty differences because participants still have a clear reminder about the status of the team-mate as a consequence of their choice.

- Explained the results of the study using RTB: The manipulation of interdependence between team members reduces the emotional difficulty differences because by making team members interdependent, both paddles are now able to contribute towards overall collective performance, without either of the team members getting sidelined.
- Provided three insights for designers who want to ensure players respond similarly towards human and AI team-mates. Designers should (1) Address both behavioral and emotional difficulty differences; (2) Address the trade-off situation and (3) Exercise caution when manipulating interdependence.
- Addressed potential objections towards the setup up of the experiment, the methodology of data collection, and the statistical analyses employed to interpret the data.

Chapter 9

Conclusion

> This chapter provides a summary and recap of this thesis, giving an overview of the results and findings of the two studies conducted. It also lists the contributions of this thesis to existing body of knowledge, and sets the direction for future work that can be done to extend this research.

This thesis examined the results of two studies that found that in order to moderate the difference in responses towards human and AI team-mates, the manipulations must address and eliminate the trade-off that exists between the attributes involved in the decision-making process. This contributes to existing knowledge by showing that the difference in responses towards human and AI team-mates can be moderated such they are treated more similarly. However, both behavioral and emotional difficulty differences must be addressed. The thesis also suggests that while interdependence between team members is a good means to use to moderate this difference, interdependence itself is a complex concept, and even small changes can affect the relationship between team members in various ways. This chapter also suggests future work that can be done to follow up on this thesis, such as examining the effect of different types of interdependence on moderating the difference towards human and AI team-mates, or to examine the different types of emotional difficulty players may have when making decisions.

9.1 Contributions

Two studies were carried out to determine what manipulations could moderate the difference in responses towards human and AI team-mates in real-time cooperative games. The first involved the game Defend the Pass

(DTP), where participants chose between sacrificing and protecting their team-mate. In this study, the visibility of scoring information was manipulated, and results showed that while differences in behavior were moderated, differences in emotional difficulty were not. The second involved a game Return the Ball (RTB), where participants chose between controlling the front paddle or back paddle. In that study, the interdependence between team members was manipulated, and results showed that differences in behavior and emotional difficulty were moderated.

The results also demonstrated that manipulating the interdependence between team members was able to eliminate the trade-off between attributes used by participants in the decision-making process. However, manipulation of the visibility of scoring information did not eliminate the trade-off between attributes used by participants in the decision-making process.

These results suggest two things about the approach that game designers of cooperative games should bear in mind when wanting to ensure that players' experiences with a computer team-mate are similar to those with a human team-mate. Firstly, it is necessary to address both the behavior and emotional difficulty involved in the decision-making process in order to truly moderate the difference in responses. This can be achieved by eliminating the existing trade-off between the attributes that exist in the decision-making process.

Secondly, while interdependence is a good means of moderating these behavioral and emotional difficulty differences, even a simple manipulation of the relationship between team members (by making them more interdependent) can result in a change in the ways that they depend on each other.

9.2 Future Work

The work of this thesis opens up directions for future work in this area. One such direction would be to examine how different types of interdependence (rather than just level of interdependence) is able to moderate the difference in how people respond towards human and AI team-mates. Findings relating to this would be useful because they could provide game designers more specific and helpful information about how to incorporate different types of interdependent relationships between team members into the design of games that require cooperation.

Another possible direction for research would be to examine the types of emotional difficulty experienced by players when faced with trade-offs, and what kind of design decisions could be taken to target these specific types of emotional difficulty if the designers had an intention of ensuring that responses towards human and AI team-mates were more similar.

Bibliography

[1] Adam Biessener. Valve's New Game Announced, Detailed: Dota 2, 2012.

[2] Jim Blascovich. A theoretical model of social influence for increasing the utility of collaborative virtual environments. In *Proceedings of the 4th international conference on Collaborative virtual environments - CVE '02*, pages 25–30, New York, New York, USA, 2002. ACM Press.

[3] Subimal Chatterjee and TB Heath. Conflict and loss aversion in multiattribute choice: The effects of trade-off size and reference dependence on decision difficulty. *Organizational Behavior and Human Decision Processes*, 67(2):144–155, 1996.

[4] L Gong. How social is social responses to computers? the function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, 24(4):1494–1509, 2008.

[5] Kevin J. Haley and Daniel M. T. Fessler. Nobody's watching? *Evolution and Human Behavior*, 26(3):245–256, May 2005.

[6] Daniel Johnson and John Gardner. The media equation and team formation: Further evidence for experience as a moderator. *International Journal of Human-Computer Studies*, 65(2):111–124, February 2007.

[7] Ralph L Keeney and Howard Raiffa. *Decisions with multiple objectives: preferences and value trade-offs.* Cambridge university press, 1993.

[8] Harold H. Kelley and John W. Thibaut. *Interpersonal relations: A theory of interdependence.* Wiley, New York, New York, USA, 1978.

[9] S. Kiesler, L. Sproull, and K. Waters. A prisoner's dilemma experiment on cooperation with people and human-like computers. *Journal of personality and social psychology*, 70(1):47–65, January 1996.

[10] Eun-Ju Lee and Clifford Nass. Experimental tests of normative group influence and representation effects in computer-mediated communication. *Human Communication Research*, 28(3):349–381, 2002.

[11] Sohye Lim and Byron Reeves. Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player. *International Journal of Human-Computer Studies*, 68(1-2):57–68, January 2010.

[12] Mary Frances Luce, John W Payne, and James R Bettman. Emotional trade-off difficulty and choice. *Journal of Marketing Research (JMR)*, 36(2), 1999.

[13] Kevin McGee and Aswin Thomas Abraham. Team-mate AI in games: A definition, survey & critique. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games (FDG '10)*, pages 124–131, New York, NY, USA, 2010. ACM.

[14] Tim Merritt, Kevin McGee, Teong Leong Chuah, and Christopher Ong. Choosing human team-mates: perceived identity as a moderator of player preference and enjoyment. In *Proceedings of the 6th International Conference on Foundations of Digital Games (FDG '11)*, pages 196–203, New York, NY, USA, 2011. ACM.

[15] Tim Merritt, Christopher Ong, Teong L. Chuah, and Kevin McGee. Did you notice? Artificial team-mates take risks for players. In *11th International Conference on Intelligent Virtual Agents*. Springer, September 2011.

[16] Tim R. Merritt, Kian B. Tan, Christopher Ong, Aswin Thomas, Teong L. Chuah, and Kevin McGee. Are artificial team-mates scapegoats in computer games. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, CSCW '11, pages 685–688, New York, NY, USA, 2011. ACM.

[17] Timothy Robert Merritt. *A Failure of Imagination: How and Why People Respond Differently to Human and Computer Team-Mates.* Phd, National University of Singapore, 2012.

[18] John Morkes, K Hadyn, and Clifford Nass. Effects of Humor in Task-Oriented Human-Computer Interaction and Computer-Mediated Communication: A Direct Test of SRCT Theory. *Human-Computer Interaction*, 14(4):395–435, 1999.

[19] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. Computers are social actors. *Conference companion on Human factors in computing systems - CHI '94*, page 204, 1994.

[20] Kristine L Nowak and Frank Biocca. The Effect of the Agency and Anthropomorphism on Users Sense of Telepresence , Copresence , and Social Presence. *Presence*, 12(5):481–494, 2002.

[21] Christopher Ong and Kevin McGee. Good team-mates do more than help the team win: design factors that impact player concern about team-mate experience. In *24th Australian Computer-Human Interaction Conference*, 2012.

[22] Christopher Ong, Kevin McGee, and Teong Leong Chuah. Closing the human-ai team-mate gap: how changes to displayed information impact player behavior towards computer teammates. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, OzCHI '12, pages 433–439, New York, NY, USA, 2012. ACM.

[23] Jason E. Plaks and E. Tory Higgins. Pragmatic use of stereotyping in teamwork: Social loafing and compensation as a function of inferred partner-situation fit. *Journal of Personality and Social Psychology*, 79(6):962–974, 2000.

[24] Niklas Ravaja, Timo Saari, Marko Turpeinen, Jari Laarni, Mikko Salminen, and Matias Kivikangas. Spatial Presence and Emotions during Video Game Playing: Does It Matter with Whom You Play? *Presence: Teleoperators and Virtual Environments*, 15(4):381–392, August 2006.

[25] Bryon Reeves and Clifford Nass. *The media equation: how people treat computers, television, and new media like real people and places.* Cambridge University Press, New York, NY, USA, 1998.

[26] J.D. Thompson. Technology and Structure. In *Organizations in Action*, chapter 5, pages 66–82. McGraw-Hill, New York, NY, USA, 1967.

[27] Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, October 1992.

[28] Bart Victor and RS Blackburn. Interdependence: An alternative conceptualization. *Academy of Management Review*, 12(3):486–498, 1987.

[29] Astrid M. von der Pütten, Nicole C. Krämer, Jonathan Gratch, and Sin-Hwa Kang. It doesnt matter what you are! Explaining social effects of agents and avatars. *Computers in Human Behavior*, 26(6):1641–1650, November 2010.

[30] Ruth Wageman. The Meaning of Interdependence. In Marlene Turner, editor, *Groups at work: Theory and research*, chapter 7, pages 197–217. Lawrence Erlbaum Associates, Mahwah, New Jersey, USA, 2001.

[31] D. Weibel, B. Wissmath, S. Habegger, Y. Steiner, and R. Groner. Playing online games against computer- vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in Human Behavior*, 24(5):2274–2291, September 2008.

Appendices

Appendix A

Interdependence Theory Matrices for Defend the Pass

Protected AI Team-mate (NO-SCORE version)		
	Team-mate successful in killing Monsters	Team-mate unsuccessful in killing Monsters
Participant successful in killing Monsters	30	20.4
Participan unsuccessful in killing Monsters	20.4	0

Table A.1: Representation of relationship when participant protects AI team-mate in NO-SCORE version

Protected AI Team-mate (NO-SCORE version)		
	Team-mate successful in killing Monsters	Team-mate unsuccessful in killing Monsters
Participant successful in killing Monsters	30	20.4
Participant unsuccessful in killing Monsters	20.4	0

Table A.2: Representation of relationship when participant sacrifices AI team-mate in NO-SCORE version

Protected PH Team-mate (NO-SCORE version)		
	Team-mate successful in killing Monsters	Team-mate unsuccessful in killing Monsters
Participant successful in killing Monsters	30	20.18
Participant unsuccessful in killing Monsters	20.18	0

Table A.3: Representation of relationship when participant protects PH team-mate in NO-SCORE version

Sacrificed PH Team-mate (NO-SCORE version)		
	Team-mate successful in killing Monsters	Team-mate unsuccessful in killing Monsters
Participant successful in killing Monsters	30	26.7
Participant unsuccessful in killing Monsters	26.7	0

Table A.4: Representation of relationship when participant sacrifices PH team-mate in NO-SCORE version

Protected AI Team-mate (SCORE version)		
	Team-mate successful in killing Monsters	Team-mate unsuccessful in killing Monsters
Participant successful in killing Monsters	30	20.46
Participant unsuccessful in killing Monsters	20.46	0

Table A.5: Representation of relationship when participant protects AI team-mate in SCORE version

Sacrificed AI Team-mate (SCORE version)		
	Team-mate successful in killing Monsters	Team-mate unsuccessful in killing Monsters
Participant successful in killing Monsters	30	26.29
Participant unsuccessful in killing Monsters	26.29	0

Table A.6: Representation of relationship when participant sacrifices AI team-mate in SCORE version

Protected PH Team-mate (SCORE version)		
	Team-mate successful in killing Monsters	Team-mate unsuccessful in killing Monsters
Participant successful in killing Monsters	30	19.63
Participant unsuccessful in killing Monsters	19.63	0

Table A.7: Representation of relationship when participant protects PH team-mate in SCORE version

Sacrificed PH Team-mate (SCORE version)		
	Team-mate successful in killing Monsters	Team-mate unsuccessful in killing Monsters
Participant successful in killing Monsters	30	27.14
Participant unsuccessful in killing Monsters	27.14	0

Table A.8: Representation of relationship when participant sacrifices PH team-mate in SCORE version

Appendix B

Interdependence Theory Matrices for Return the Ball

Participant uses Front paddle with AI team-mate (INDEPENDENT version)		
	Participant successful in returning ball	Participant unsuccessful in returning ball
Team-mate successful in coordinating with participant	3.5	0
Team-mate unsuccessful in coordinating with participant	3.5	0

Table B.1: Representation of relationship when participant uses Front paddle with AI team-mate in INDEPENDENT version)

Participant uses Back paddle with AI team-mate (INDEPENDENT version)		
	Participant successful in coordinating with team-mate	Participant unsuccessful in coordinating with team-mate
Team-mate successful in returning ball	4.5	4.5
Team-mate unsuccessful in returning ball	0	0

Table B.2: Representation of relationship when participant uses Back paddle with AI team-mate in INDEPENDENT version)

Participant uses Front paddle with PH team-mate (INDEPENDENT version)		
	Participant successful in returning ball	Participant unsuccessful in returning ball
Team-mate successful in coordinating with participant	3.6	0
Team-mate unsuccessful in coordinating with participant	3.6	0

Table B.3: Representation of relationship when participant uses Front paddle with PH team-mate in INDEPENDENT version)

Participant uses Back paddle with PH team-mate (INDEPENDENT version)		
	Participant successful in coordinating with team-mate	Participant unsuccessful in coordinating with team-mate
Team-mate successful in returning ball	4	4
Team-mate unsuccessful in returning ball	0	0

Table B.4: Representation of relationship when participant uses Back paddle with AI team-mate in INDEPENDENT version)

Participant uses Front paddle with AI team-mate (INTERDEPENDENT version)		
	Participant successful in returning ball	Participant unsuccessful in returning ball
Team-mate successful in coordinating with participant	3.9	0
Team-mate unsuccessful in coordinating with participant	3.5	0

Table B.5: Representation of relationship when participant uses Front paddle with AI team-mate in INTERDEPENDENT version)

Participant uses Back paddle with AI team-mate (INTERDEPENDENT version)		
	Participant successful in coordinating with team-mate	Participant unsuccessful in coordinating with team-mate
Team-mate successful in returning ball	4.8	4.5
Team-mate unsuccessful in returning ball	0	0

Table B.6: Representation of relationship when participant uses Back paddle with AI team-mate in INTERDEPENDENT version)

Participant uses Front paddle with PH team-mate (INTERDEPENDENT version)		
	Participant successful in returning ball	Participant unsuccessful in returning ball
Team-mate successful in coordinating with participant	3.9	0
Team-mate unsuccessful in coordinating with participant	3.6	0

Table B.7: Representation of relationship when participant uses Front paddle with PH team-mate in INTERDEPENDENT version)

Participant uses Back paddle with PH team-mate (INTERDEPENDENT version)		
	Participant successful in coordinating with team-mate	Participant unsuccessful in coordinating with team-mate
Team-mate successful in returning ball	4.9	4
Team-mate unsuccessful in returning ball	0	0

Table B.8: Representation of relationship when participant uses Back paddle with AI team-mate in INTERDEPENDENT version)