FORMAL MODEL OF TRUST DYNAMICS FOR SHORT-TERM HUMAN-ROBOT INTERACTION

WADHAH A. ABDULHUSSAIN

SCHOOL OF COMPUTING
UUM COLLEGE OF ARTS AND SCIENCES
UNIVERSITI UTARA MALAYSIA
2015

Permission to Use

In presenting this dissertation in partial fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this dissertation in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my dissertation.

Requests for permission to copy or to make other use of materials in this project dissertation, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

.

Perkembangan dalam bidang teknologi robotik pada tahun sebelum ini telah membuka pasaran serta cabaran yang luas dalam bidang tersebut. Interaksi dengan teknologi canggih telah membawa kepada perbincangan yang penting tentang bagaimana interaksi tersebut dimulakan itu mampu mewujudkan interaksi yang lancar antara manusia dengan robot (interaksi manusia-robot)dan mampu bertahan untuk masa yang lama. Salah satu faktor penting yang mempengaruhi tahap interaksi antara manusia dengan robot adalah tahap kepercayaan. Kepercayaan adalah keyakinan bahawa pergantungan kepada individu itu tidak akan membawa kesan negatif atau mendatangkan bahaya. Dari segi psikologi pengkomputan, model formal (model pengkomputan) telah digunakan untuk mendapat gambaran tentang fungsi kognitif serta corak tingkah laku manusia. Oleh itu, kajian mengenai pemodelan formal kepercayaan dalam interaksi manusia-robot dibina untuk menjawab persoalan bagaimana kepercayaan terbentuk semasa manusia berinterkasi dengan robot. Menerusi sorotan berkaitan, terdapat 18 faktor asas yang berkaitan dan ini meliputi; personaliti, penampilan fizikal, kepercayaan tingkah laku, isyarat tingkah laku, tahap automasi, pengalaman positif, ketelusan, persepsi risiko jangka pendek, persepsi risiko jangka panjang, kepercayaan kepada pergantungan jangka pendek, persepsi kecekapan, penyamaran positif, pengalaman positif jangka panjang, kepercayaan jangka pendek, kesangsian jangka pendek, kepercayaan jangka panjang dan ketidakpercayaan jangka panjang. Faktor ini telah digunakan sebagai pengetahuan asas untuk membangunkan kepercayaan manusia terhadap robot. Satu model formal telah dibangunkan menggunakan set persamaan pembeza. Seterusnya, terdapat lima kes yang berbeza telah digunakan bagi membentuk simulasi untuk menerangkan proses kepercayaan dalam interaksi manusia dan robot; iaitu 1) tahap kepercayaan yang tinggi, 2) tahap kepercayaan sederhana tinggi, 3) tahap kepercayaan sederhana, 4) tahap kepercayaan sederhana rendah dan 5) tahap kepercayaan yang rendah. Model yang dibangunkan telah diuji dengan menggunakan analisis matematik (analisis kestabilan) dan pengesahan automatik (bahasa surih masa).

Abstract

Rapid advance of robotic technologies in the last years have opened numerous venues and great challenges in the field of robotics technology. Interacting with those advanced technologies has carried huge debates on how such interactions can be instigated to create fluent interaction between humans and robots that can last for long time (human-robot interaction). One of the crucial factors that majorly influence the level of interaction between human and robot is the level of trust. Trust is the feeling of confidence that the reliance on other partner will not yield negative or dangerous consequences. In computational psychology domains, formal models (computational models) were used to acquire deep insights of human cognitive functions and behavior patterns. Therefore, this study implements formal model of trust in human robot interaction to answer how trust can be a reason to initiate interaction between human and robot. From related literature, eighteen basic factors have been established that include; personality, physical appearances, believable behavior, behavior cues, level of automation, positive experiences, transparency, perception, long term perceive risk, short term perceive risk, reliable behavior, perceive competency, positive deception, long term positive experiences, short term trust, short term distrust, long term trust, long term distrust. Those factors provide the fundamental knowledge of developing trust in robot. A formal model was developed based on a set of differential equations. Next, Five different cases were implemented to simulate various scenarios that explain the development of trust in HRI; namely, 1) high level of trust, 2) moderate high level of trust, 3) moderate level of trust, 4) moderate low level of trust, and 5) low level of trust. The developed model was verified by using mathematical analysis (stability analysis) and automated verification (temporal trace language).

Acknowledgement

The most praise, gratitude, and thankfulness goes to the Almighty Lord, the most merciful and compassionate, for reconciling, and providing me with patience and strength to accomplish my MSc dissertation.

My sincerest gratitude and appreciations go to my supervisor, Dr. Azizi Ab Aziz, for his guidance, encouragements, and the unique hospitality since he treated our group as his family. My gratefulness for the valuable knowledge we gained from him to complete this valuable thesis. Definitely, I will miss your words "Sleep enough otherwise u will be in trouble", "I was student I knew all these tricks", "We must push the boundaries", and "I don't care". Lastly, I would like to thank him for keeping my academic plans on schedule, and encouraging me to keep up ready for the next challenge in my life which is chasing PhD dream.

To my extended family in Iraq, a special heartfelt thankfulness to the first inspirer, my father Eng. Dr. Abdulhamza Abdulhussain, for his financial, moral, efforts, motivation, sacrifice, and unlimited support since I started to pursue my master program in Malaysia at august 2013. I wish to make you feel proud of me. I also would love to express my sincerest gratitude to the soft-hearted and genuine love that I consider myself lucky to have such a great mother, Khairiya Taj, who has believed in me and never stopped praying for me. Your prayers always enlightened my way and blessed my life. I believe that my parents' satisfaction of me was the key word to my successful. So I dedicate and acknowledge this dissertation to whom I'm beholden for the rest of my life, my wonderful parents.

Moreover, I would love to express my thankfulness to my brilliant siblings, the most precious persons that I have in this world, (Zena, Hussain, Alaa, Hiba, Firas, Noora) for their continuous encouragements, supports and prayers.

I also would like to extend my thanks and appreciations to all my friends who have contributed to complete this thesis successfully especially my close friend Hayder M. Ali Phd Candidate for his help, and my best friends Zaid Kubba for his inspiration, Wael Hasan, Mustafa, Maytham, Abbas Saad, Joke, Faisal Saido, Harith Azam, and Yassir. We have spent such awesome time together; and by knowing you, I never felt alone or in alienation especially when I am 4234 Miles (the distance between Baghdad and Alor setar) away from our country.

I would love to extend my thanks to my lovely and closest friends back at home Ahmed Hashim, Nawar Kahtan, and Nawar Zuhair, for their continuous asking, calling, and Duaa for the whole 2 years. Many thanks go to everyone stood by my side and gave me advices, who ask me to look at the glass half full and to live with no regrets.

Lastly, I would love to say “Terima kasih” to Malaysia and I strongly love you. May Allah protect this land.

Wadhah A. Abdulhussain

UUM, Sintok, April, 2015.

Table of Contents

List of Tables

List of Figures

List of Appendices

CHAPTER ONE
INTRODUCTION

1.1 Background

In all facets of life, people normally communicate frequently and collaborate with each other in several ways and manners. A sample form of this communication and collaboration exists between human and artefacts such as robot, avatar, animation and software agent. The most essential key for interacting with other people or peers most times is based on their interpersonal trust.

In light of this, trust is a pivotal key for cooperative efforts in all aspects of our everyday life (Sanders, Wixon, Schafer, Chen, & Hancock, 2014). Human-interpersonal trust can be well-defined as the enthusiasm or willingness of a party to be exposed to the outcomes of another party. This is based on the expectation that the other party will execute a particular action significant to the trustor, without the capability to control or monitor that other party (Mayer, Davis, & Schoorman, 1995).

Based on previous clarification, specific expectation to the actions of parties, trustee and trustor have been addressed, where trust is typically the characteristic of the trustee (e.g., goodwill, honesty, morality, expertness and care). Trust is a much disputed and common topic in numerous research studies. Hence in addition, the term trust has been extensively examined in different fields such as economics, business, marketing, politics, e-commerce, psychology, sociology, medicine, nursing, and computing science (Masthoff, 2007).

A few numbers of researchers have mentioned in their studies that trust could be represented as different types such as trusting intention, trusting behavior, trusting beliefs, disposition to trust and finally, institution trust (McKnight & Chervany, 1996; Trang, Zander, & Kolbe, 2014). Trusting beliefs which helps to make the perspective of trust clear can be defined as the extent to which one believes (and feels confident in believing) that the other person is trustworthy in the situation, while, trustworthiness in this case can be described as the capability and willingness to act in the other person's top interests (McKnight & Chervany, 1996).

Secondly, the definition of trusting intention can be likened to when one peer eagerly intends to depend on the other peer with a sense of relative security, in spite of the absence of control over the party, and also despite negative consequences being obvious. Thirdly, trusting behavior is defined as when a person gladly depends on another person with a sense of relative confidence, in spite of negative outcomes being possible. Dependence in this context is classified as a behavioral term, which distinguishes trusting intentions which is willingness to depend from trust related–behavior. Fourthly, institution based trust means an aspect of one's life or one's beliefs with emotions of relative security, where helpful conditions are in places which encourage situational success in dangerous or risky endeavors. Finally, disposition to trust is defined as the extent or degree to which there is a consistent tendency to be willing to depend on the partner in certain circumstances. (McKnight & Chervany, 2001).

Based on the above mentioned definitions, the construct of these different types of trust emanates from sociology tradition theories which state that people can depend on others because of the situation, structure, or role that guarantee or assure that things will go on well or would be fine. However, generally speaking, trust in artificial artefacts have been examined in recent studies and the issue of trusting in robots have significantly become a crucial case due to the noticeable presence of robots in our environment of our day-to-day life (Atkinson et al., 2012; Grodzinsky, Miller, & Wolf, 2012; Taddeo, 2010).

This issue of trust plays a very important role and one of the greatest challenges of effective collaboration between robots and human is the improvement of appropriate levels of joint trust in robots (Desai, Stubbs, Steinfeld, & Yanco, 2009). Irrespective of the domain of the application, the environment, or the given task, a human's trust in their non-human collaborator is an essential element required to ensure that any functional relationship will ultimately be effective.

The primary factors that impacts initial robot trustworthiness are reputation, physical form (e.g., anthropomorphism), robot type, robot personality, intelligence, as well as, level of automation and perceived function (e.g., behavior) (Daellenbach & Davenport, 2004; Dautenhahn, 2007; Lee & See, 2004). Since initial trustworthiness is assessed prior to interaction, the perceived functionality of the robot can be drawn from its reputation or viewing the robot's physical form. This is also obvious as seen in literature on interpersonal trust, which suggested that trustworthiness is established through assessing the partner's ability, benevolence, and integrity (Colquitt, Scott, & LePine, 2007). However, automation literature suggests that trustworthiness is a key element when researching on trust and it is founded on the characteristics of the automation (Daellenbach & Davenport, 2004; Lee & See, 2004).

Specifically, the robot's physical form may be the primary influence of initial trustworthiness due to the importance of features on perception of trustworthiness. For example, the e-commerce domain suggests that attributes such as color, brand, and the layout of computer interface can increase or reduce the perception of trustworthiness (Kim & Moon, 1998; Nickerson & Reilly, 2004). In addition, the human cognitive aspect of trust arises from our ability, based on various dimensions of commonality, to make reasonable inferences about the internal state of other agents such as beliefs, dispositions and intentions in order to predict future behavior and judge the risks versus the benefits.

Thus, based on the above, trust between human and robot is a crucial issue and studying that phenomenon is highly important to provide a clearer and richer understanding and also to answer the question why human trust often occurs in machines such as robot, mobile, watches and so on. In order to achieve this, modeling the dynamic mechanism of trust between human and robot is needed. Modeling dynamic is a representation of a phenomenon over a period of time which can be achieved by implementing computational model analysis.

Computational models are models that can simulate a set of dynamic processes perceived in a natural world in order to obtain knowledge of these processes, and then predict the outcome of the natural processes by giving specific set of input parameters. Hence, this study has developed a computational model to understand the condition of trust in human-robot interaction (HRI) based on identified important attributes and relationships related to human-robot interaction (Shiflet & Shiflet, 2014).

1.2 Problem Statement

No doubt, robots or any intelligent artefacts are becoming more important in various environments and applications with the purpose of providing aid and making huge development in human daily activities. For instance, lately, robots and virtual intelligent agents are being widely used in healthcare, education, edutainment, and even for military purposes. Therefore, exploring issues that are related to maintaining the interaction between these new technologies and human beings is extensively researchable to get a deeper insight to the phenomenon interaction taking place in human robot interaction (Fasola & Mataric, 2013; Hancock, Billings, Schaefer, et al., 2011; Sidner, Kidd, Lee, & Lesh, 2004).

However, trust is one of the important issues in maintaining the interaction between human and robot and it is has been extensively studied in recent literature (Robinette, Wagner, & Howard, 2014; Sanders et al., 2014; Taddeo, 2010; Trang *et al*., 2014). Also, trust is stated as an important factor in human-robot interaction because of the effects that the existence or lack of trust may have on interactions (Billings, Schaefer, Chen, Kocsis, *et al*., 2012). Regarding to the trust, recent studies have shown that the performance of robot is a very critical issue in building human trust (van den Brule, Dotsch, Bijlstra, Wigboldus, & Haselager, 2014). And therefore, trust in robot is a well investigated area of human-robot interaction. Hancock *et al.* (2011) explicated that antecedents of trust can be separated into robot capabilities or performance-based characteristics (e.g., behavior, reliability), and robot features or attribute-based characteristics (e.g., anthropomorphism).

Thus, building trust for human while interacting with robots is dynamically developing and maybe changing based on the duration of interaction between human and robot. Therefore, modelling the dynamics of trust in human robot interaction is strongly needed to get a deep understanding on how people develop their trust while they interact with robot. This will provide guidance to designers of robots and other intelligent artefacts in designing a more trustworthy artefact that can increase the level of human trust on their designs. However, this trustworthy interaction can be influenced by many factors such as perceived risk, stress, unstable workload and undefined task (Hancock, Billings, Oleson, *et al*., 2011).

Wells and Deguire, (2005) presented an innovative robotic application that can detect and assist in rescue missions; the special weapons observation reconnaissance detection system (SWORD), which was developed and deployed in 2007 to assist and act as a supportive agent (autonomous agent) to the American army in operations of the American troop in Iraq. Nevertheless, these assistive robots were never used because the soldiers did not trust them to work correctly in dangerous situations. The disuse of Talon robot in such military purposes was as a result of the existence of distrust from the soldiers towards a robotic teammate in critical situations and when in need to make a decision. This eventually led to a failure in the desired interaction between the human and robot (Ogreten, Lackey, & Nicholson, 2010).

Furthermore, Jianxin and Bond (1992) pointed out that there is relationship between trust, cognitive, situation awareness and perceived risk. This implies that trust has more to do with cognitive and mental abilities. The concept of cognitive models have been widely used to formulate dynamic phenomenon in mental state systems (Freeman, Exell, Meadmore, Hallewell, & Hughes, 2014; Khan, Bashir, Nasir, Shah, & Mehmood, 2014; Marsella, Gratch, & Petta, 2010; Masthoff, 2007).

However, less attention has been given to the study of dynamics of trust in human-robot interaction in literature. Also, none of these studies explicitly interpret how the mechanism behind trust is obtained. Thus, there is further need to clearly understand the phenomenon and factors influencing the interaction that can enhance or produce trustworthy robots. Hence, this study developed a formal model (computational model) of dynamics trust in human and robot interaction.

1.3 Research Questions

Based on above explanation, this study will answer the following research questions:

i. How to develop a formal model to represent relationship of temporal dynamics in trust for human- robot interaction?

ii. How to evaluate the proposed model?

1.4 Research Objectives

Generally, this study aims at developing a cognitive model of trust, related to the dynamics in human robot in short term interaction. In order to answer the two research questions, the following objectives are designed:

i. To identify the important factors of the dynamics trust in human-robot in short term interaction.

ii. To develop a formal model of dynamics factors obtained from previous objective.

iii. To evaluate the proposed model by using selected cases via verification (mathematical and automated temporal verification).

1.5 Scope of the Study

The scope of this study will be conducted based on the important factors for trust that is related with short term interaction between humans and robot. And the relationships between those identified factors will be selected based on the related literature review, empirical evidence and expert view. Then, the important identified factors will be formalized by using a set of differential equations and will be simulated in numerical programming language tool (Matlab) where five cases of trust dynamics simulation will be selected to generate the simulated traces. Finally, the developed model will be evaluated using two techniques: mathematical analysis and automated verifications.

1.6 Significance of the study

This study aims at identifying important factors of interaction based on the trustworthiness of the robot. These factors have significant impacts on dealing with sociable robot. The selected factors will be modeled for two purposes: First, theoretical significance in computational model of trust (in HRI) especially to understand important behavioral perspectives that can enhance human robot interaction trust. Second, practical significance: to help robotic system designers in identifying and developing factors that can increase human trust when interacting with a sociable robot as a basis to develop trust evaluation within digital artifacts (agent/robot).

1.7 Summary

This chapter explains the general background of trust and the impact of trust in human-machine interaction. The problem statement has been formulated, as well as research questions, objectives, significance, and scope of the study.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

In light of trust definitions stated in the previous chapter, the development of an appropriate level of mutual trust between human and robotic systems is considered as the most significant challenge for performing successful collaboration (Desai, Stubbs, Steinfeld, & Yanco, 2009; Groom & Nass, 2007). Regardless of the area of application, be it in a structured, controlled or unstructured environment, or in consideration of the type of task, whether it attributes by simplicity or complexity, a human's trust in their non-human assistive teammate is needed to make sure that all functional tasks will be working effectively.

Many researchers have claimed that robots belong to the automation field; therefore the definition of the term automation is needed. Automation can be defined as "the execution by a machine agent (usually a computer) of a function or task that was previously carried out by a human" as coined by Parasuraman & Riley (1997). Usually these kinds of tasks are characterized as dangerous and stressful for humans to cope with, such as explosive detection in military fields, or search and rescue in the battlefield, meanwhile the humans will only need to play a supervisory role only.

However, researchers in social robotics domain have described a robot as an autonomous system which is able to perform certain commands automatically by a human operator (Feil-Seifer & Mataric, 2005). Also, Scholtz (2003) suggested that the type of interaction, the number of systems that might be interacting with the humans, the nature of environment, and the physical nature of the robot (e.g., complex, dynamic control) help to make a difference between HRI (Human Robot Interaction) and human-computer or human-machine interaction.

In addition, Yagoda (2011) defined a robot as the interaction between intelligence and autonomy of technology. The author argued that based on this interaction, robots are distinctly different from other machines, such as a refrigerator (low intelligence, high autonomy) or an expert system (high intelligence, low autonomy). Also, Fong, Thorpe, and Baur (2003) defined that a robot is a complex, dynamic system that shows a degree of applying automation and cognition as it operates in a real-world environment. From this perspective, a robot adapts and deals with the environment based on a set of sensors, and these sensors have the ability to move in order to manipulate an uncertain physical world (Groom & Nass, 2007; Lin, Bekey, & Abney, 2008). This allows robots to be deployed in unstructured environments with task-oriented goals.

Obviously, the robot is a reprogrammable and flexible agent and in addition it has mobility capacity (Lin et al., 2008); (Holm, Lukander, Korpela, & Sallinen, 2009). Schaefer (2013) mentioned that there are up to 300 related citations which indicate the concept of trust among different research areas with 32 of them being definitions related to HRI area.

However, it is difficult to interpret human-robot trust from these definitions only, but there is no doubt that such definitions can provide a number of common attributes that are considered as essential to understand trust concept across various domains. Billings, Schaefer, Llorens, and Hancock (2012) described that a large number of definitions refer in particular to expectations, confidence, risk or uncertainty, reliance, and vulnerability. Figure 2.1 is adapted from Billings et al. (2012) which represents a collection of the elements of trust definitions.

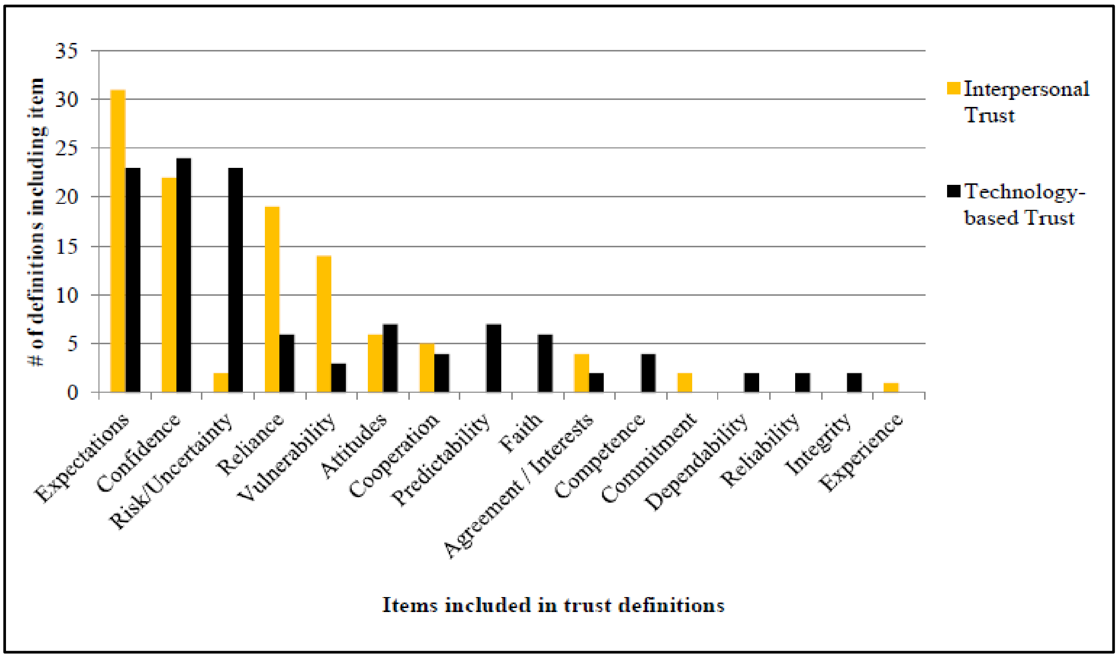

Figure 2.1: Variations in trust definitions

However, the definitions of trust among different types of interactions (*human-human*, *human-automation/machine* and *human-robot*) clarify an appropriate level to understand the most important components that should exist to construct trust relationship. Mayer et al. (1995) defined human-interpersonal trust as the willingness of a party to be vulnerable to the outcomes of another party based on the expectation that the other will perform a particular action important to the trustor, regardless of the ability to monitor or control that other party. This definition has illustrated the significance of two elements which are vulnerability and expectation between the trustee and the trustor.

In the same spectrum, Lee and See (2004) explained that both uncertainty and vulnerability are also important to the definition of trust in autonomous systems, when the authors defined trust as an attitude that an agent will help to achieve an individual’s goals in a situation characterized by uncertainty and vulnerability. However, Lee and See’s definition of trust highlights the importance of uncertainty and vulnerability as it relates to a situation but not specifically the actions of the trustor or the autonomous agent.

Numerous authors have considered trust concept as an essential element which enables human to predict the action of the partner (trustee). As well-known in various studies, any subject who trusts in other entities is called trustor while the trusted entity is called the trustee where the trustor relies on the trustee’s acts to achieve the desired outcomes. Trust plays an important and necessary role in order to instigate an interaction between humans and other party (such as human, machine, and animal) in uncertain environments. Typically, if a human doesn't trust the trustee, the interaction between them will be almost impossible. On the other hand, in internet services, users trust systems that provide protection from malicious entities and facility to access resources and so on (Aljazzaf, Perry, & Capretz, 2010).

Trust is the willingness of a trustor to rely upon the trustee's outcomes based on an expectation of good performance. Previously, trust has been defined by many authors as the capability to forecast other's dispositions. Also, trust has been defined as the total belief in the trustee to act securely and reliably in specific domain with no fear of risks due to this dependence. The authors considered that trust is configured from several features such as security, dependability, reliability, and competence, according to different environments where trust will arise. Trust is also the acceptance of human(s) to depend on the partner (trustee) to do a particular task in an uncertain environment despite lack of information about the trustee (Barbalet, 2009).

2.2 The Nature of Trust

There are some studies that regard trust not to mean an act but rather, a mere decision (Hardin, 2001), whereas other studies regard trust as a multi-layer concept that includes disposition, decision and act (Castelfranchi & Falcone, 2010). On the other hand, Schoorman, Mayer, and Davis (2007) have claimed that trust sometimes exists from one side (not mutual). In addition, there is a significant attribute called the degree of trust as mentioned by Castelfranchi and Falcone (2010). The authors presented the idea that the confidence that the trustor has in the personal trust, eventually represents an essential element to the degree of trust.

Trust is by nature contextual, that is, mutual trust between individuals and other humans or artificial artefacts such as computers or robots to play certain roles. Some authors claimed that trust differs across domains, just like trustees have different abilities in different domains. One study provides a broader concept of situational trust by defining trust as a measurable belief that the trustor has on the competence of the trustee in behaving in a dependable way, within a given context and relative to a specific task (Dimitrakos, 2001). On the other hand, one of the earliest research on computational trust (Jøsang, Ismail & Boyd 2007) considered trust to be situational, and provides this clarifying example: '*I may trust my brother to drive me to the airport, I most certainly would not trust him to fly the plane'* (Urbano, Rocha, & Oliveira, 2013).

There are five types of trust with their sub constructs which can be measured with subsets of each of them. These types illustrate the facets that can highly contribute to give clarification for trust definitions, such as risky or uncertain conditions, feelings of security or confidence, and absence of control and so on (McKnight & Chervany, 2001).

The first type of trust is *Trusting Based on Intentions*. This means that the individual intends to depend or is willing to depend on the other party's acts with a feeling of relative security (the meaning of the term ‘relative security’ is that every person has various degrees of security feeling or confidence to depend on others) in spite of the lack of control over that party, and even the possibility of negative consequences occurring.

In fact, the definition of Trusting Based on Intentions represents four elements combined: the disposition to rely on another is essential and necessary to trusting based on intentions, by depending on other entities the trustor will be vulnerable to the results of trustee, a feeling of security means the trustor feels comfortable and confident (without concerns or being fearful) towards the possibility of depending on others and feelings of security reflect the affective side of trusting based on intentions to feel secure, confident, assured or expectant. Hence, trusting intentions involves willingness without condition over the other party.

The second type of trust refers to *Trust-related Behavior* where trust-related behavior means when one gives another person a fiduciary obligation by acting in such a manner that the other could betray them. Trust-related Behavior means that a person voluntarily depends on another person with a feeling of relative security, even though there are possibilities of negative consequences. According to many studies, trust has been defined as a behavior and trust-related behavior is acceptance of risk. A trustor who behaviorally depends on a trustee gives the trustee some measure of power, because dependence is the obverse of power. One who depends on the other person places the other in a position of dependence-based power over one.

Therefore, by definition, behaviorally trusting another voluntarily gives the other dependence-based power over them. In addition to the definition of trust-related behavior, many procedures can make one depend on another, for instance: cooperation, information sharing, transacting business, informal agreements, decreased controls, autonomy, amongst many others. Trust-related behavior makes the trustor dependent upon the trustee just like cooperating with another makes one dependent on the other.

Trusting Beliefs is the third type of trust and this means the extent to which the trustor believes and with feelings of relative security that the trustee has beneficial traits to be trustworthy in order to serve the trustor's interests. Based on the definition of trusting beliefs, there are four sub-constructs illustrated with the relationship between the trustor and the trustee. The fourth type is *Trusting Belief-Competence* which means that the trustee securely believes the other person has the positive potential to do a specific task in the certain environment. For instance, when a consumer believes that the vendor of internet can provide the services and products in a proper and convenient way.

Trusting Belief-Benevolence is another sub-construct meaning that the trustor securely believes that the trustee partner cares enough and has the motivation to act well. *Trusting Belief-Integrity* which means the trustor securely believes that the trustee will make good decisions, tell the truth, and fulfil promises, such as when the trustee delivers products or services or keeping information private. The final sub-construct is *Trusting Belief-Predictability* which means the trustor will securely believe the trustee's actions and have the ability to make forecasting in order to make a decision.

The aforementioned constructs of trusting beliefs provide a robust foundation for trusting intentions in the other partner in conjunction with achieving the definition for willingness. However, just being willing is not sufficient because they may not be capable to help. On other hand, any entity with integrity will prove a willingness to help by acting effectively and ensuring that the trustee's behaviour will not change over time. A trustee who consistently (*predictably*) shows to be willing (*benevolent*) and have the ability (*competent*) to serve the trustor’s interest in an honest, ethical manner (*integrity*) is indeed worthy of trust.

The next type of trust to be considered is *Institution based Trust*. Institution based trust means the trustor believes that the existing circumstances will lead to circumstantial success in an uncertain environment and with feelings of relative security. This construct arises from the sociological assumption that people can rely on others because of structures, conditions, or roles that provide assurances that everything will go well. Institution-based trust refers to beliefs about those protective structures and not about the involved people. However, institution-based trust affects interpersonal trust by making the trustor feel more comfortable about trusting others.

There are two sub constructs of Institution-based Trust; *Structural Assurance and Situational Normality*. The main idea of structural assurance is that the trustor strongly believes that situational success will be done according to the presence of legal resources, promises, protective structure, guarantees, contracts, and procedures, as well as, structural conditions readily building trusting intentions. For instance, the person who is using the internet would encrypt his private information (structural assurance) to protect that information from getting lost.

Situational Normality on the other hand is another sub construct which means the trustor in a non-safe environment will securely believe that the situation is ordinary to situational access. Moreover, normality mirrors the idea whereas trust is the perception that explicates the circumstances in the situation that is working as normal. Thus, situation ordinary or normality is probably relevant to both intentions and trusting beliefs. An example is when employers who feel in good condition about the entire settings and roles in their existing work will probably have trusting beliefs about the people in that area.

The last type of trust considered is *Disposition to Trust*. This refers to the extent or the degree to which there is a consistent tendency to be willing to depend on the partner in certain circumstances. The definition of disposition to trust does not exactly refer to a person's trait, but rather, it means that the individual has a general propensity to be willing to depend on others. For instance, some employees when asked whether they trust their new boss or not, they replied that they generally trust new people. In the same way, the trustor tends to be willing to depend on others generally.

Disposition to Trust has two sub constructs, *Faith in Humanity* and *Trusting Stance*. Faith or conviction in humanity explicates assumptions about people, whereas trusting Stance is same to personal plan. From psychological perspective, faith in humanity refers to one assuming in common that others are normally honest, competent, expectable, and benevolent. Faith in humanity is unlike trusting beliefs in which it discusses general people while trusting beliefs means a specific people. For instance, when someone is drowning, he/she believes that unknown people will come to provide assistance.

Trusting Stance is that one person will yield better results or outcomes by working with people who are trustworthy enough, which means it is a plan or personal selection to make trust with others. It is relevant to trusting intentions, which might not be made only on beliefs about peers, while Faith in Humanity (FIH), much probably to be an antecedent of trusting beliefs (in other people) because it relates to assumption about peoples' attributes.

2.3 Forms of Trust

The different forms of trust are discussed in this section as listed below.

2.3.1 Human-Human Trust

Trust is a dynamic process consisting of interaction between two individuals, the first one is the trustor and other one is the trustee. Moreover, in a bid to understand the importance of developing trust principles, there is an urgent need to consider the characteristics and the traits of the individuals. Researchers have suggested three trustee characteristics that lead to mutual trust between individuals: ability, benevolence, and integrity. Ability is an essential element of trust and a set of skills and characteristics that belong to the trustee while benevolence is the belief that the trustee has good willingness to help and to act efficiently in some specific task according to some information about the history of the trustee. Integrity on the other hand is the belief that the trustee committed acceptable values to the trustor (Mayer et al., 1995).

Humans are continuously facing many complex and interlinked social situations and hence, there is need for the trust principle, despite the fact that the behaviour of others' is sometimes unpredictable, or maybe there exist some degrees of uncertain situations towards the other person, or the presence of interdependent situations such as uncertainty and risk. All of these considerations enable humans to feel confident with their partner and thus provide an opportunity to work in proper social contexts and that the other person will work in a positive way and he/she is trustworthy.

Trust is essential in a set of interpersonal social interactions; therefore it has been examined in several fields of disciplines such as economics, medicine, law, religion, sociology, and psychology. In psychological fields, researchers have focused on identifying characteristics that make people influenced to be somehow trusted (Simpson, 2007). In addition, individuals vary in characteristics and traits (attributes) and this will have an impact on human judgement whether to trust or distrust in making decisions for a specific task

Righetti and Finkenauer (2011) proposed that in interpersonal interactions and in dynamic relationships, people can detect the rate of another person's self-control (such as a dispositional trait and temporary state) and eventually have an influence on their perspectives of other person's trustworthiness. Trust arises in a situation when three conditions are met. Firstly, the necessary condition is interdependence between two humans: A person's outcome cannot be achieved without the reliance upon another person.

The second condition is risk awareness: A person realizes that something will be lost in the case that his/her partner does not act in a positive way. In other words, when two persons are sharing same interests (interdependent interests) there is room for slight risk because the desired outcomes will be their own aim. Thirdly, a necessary condition is free choice: this includes interpersonal uncertainty where people have the free choice to depend on the partner in order to achieve the best outcomes. Barnes and Jentsch (2010) suggested that drivers of interpersonal trust may also drive towards human-robot trust.

2.3.2 Human-Animal Trust

The concept of human animal interactions provides a basic platform to a new way of how robot and human interacts (Coeckelbergh, 2011). Human-animal interactions are distinct forms of partnership which typically bring direct benefits to the human physically, emotionally, and at cognitive levels (Wilson, 1994). The relationship between humans and animals entirely depends on the functions of animals during the period of interaction (Coeckelbergh, 2011). There are four primary types of partnerships: companionship, for instance the pets as it is well known in adoption of some kind of dogs to guide the blind individuals, service, for instance assisting people to live more independently, therapeutic assistance, and high-risk teams such as search and rescue and military operations (Finkel, 2012; Helton, 2009).

It is necessary to mention that humans must spend enough time with their animals regularly in order to increase the ability of human’s prediction about all possible reactions of their animals in most cases. Choosing animals is to meet different needs based on their natural abilities, characteristics, and functional capabilities in the context of the determination of robotic roles. Hence, animals are regarded highly and valued in these partnerships because they are able to replace or increase human skills. Trust is fundamental for building effective interactions between entities, it arises by having two ideas: the first idea is to know how a partner will respond and secondly, trusting one’s self to translate a partner’s behavior.

A human must trust that their animal partner will perform the mission they have been trained to do. However, at the end, the animal will behave like an animal, displaying tendencies and behaviours that are based upon its instinctive reactions (Keaveney, 2008). Human-animal trust involves risk and uncertainties due to the human not always being sure of the animal's response, or indeed vice versa. Therefore, a productive interaction between humans and animals depends on cooperation and mutual trust (Armstrong Oma, 2010).

Typically, robots are designed with anthropomorphic traits such as zoomorphic attributes and behaviours. In order to achieve this industrial aim, the designers in robotic industries field consider the physical appearances and functional capabilities which are essential to accomplishing trust condition. However, human-animal interaction has offered and extended the perspective of increased trust in robotic systems. Robotic industrials have the tendency to simulate physical and behavioural aspects of animal realm to enhance interaction by trust condition and make humans rely on robotic systems in their daily life (Hancock, Billings, Oleson, et al., 2011).

Humans & Robots **Humans & Animals**


Tools


Replacing / Enhancing Human Capabilities


Augmentation Human Capabilities


Providing Comfort

Team Member

Figure 2.2: A Comparative Representation of the Transition from Tool to Team Member between Human-Animal Interaction and Human-Robot Interaction

When a robotic system is regarded as a tool, the robot is assigned to perform certain tasks while the human will take the supervisory role. The reliability of the robot's functional capacity is critical for developing trust between human and robotic partner. In the same spectrum, any failure in the robot's functional capacity may lead to increased external risk to the human population within the area. At the next level, robots are designed to replace or improve human capabilities. Hence, the robot is assigned as an industrial robot in the manufacturing field according to its high accuracy to develop products, as well as providing more safety of human employees such as lifting or moving heavy objects, work load and so on. The human takes a supervisory role, and trust is related to dependability and efficiency.

When the robots are designed to increase human capabilities such as robotic surgical systems, at this level, trust becomes essential due to the potential of the robotic systems to extend human operator’s capacities. In addition, there is a trust relationship with a bystander (e.g., surgical patient, assisting nurses, etc.). Communication and feedback, as well as the aforementioned reliability and dependability of the robot, are essential to the trust relationship at this level of Human Robot Interaction. The continuous development of the robotic system to create robots can be involved in human team and interact with them intelligently in order to play a team member role in HRI.

At the fourth level, the robot will provide emotional support, for instance, social robot giving care to elderly people. Social and therapy robots are often included in this level of HRI. In order to achieve this goal, most robots are designed with zoomorphic or anthropomorphic traits and behaviours to reach this goal, as well as instil trust. Both the physical appearance and functional capabilities are essential to trust development. At this time, it can be argued that robot development has not yet reached the level of a team member. However a human-animal interaction has provided a structure to achieve this purpose. Robot tends to emulate physical (e.g., mobility characteristics) and behavioural aspects of animals ranging from tool to integrated team member. Thus, examining the nature of human-animal relationships can increase our understanding of how humans may interact with and trust certain technology, such as robots (Billings, Schaefer, Chen, Kocsis, *et al.*, 2012).

2.3.3 Human-Machine/Robot Trust

A human's trust in machine/robot is an important element typically needed to ensure that any functional role of application will be effective, as it is well known that robot belongs to the automation field and automation here is defined as the execution by a machine agent instead of humans (Parasuraman & Riley, 1997). Automation agents are always implementing specific tasks in structured environment and monitored by human operators including the assistance as cognitive aides (e.g, airport baggage screener), perceptual aids (e.g., collision avoidance systems in cars) and control aids (e.g, adaptive cruise control).

Social robotics researchers have developed Mahoney's definition to embed the perspective of robot as an autonomous system that acts based on certain information from human operator and may not be completely controlled by human supervision. Hence, robots are designed to enhance human capabilities especially in the manufacturing domain where it helps to maintain safety of human employee such as in lifting of heavy things. From those aforementioned facts, the human role is merely for supervising, so it can be seen clearly that trust perception is related to dependability and competences.

Yagoda (2011) viewed a robot as "the interaction of intelligent and autonomy of the technology". She justified that according to interaction, robots are definitely different from machine, such as a refrigerator's categories being classified as high level of autonomy and low level of intelligence while an expert system being classified as high level of intelligence and low level of autonomy. Researchers have suggested that the robot is a complex system characterized by a degree of automation and cognition and robots come in various forms depending on the purpose of using the robotic systems. In addition, robots have the ability to work and move in unstructured or uncertain environments (Schilling, 2013). The reliability of the potential of robotics functional is a very important part for trust enhancement due to some cases such as failure which will lead to decrease in sensation in trust.

The expected risks and uncertainty environment are two important effective aspects on the calibration of trust development (De Santis, Siciliano, De Luca, & Bicchi, 2008). The internal and the external risks of the environment can have an effect on trust development. In the field of automation, Beckles, Welch, and Basney (2005) have considered the internal risks such as wrong data in the automation itself, as a result that the system will be untrusted from the operator's side. Also some researchers have examined automation based on pilot alerting systems for the external risk about physical terrain; they concluded that trust is an important condition to interact with these automation systems according to the safety risk in human-automation interaction.

2.4 Human Robot Interaction

Human robot interaction is an interesting and quite challenging research area, where various interconnected fields need to be studied to explicate clear understanding about this field of the study like cognitive science, social science, artificial intelligence, psychology, computer science, robotics engineering, and human-computer machine interaction (Dautenhahn, 2007a). Thus, the main objective is to understand and embody the natural way by which a human can communicate and interact with a robot. As well as, due to the embodied nature of the interaction, where human and robot need to control their interaction in real time and space which means face to face interaction, the quality of the communication is highly related, but different from human-computer interaction. Meanwhile, communication between human and robot may occurs in various forms, but these forms are mostly affected by whether the two parties; human and robot, are near or roughly close to each other or not. Therefore, interaction and communication process can be divided into two general categories which are;

- **Remote interaction**: The human and robot are not in the same place and are disconnected spatially or even temporally such as, the Mars rovers are separated from earth in both time and space.

- **Proximate interaction:** The human and robot are located in the same zone such as service or domestic robot (may be in the same room as human).

Based on these general categories, it is useful to differentiate between different applications which need mobility, social interaction, and physical manipulation (Goodrich & Schultz, 2007). However, in recent years, increasing attention have been made in human robot interaction, where several kinds of robots have been made in the society of robots such as assistive robots in healthcare aspects (Rodić & Jovanović, 2014), domestic robot (Casiddu *et al.*, 2014), personal robots (Schwarz, Stückler, & Behnke, 2014), educational robot (Lu, Wang, & Liu, 2014), and many others. In addition, the researchers in the field of human-robot interaction are trying seriously to develop socially interactive robot that can interact and communicate with humans naturally and with intuitive manners (Rodić & Jovanović, 2014).

2.5 Theories and Important Concepts of Trust in Human Robot Interaction:

Human Robot Interaction vision has been to create and study robots that humans rely on based on their needs to do task instead of humans (such as cooking and cleaning, or to more intellectual and social endeavors of entertainment and caregiving), and the human individual will represent a supervisory role (operator). Also, the vision aims to study if the engagement between human and robot is a highly significant element in the daily life. Current examples of long term interaction between humans and robots include behavioral therapy, lifestyle modification, pharmacotherapy, surgical interventions, and so on.

Recently, there is a debate on the important factors in creating a relationship where three factors were regarded as the most important: engagement of the user, trust of the system, and motivation to use the system where the sociable robot has different potentials with which to assist in engaging humans in natural social modalities. There are several key features that are desired for robots to be used in a long-term interaction of Human Robot Interaction (HRI), such as the ability to look at the user's eyes is important to make the interaction. In addition, some set of features that enable social interaction include observing robot behavior such as eye contact, head, arm, and hand gestures, speech, speech recognition and so on are needed, but the accurate set of features depends on the expected type of interactions.

Hence, it is important to understand how users interact with robots over long periods. There are numerous reasons for investigation in the reasons leading to long term interaction. Longitudinal investigations are highly useful to examine changes in user behavior and experiences and explore the methods to achieve the potential of long-term interactions such as the novelty or modernity effects (Leite, Martinho, & Paiva, 2013). The robots created positive results in the area of long-term interaction of social robots in therapeutic or health fields. Robots served different groups of people such as elderly, autistic children and adults. In addition, the social features in the robot's capabilities have been playing an important role for long-term engagement. Even though social robots are well accepted by children or other users in educational environments, the robots should be able to perform simulation for the complex social behaviors in order to engage children in the long-term.

Some examples of social robots are listed in Table 2.1 , primarily to illustrate the main aspects that robots play as important roles in order to interact over a long period and these aspects include health care and therapy, education, work environment, and home environment (Leite *et al*., 2013).

Table 2.1: Examples of Social Robots

References	Robot's Name	Description and Design	Domain
Francois et al. (2009)	AIBO	Dog-like behavior such as wag the tail, responds to touch.	Health care and therapy
Sabelli et al. (2011)	Robovie	Remotely conversation and child-like behavior. Interact by interviews, direct observations.	Health care and therapy
Hyun et al. (2010)	iRobiQ	Move head and arms, navigate in the environment, express, emotions. Interact with kids by doing interviews	Education
Kanda et al. (2010)	Robovie	Guiding, rapport building, identify repeated users, advertisement	Work environment and public space.
Sung et al. (2009, 2010)	Roomba	Vacuum cleaning, move around the house.	At home environment
Klamer et al. (2011)	Nabaztag	Personalized health conversations; users interact by using yes- and no-buttons	At home environment

However, there are some guidelines to improve the design of social robots for long-term interactions.

- **Appearance:** Embodiment might play an important role in the first impressions and future expectations that people will create about a robot (such as human or animal form). A robot appearance should consider the application domain in addition to the robot's behavioral and social capabilities.
- **Continuity and Incremental Novel Behaviors:** Researchers and designers in the fields of robotic systems trying to perform a set of modalities towards improving and enhancing the robot's capabilities towards creating effective interaction between human and robots.

- **Affective Interactions and Empathy:** It means the expression of emotions and other non-verbal behaviors especially in social robots which have been the most extensive features used in recent studies.
- **Memory and Adaptation:** When robots are able to identify or distinguish different users and call them by their names. Another relevant aspect that remains nearly unexplored is memory. Memory will definitely make social robots more flexible and personalized to particular users, regardless of the application domain in which they are assumed to operate.

2.5.1 Human Robot Short-Term Interaction

Human robot interaction have become an interesting area for the designers of robotic systems due to the ability to achieve a high potential of attracting and engaging users by applying significant factors for both the implementation and appearance of robot conducive for effective interaction. According to many studies, HRI can be divided into two types: *long-term interaction* and *short-term interaction*. This study is aimed at using the concept of short-term interaction between humans and robots. This interaction occurs when many people interact with a robot for a short time (limited time). For instance, the interaction occurring between hotel receptionists, information kiosks, or with tour-guides. In the same vein, people spend only limited time with a robot and in this limited time, robots are supposed to be able to interact with people in novel modalities.

For example, an interactive museum tour-guide robot (mobile robot) was used as a case study to illustrate the concept of short term interaction, these robots implemented the function of interactive tour-guides, deployed in museums to guide people and explain to them what they see along the tour. The most critical task for this robot is to attract people to participate in a new tour and maintain their attention. Once a person requests a tour by the touch-sensitive screen on the robot, the robot will start to move from one place (exhibit) to another. People might block the robot's path. The robot called *Rhino* would honk a horn to ask for space in order to attract their attentions, whereas *Minerva* applied different strategy to make people aware of its intentions, in both of these cases, people interact with robots in short term interaction.

On the other hand, Minerva was developed in terms of enhancing the interaction with humans. Minerva has humanoid features which were certainly conducted to significantly impact people's perception and eventually the effective interaction of the robot. The typical interaction with a tour-guide robot lasts for less than fifteen minutes, during which a person has to fully grasp the concept and understand how to make use of the machine. In emotional aspects, Minerva is capable to express different states such as "happy", "neutral" and "sad" expression. For example, if someone is blocking the path of robot, then the robot will ask a person to stand behind it. If the person still does not move, then Minerva frowns and becomes even more demanding. In addition, observation of interaction between museum visitors and Minerva usually results in a quick response from anyone in the way. Thrun, Schulte, and Rosenberg (2000) observed that the humanoid features such as the motorized face and the utilization of emotions resulted in increase in the believability of the robot and enabled people to understand what the robot demanded.

2.5.2 Cognitive Modeling Approach

Cognitive modeling entails simulating a set of cognitive processes that has its observable features in the real world situation so as to have an extensive knowledge of its trends and possible prediction of the natural processes underlying such process through the implementation of certain input parameters. Though, previous studies in this area have explored symbolic framework to investigate reasoning ability and other cognitive functioning processes. But, evidence from these studies provides little correlation between the mental simulations and the intrinsic brain mechanisms. Meanwhile, the emanation of connectionist models enhance the exploration of the mental processes and their correlation to the basic anatomical concepts and physiological mechanisms, but in view of their significance in the study of cognition, little contribution has been made to the involvement of the emotional processes.

Cognitive modelling has become an effective and flexible approach to the study of cognitive processes. It has been employed in investigating the fundamental nature of various cognitive functionalities and psychology for the development of intelligent systems capable of maintaining social interactions with humans through the understanding of human cognitive functions and behaviours (Asada, 2013; Bonchek-Dokow & Kaminka, 2013; Bosse, Both, Duell, Hoogendoorn, Klein, Lambalgen, *et al*., 2013; Breazeal, Buchsbaum, Gray, Gatenby, & Blumberg, 2005).

Furthermore, it has been used to simulate related behaviours in specific domains of interests by assigning the corresponding computational processes on to cognitive functions to produce executable computational models by which the detailed simulations are performed. To sum it up, there are numerous cognitive models which have been established in different facets to explain an intrinsic notion of cognition. Table 2.2 further illustrates the summarization of related cognitive models.

Table 2.2: Related Research on Cognitive Computational Modelling

Author/Year	Title	Technique	Discussion
(Farrell & Lewandowsky, 2010)	Cognitive Models as Aids to Better Reasoning in Psychology.	Differential equation	Cognitive modeling was used to explore human reasoning.
(Bosse, Pontier, & Treur, 2010)	A cognitive model based on Gross' emotion regulation theory.	Formal logic	The authors employ cognitive modeling technique for emotional regulation.
(Bosse, Both, Duell, Hoogendoorn, Klein, Van Lambalgen, et al., 2013)	An Ambient Agent System Assisting Humans in Complex Tasks by Analysis of a Human's State and Performance	Differential equation	The authors employ cognitive modeling to explore human knowledge for intelligent assistive system.
(Thilakarathne & Treur, 2013)	Cognitive modelling of Human-Robot interaction estimating other's internal state.	Differential equation and formal logic	Cognitive modelling was used to explore human-robot interaction through intention estimation.
(Thilakarathne & Treur, 2013).	A cognitive model for intentional inhibition of actions.	Differential equation and formal logic	The authors employ cognitive modeling to explain the concept of selection and intention inhibition.
(Mui, Mohtashemi, & Halberstadt, 2002)	Cognitive Model of Trust and Reputation	Formal logic	The authors employ cognitive modeling to understand the dynamics of trust and reputation.
(Both, Hoogendoorn, Klein, & Treur, 2010)	Cognitive Modeling and Analysis of Therapeutic Interventions for Depression	Differential equation and formal logic	The authors employ cognitive modeling to demonstrate the concept of interventions therapy for depression.

2.6 Computationally Modeling Interpersonal Trust

Lee, Knox, Wormwood, Breazeal, and DeSteno (2013) constructed a computational model which has the capability of predicting the degree of trust that a human has toward their novel partner by depending on trust related nonverbal cues. A set of nonverbal cues have been identified to demonstrate the human mind's readiness to assess the trustworthiness of a social robot. Lee et al. (2013) utilized the construction of hidden Markov models to investigate temporal relationships among the trust related nonverbal cues.

As mentioned before, trusting behaviour represents a person's willingness to assist the other partner to do a certain task, hence, a computational model is capable of predicting similar humanlike accuracy and these predictions are based on the nonverbal behaviours.

Lee et al. (2013) identified four nonverbal cues: *face touching, arms crossing, leaning backward*, and *hand touching*; the findings of these nonverbal cues did not offer significant predictive ability when examined in isolation. Although, in computational model of interpersonal trust experiments, trusting behaviour was measured as participants' exchange action with their partner during an economic game. To confirm these findings, there was a verification of the nonverbal cues set through a human-subject experiment. By utilizing social robotic platform, they took advantage of its programmable behaviour to penetrate which cues were engaged for each participant.

2.6.1 Computational Modelling

The concept of the computational model refers to the mechanism of simulating a set of natural phenomenon (such as fatigue, stress, anxiety, trust, distrust, phobia and so on) in order to have a deep understanding of these processes and to predict the outcome of the natural processes which are represented by a specific set of input parameters. These models are priceless since they allow researchers to study and having revision on complete relations that perhaps may not be arranged out by righteously experimental approaches, and to create rough calculations that cannot be made without difficulty by extrapolating from the existing data (Ellner & Guckenheimer, 2011).

A constructed computational model is accomplished by stimulating certain key behavior in the particular area and researchers have considered it as highly interesting. For instance, in the field of neuroscience, theoretical neuroscience applies computational modelling to help in understanding and explaining the mechanisms of cognition. This means developing clear mathematical models of the processes which go on in the brain when we act, think, perceive, learn, or remember a certain task. In spite of the evolution of software and influential brain imaging machines that permit scientists to investigate into broader details of our brain activities, these technologies are still unable to describe the detailed interaction between all of those involved activities.

2.6.2 Computational Cognitive Models

Computational cognitive model is an approach to help robotic systems designers to create a robot that is able to interact with humans cognitively and naturally by analyzing human's traits and behavior which are targeted from the interaction. Sun

(2006) built computational cognitive models with specific high level of human skills which were identified to be critical for successful collaboration and interaction in order to build better human-robot interaction. These cognitive models were used as reasoning mechanisms on the robot and in addition allowing the robot to make decisions which results in efficient interaction with humans. Sun (2006) proposed that by adding computational cognitive reasoning components into intelligent systems, robots will be conducive to the three following advantages:

- Giving the system cognitive models will be fruitful in order to foster the human system interaction by creating a form of cognitively reasonable representations and qualitative reasoning.
- Applying cognitive models on intelligent systems such as robots will be more successful in building more effective interaction in human robot interaction.
- The intelligent systems are interacting with humans by providing natural behavior which resembles human behavior and eventually result in developing human robot interaction.

Meanwhile, cognitive models of human performance have been developed in creating the intelligent system to be able to act and mimic human behavior and it can be directly compared between intelligent system performances and humans.

2.6.3 Cognitive Model for Trust Dynamics of Interfirm:

The main objective of conducting this model is to provide an integrated view that may address trust formation and maintenance and show how trust between firms evolves over time. This model is built based on using Fuzzy Cognitive Maps theory (FCMs) and is aimed at showing how the trust of interfirm evolved. As it is well known, trust is characterized by complexity nature and this nature needs to be analyzed and evaluated, thus, there is need to use soft computing modelling techniques. As defined previously, trust is a dynamic phenomenon (dynamic process) and this dynamic nature of trust forms the great challenge when measuring trust or predicting trustworthiness. The familiar definition of trust is that one believes and is willing to be dependent on another agent. Admittedly, trust is identified as a psychological and cognitive state based on specific beliefs in uncertain environments.

FCMs are suitable for causal reasoning to perform some cognitive operations that lead to trust decisions. The nodes in FCMs refer to the most related features in the environment, while links between these nodes denote the relationships between those features. The hypothesis in this model is: team trust in the interfirm is a function of triple essential components; the trustor, the trustee and its surroundings to illustrate the effects from these different factors and the map is constructed using three levels. Any change in any of these three components influences trust and the development of trust (Zhu & Zhang, 2008).

2.6.4 Cognitive Model of Interfirm Trust Dynamics Based on FTCM

Fuzzy cognitive maps (FCMs) model has two significant features. First, causal relationships between nodes take trivalent weight values [-1, 0, 1] and it would be more practical to utilize fuzzy weights like strong, medium, or weak for complicated FCMs. Secondly, it supports dynamic involving feedback, where the effect of change in the concept node influences other nodes, which in turn can affect the node to perform the change (Zhu & Zhang, 2008).

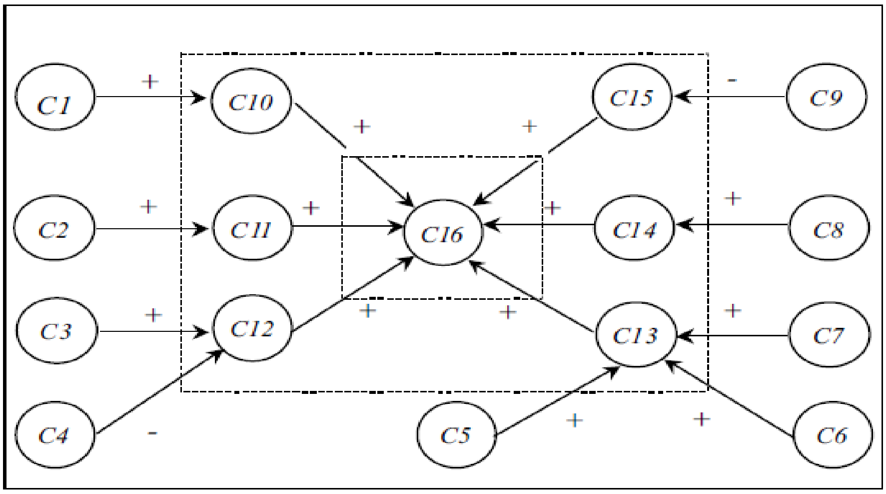

Figure 2.3: The Cognitive Model of Interfirm Trust

The valid nodal values of 1 denotes a positive influence or increasing in beliefs, while -1 denotes a negative influence or decreasing in beliefs, as well as 0 of no effect.

2.6.5 Model of Operator Reliance on Automation

Another model of trust which is developed by Desai (2012) is the operator reliance on automation model. This study has described and clarified trust relationships related in HRI.

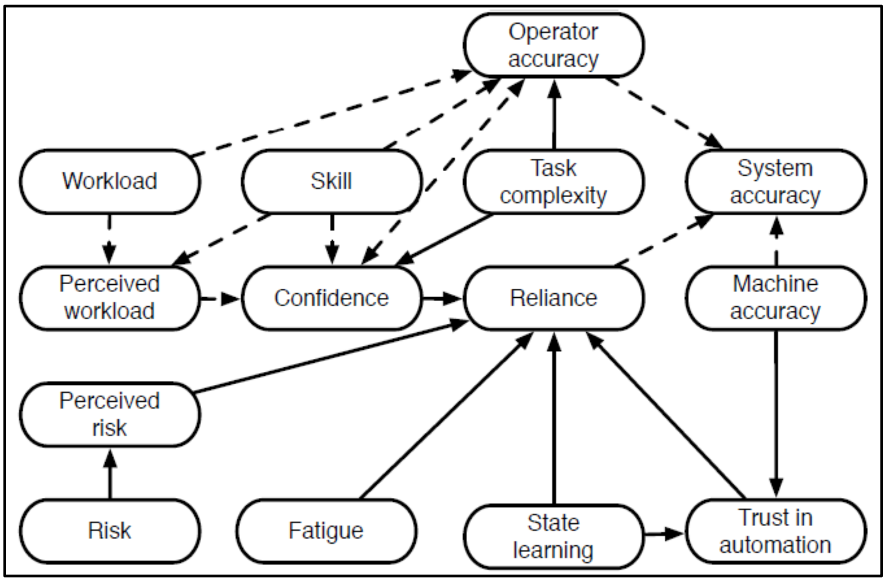

Figure 2.4: Model of Operator Reliance on Automation

The constructed model attempts to identify automation reliability as the most effective factor in trusting automated system, such as, they claiming that low reliability results in low operator trust and vice versa. In addition, the study considered risk and reward to be motivating factors for achieving better performance. They concluded that the existence of risks and failures may lead to low motivation to interact with automated teammate. Moreover, self-confidence is another important factor which is denoted by many researchers, since the operator has confidence in their own capabilities more than the automated system and therefore operators tends to prefer not to interact with a robotic system. Also the experiences of users (user who have interacted with similar system) play an important role in providing an appropriate level of confidence for non-human teammates.

2.6.6 Theoretical Model of Human-Robot Trust Development

Trust is a pivotal key component in human robot interactions and it can impact the outcome of interaction. Hence, the factors that may influence trust in robotic systems have been investigated to explore the dynamic state of trust in HRI. Sanders, Oleson, Billings, Chen, and Hancock (2011) highlighted the baseline for developing initial model of trust from empirical and theoretical literature review. A constructed trust model involved four facets, robot performance, robot reliance, individual differences, and collaboration, as factors of trust in human robot interaction. These facets were indicated to antecedents of trust.

Sanders *et al*. (2011) study aimed to build an initial human-robot trust model, and the initial model provided a lot of trust related factors, starting from the attributes of robot such as anthropomorphism, likability, perceived intelligence, and perceived safety to human interaction factors (e.g., usability, social acceptance, user experience and societal impact).

2.7 Human- Robot Trust Model

Robotic systems are becoming an effective gadget in order to facilitate and enhance our daily life, and the need to develop an appropriate level of trust in such a system has become an increasingly disputed subject for the designers and researchers in the robotic systems field. However, the targeted aim is to provide a robot being perceived by humans as an interdependent partner. In order to achieve this goal, humans must accept and trust a robot before effective interaction can occur.

Billings, Schaefer, Chen, and Hancock (2012) constructed a trust model based on literature review, and the experimental and theoretical studies in human robot interaction domain especially that unmanned robot that has been adopted in the combat operations (military operations) and to which extent the acceptance of soldiers to rely on the robot's behaviour (performance) might be influenced by different surrounding factors. As we have mentioned previously, trust can be one of the most effective factor on the success of human-robot collaborations and the existence or absence of the trust factor has a great influence on the use of robotic systems (Freedy, DeVisser, Weltman, & Coeyman, 2007).

The constructed model was seeking to identify crucial antecedents (factors) of trust in human-robot interaction. Based on the collected information, from literature review, experimental and theoretical studies in trust researches, the model was conducted by identifying effective factors on trust in robotic systems which were then classified as: factors related with robot (such as, robot characteristics, performance, appearance, proximity), factors having link with human considerations, and environmental characteristics. The trends in the literature indicate that higher trust is associated with higher reliability and high level of optimal implementation (Billings, Schaefer, Chen, & Hancock, 2012).

2.7.1 Factors of Human-Robot Trust Interaction

The ability to trust in robots is an ever-present challenge, especially in the aspect of performing supervisory control tasks. Trusting in robots may be misleading due to the fact that trust has the concept of a prescribed behavior and hence the word 'calibration' seems to be a more fitting expression because it implies that when a supervisory control task is in progress, the operators will intervene only when they have a cause to believe that they have a superior decision in relation to the robotic system's decisions (Dzindolet, Beck, Pierce, & Dawe, 2001).

In addition, it is suggested that humans will exaggerate their own particular choices in correlation to robotic arrangements. However, the propensity for people to over depend on robotic frameworks has been demonstrated by various studies (Chen & Terrence, 2008, 2009; Dzindolet, Pierce, Beck, Dawe & Anderson, 2000; Chen, Barnes & Sciarini, 2011).

Previous studies have mulled over variables influencing and developing trust while interacting with robotic systems, and that includes both external (*extrinsic*) factors, in that they are identified with the attributes of the environment or the robots, and internal (*intrinsic*) factors, which are identified with the qualities of the operator. Using one of the important concepts which is the *Extrinsic Factors,* scholars proposed that the mental connection of the choice to trust decides the propensity of the operator to neglect (under depend) or abuse (over depend) robotic frameworks (Parasuraman, Sheridan & Wickens, 2000).

This sort of unreliability has an essential effect on the operator's perception and reaction to framework cautions when the performance are frequently false-alert inclined or miss-inclined, taking into account the limit settings of the caution. The operators of robotic assignment (framework failure observation) execution are corrupted when the false alert rate of the alarms for the robotic task was high. As it were, high false alert rate diminished operator's compliance with robots (compliance was characterized as "the tendency to follow robotic information when it gives an alarm") (Levinthal & Wickens, 2006).

Lee and See (2004) characterized trust as the mentality that an operator will help to achieve certain objective in a circumstance described by instability and helplessness. At the point when people see robotic frameworks as colleagues, their disposition of abuse or neglect ought to grow as operators get to be familiar with the framework; consequently, suggesting that the "trust" choice is subject to the mentality of people towards robotic frameworks.

This intrinsic definition has interesting ramifications recommending that people view robotic frameworks similar to a human. This makes you wonder whether we can outline frameworks to perform human-like assignments, as well as to really act like people. For example, if robots can be made to be seen as human, then teaming relations would be more common as well as intuitive.

Placing focus on some specific factors that influence trusting in robots, the following sub-sections highlight how personality, physical appearance, believable behavior, behavioral cues, positive experiences and level of automation affect trust in robots. Some accompanying rules on planning frameworks that advance fitting trust in the robots is also highlighted.

2.7.2 Trust and Personality

Robotic systems became introduced into military context to enhance fighter’s capacities in complex and dynamic environments. While these systems are designed to complement the potential of humans, for instance, aiding in battlefield situation, decision making and so on, usually, robotic systems are misused or disused because the user does not have an appropriate level of trust in the robotic teammate and individual differences in terms of personality. One of the most important factors can be influencing human's trust toward robotic counterpart.

In addition, personality factor may also impact the development of trust in a robot, for instance, an individual's self-confidence (a user would like to trust in his or her own capabilities more than robot). Research has also indicated that there may be relationships between some of the Big Five personality traits and trust development in robotic systems (Oleson, Billings, Kocsis, Chen, & Hancock, 2011).

Oleson et al. (2011) provided support that the principal influences on human trust of a robot partner included attributes of human and the robot as well as the environment (situation) in which cooperative work was to be performed. Cramer, Goddijn, Wielinga, and Evers (2010) gave a further clarification in a study of vehicle assistive information systems showing that human trust in the technology as reported by participants was strongly related to personality type, the situational context, and the manner in which the autonomous agent presented information.

Individual personality factors also appear to influence the choice to become reliant on an intelligent, autonomous agent depending on the details of specific use case scenarios. These personality factors include likelihood of accepting innovation, perception and acceptance of different types of risks, and factors such as

Extraversion, *Openness*, and *Conscientiousness*. Certain personality factors, including high scores for Extraversion, Openness, and Conscientiousness are very important in some situations while in others, the tolerance for certain kinds of risk is dominant when it comes to deciding to become reliant on an intelligent, autonomous agent. This suggests that certain people may more easily accept a dependency on an agent, and conversely, others are likely to be extremely resistant (Atkinson & Clark, 2014).

2.7.3 Trust and Physical Appearance

Indeed, the majority of people around the world are not familiar yet with the interactions of robotic systems (the interaction with highly realistic looking robot revealed that people get more familiar with the robot over time), therefore, recently there are many scholars tending to examine different related factors in the robotic system field. Scholars have shown that the physical embodiment of the robot (presence of the robot) has an important impact in human robot interaction and how the physical appearances of robot would influence a human perception about the robot (Haring, Matsumoto, & Watanabe, 2013). Wainer, Feil-Seifer, Shell, and Mataric (2006) have illustrated by conducting a study that the appearance of robot (humanoid robot, animal like robot) has a significant role in developing human robot interactions.

Researchers have proved that robot with human-like appearances are more familiar for people to interact and respond to the robot's instruction quickly in short term interaction. Haring, Watanabe, and Mougenot (2013) mentioned that people prefer to interact with particular robot appearances and behaviour based on their personality traits. It has been shown that more introverted individuals tend to prefer mechanic robot appearance and extroverts preferred more humanoid robots. Also there are other factors like prior experiences with robots, prior relationships with non-human agents like pets and the subject’s personality which may influence the measurement of human trust in their novel agent partner (robot).

Scholars believe that the physical appearance of the robot can provide cues of the robot's functionality in human robot interaction (Fischer, 2011; Goetz, Kiesler, & Powers, 2003). For example, cleaning robots may be designed to look like maids, or administrative assistants may have the presence of a secretary. However, the absence of match appearance and functionality may be conducive to misuse or disuse of robotic systems and the users having no willingness to rely on such systems.

Furthermore, robotic systems which are interacting directly with humans also do not confer the opportunity for function-appearance alignment. Therefore, it is important for the user of a robotic system to fully understand the intent or purpose of the robotic system. This higher-level understanding of the robot's functions can be distinguished from the task model relating to specific actions of the robot. Such an understanding will help users of robotic systems to put the actions of the robot in the proper strategic context. This is important because in the future, robotic systems may interact with multiple users who may have different levels of baseline knowledge about the particular robot (Lyons, 2013).

Zeller and Smith (2014) designed a HitchBOT, the Canada’s first hitchhiking robot who unaccompanied, travelled across several places by relying on the strangers’ help solely. The HitchBOT’s physical appearance was designed to instill trust, and to make people want to engage and connect. The robot was designed at approximately the size of a 6-year old child, appealing to human behaviour associated with empathy and care.

The overall design, with a plastic body, pool noodle arms and legs, and matched rubber gloves and boots was meant to be funny. The low-tech look of the robot was intended to refer for the approach ability rather than introducing a complex and sophisticated gadget. Hence, this confirms the results of the study which concentrated on the physical appearances of a robot face and justified that participants acted as if they attributed human-like motivations to the robots (Haring, Matsumoto, *et al.*, 2013).

2.7.4 Trust and Believable Behaviour

Recently, studies in trust field have illustrated the existence of four significant categories of beliefs about the potential of trustee: Competence, Predictability, Safety, and Openness. Our goal is to tackle individually these human interpersonal belief structures to see how they specifically relate to a human's decision to be reliant on the intelligent and autonomous agents (Atkinson & Clark, 2014).

By taking the first category, competence; these are beliefs that represent a potential trustee's expertise, ability, skills, aptitude, etc.; these, and related constructs, are synonymous. The foundational studies reviewed and unified (as well as they could be) by others (Mayer et al., 1995; McKnight & Chervany, 2000), result in the notion of "competence" we employ here. With respect to intelligent, autonomous agents, we define the construct of competence as "detailed functional and specific knowledge in some domain, and the skills to apply that knowledge to problems of interest." The second category is predictability and these are beliefs that represent those facets of character that enable one to predict a potential trustee's behavior specifically with respect to completing a task or performing some service of value.

Predictability is at the core of hopefulness or optimism that a desired outcome, to be brought about by a trusted agent will occur. It has been shown to be especially important for trust in automation (Feltovich, Bradshaw, Clancey, & Johnson, 2007; Marble, Bruemmer, Few, & Dudenhoeffer, 2004; Muir, 1994). If someone makes a mistake, e.g., while performing math calculations, people will nevertheless predict that future calculations by that person will be reliable.

However, if the mistake is caused by machines, then people lose confidence in their ability to predict the future reliability of the machine (Beck, Dzindolet, & Pierce, 2002). Predictability is at the heart of accepting robot initiative in collaborative tasks (Marble *et al.*, 2004) and social regulation of joint activity among human and other autonomous agents (Feltovich *et al.*, 2007).

The third category is safety which refers "*to the beliefs that represent the risks of harm (performance, social, financial, physical)*". Safety is an obvious consideration in reliance on highly complex automated systems. Reliability, performance, operating characteristics, and error margins are all engineering concepts applicable to a veridical assessment of safety. However, safety becomes more difficult to judge as complexity increases, and in operational situations, humans tend to rely on heuristics in such cases (Lee & See, 2004). A meta-analysis of factors affecting trust in human-robot interaction showed that robot performance characteristics (e.g., failure rate) were found to be strongly associated with trust (Hancock et al., 2011).

However, qualities of the robot per se, such as manner of interaction, physical proximity, shape, "personality" and so forth, some of which (e.g., proximity) manifested relation to safety, could not be included in the meta-analysis due to an insufficient number of study samples (thus, indicating a need for further research).

Finally, speaking on the last category, openness; these are beliefs that represent the perceived ability to understand the potential trustee. In human interpersonal trust, this includes a judgment of the trusted agent's honesty, communication, and other behavioral and character attributes. With respect to automation, openness is often synonymous with "transparency," that is, the ability to understand what a device does and how it works. Madsen and Gregor (2000) explored the idea that the perceived understanding ability and the perceived technical competence of a system were key principal components in trust of automation.

2.7.5 Trust and Behavioral Cues

An important aspect of social behavior is the ability to convey and perceive trustworthiness. Trustworthiness of a trustee can be interpreted by using several sources and task performance could be the considered as major objective source for trust, because in general, previous behavior is a good predictor of future outcomes. Consequently, ideal trustworthiness is estimated from prior observation. In contrast, behavioral style can be defined as all the externally observable nonverbal aspects of the way a robot behaves while engaged with a task, such as the trustee's body language, facial expression, looking behavior or the way the trustee moves and acts.

Nonverbal behavioral cues of confidence or doubt can be easily displayed and changed, and this should affect the trustworthiness interpretation of a trustee. For instance, it has been shown that manners displayed during an interview, such as position and gestures, have a significant effect on the interviewer’s perceived trustworthiness (Hancock, Billings, Schaefer, Chen, Visser & Parasuraman, 2011). In the experiments presented by Van den Brule, Dotsch, Bijlstra, Wigboldus, and Haselager (2014a), it was confirmed that a social robot’s trustworthiness is chiefly influenced by its performance on a task. Features of a robot’s behavioral style (such as hesitations and gaze behavior), and motion fluency are noticed by participants. There is also preliminary evidence that these manipulations may lead to changes in the participants’ behavior, considering the exploratory nature of this analysis. Motion fluency was identified to affect a robot’s trustworthiness in a non-collaborative VHRI experiment. In the interactive setting, a robot’s task performance did seem to have a significant influence on participants’ monitoring behavior during the experiment.

2.7.6 Trust and Positive Experiences

To say that one has had a positive experience simply means that while having a sort of encounter or background about any object, it was very profitable or promising. The same goes for the success stories recorded in the usage of robots which automatically transfers to the ability to trust and believe in the capability of robots. In a press release by the International Federation of Robotics, it was stated that “The use of robots always guarantees fast return on investments and dramatic improvements in terms of quality. And this is true both in the car and the general industry, both in emerging countries and in nations having a long industrial tradition."

Robots have been seen to have a wide range of applications such as general-purpose autonomous robots, factory robots, military robots, healthcare care robots, research robots, entertainment robots amongst many others. Since 1998, a total of about 150,000 service robots for professional use have been reported in statistics. With about 9,500 units, service robots in defense applications accounted for almost 45 percent of the total number of service robots for professional use sold in 2013. Almost 5,100 milking robots were sold in 2013 compared to 4,750 units in 2012. 760 units of other robots for livestock farming such as mobile barn cleaners or robotic fencers for automated grazing control were sold in 2013. Other robots used for livestock farming as well as agricultural robots are getting grounded in the market.

The most important applications of medical robots are robot assisted with surgery and therapy with more than 1,000 units sold in 2013. Medical robots are the most valuable service robots therefore suppliers of medical robots also provide leasing contracts for their robots. Medical robots as well as logistic systems are well established service robots with a considerable growth potential. So far, service robots for personal and domestic use are mainly in the areas of domestic (household) robots, which include vacuum and floor cleaning, lawn-mowing robots, and entertainment and leisure robots, including toy robots, hobby systems, education and research. Handicap assistance robots have taken off to the anticipated degree in the past few years. In 2013, a total of about 700 robots were sold, up from 160 in 2012 which is an increase of 345 percent.

Numerous national research projects in many countries concentrate on this huge future market for robots and projections for the usage of robots within the next decade show very promising statistics. In addition, Takayama and Pantofaru (2009) have proved that human who have had previous experience with a robot normally tend to interact in close space with the robots and this denotes that the user have trust and predictability in the robotic systems behavior.

2.7.7 Trust Calibration Tools

Lee and See (2004) suggested that the capacities and confinements of the robotic frameworks be passed on to the operator, when achievable, in place for the operator to create proper trust and dependence. Bagheri and Jamieson (2004) exhibited that when operators were mindful of the connection related nature of robots unwavering quality, their rate of robots disappointments expanded altogether without influencing their simultaneous following and framework administration errands. The authors ascribed this change in execution to a more viable attentional designation system.

Rovira et al. (2007) explored the differential impact of robots unwavering quality and distinctive sorts of robots (decision support and information support) on execution amid summons and control undertaking. Their outcomes affirmed a differential expense of robots instability for three manifestations of choice robots, as contrasted and data robots when the general robots unwavering quality was at 80 percent. At 60 percent general robots dependability, in any case, there was a decrease in execution for both data and choice robots amid inconsistent trials. This discovery recommends that the kind of robots utilized is superfluous when robots reliability is beneath a certain limit.

Taking into account their discoveries on the diverse sorts of robots, Rovira *et al*. recommended that decrements in choice making execution will be lower when operators can make inquiry about the robots, review crude data sources, and check or invalidate the robotic guidance.

In another study, Seppelt and Lee (2007) found that drivers' dependence on the robots (e.g, voyage control) was more suitable when the presentation was available than when it was definitely not. Lee and See (2004) added to the accompanying rules on planning frameworks that advance fitting trust in the robots to include:

- Not more prominent trust; showing the past execution of the robots;
- Showing the procedure and calculations of the robots by uncovering Transitional results in a manner that is fathomable to the operators;
- Simplifying the calculations and operation of the robots to make it more justifiable;
- Showing the reason for the robots, outline premise, and scope of use in a manner that identifies with the client's objectives;
- Training operators in regards to its normal unwavering quality, the components representing its conduct, and its planned utilization;
- Carefully assessing any humanizing of the robots, for example, utilizing discourse to make an engineered conversational accomplice, to guarantee suitable trust.

CHAPTER THREE

RESEARCH METHODOLOGY

This chapter illustrates the research methodology adopted in this study. The methodology are presented in four phases of the design plan to examine the important features and attributes, obtained from psychological and cognitive grounded theories, that might be the baseline in establishing the formal model of dynamic trust in human-robot interaction. However, the phases are described and the sequence is shown in Figure 3.1

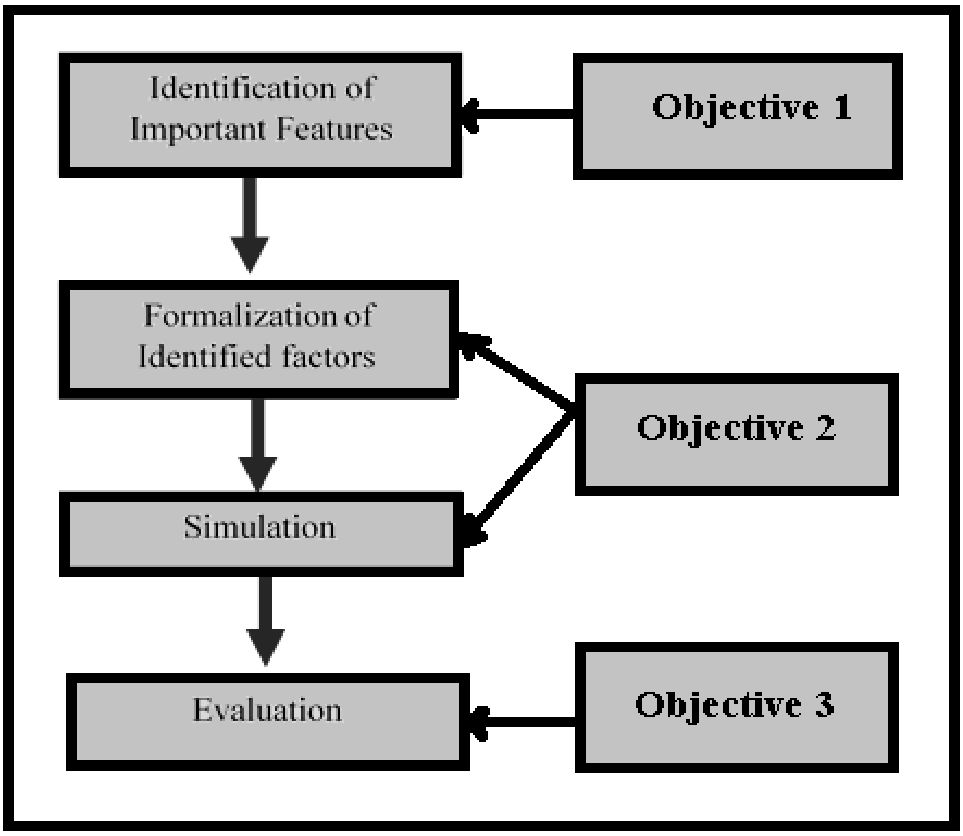

Figure 3.1: General Research Methodology Phases to Develop a Cognitive Model

The following sections give the description of the phases involved in the methodology:

3.1 Phase 1: Identification of Trust Factors in Human Robot Interaction

Dynamic properties of trust in human robot interaction are the group of psychological and cognitive factors that contribute directly or indirectly in developing dynamic trust based on the way of interaction between human and robot. Moreover, it gives insight to the fundamental mechanism of trust and normally these factors are in an informal representation, as well as, it called local dynamic factors. The outcome of this phase is a set of local dynamic properties, based on literature review, expert view, and empirical evidences, that will be as input to the next phase in constructing the formal cognitive model.

3.2 Phase 2: Formalization of Trust Factors in Human Robot Interaction

This phase involves the formulation of the local dynamics factors of trust into formal representations as well as aimed to be presented in form of executable dynamic properties to produce executable model that specify the dynamics of the process. It will involve logical (*First Order Logic*) and mathematical (*Differential Equations*) specifications of the various local dynamics features to create executable dynamics properties. Meanwhile, the formal model will be the outcome of this phase to represent the dynamic relationship between various features or attributes of trust in human-robot interaction.

3.3 Phase 3: Simulation

This is the phase where the formal model is simulated to generate simulation traces. The design and development of the simulated traces will be obtained using programming language and simulation tool based on the dynamic factors that have been previously formalized. In addition, it provides an insight in the sequence of events over time in specific instances of the process.

3.4 Phase 4: Evaluation

In this phase, the process of evaluation against simulation traces is specified. There are two mechanisms of evaluation that will be used to verify the validity of the cognitive model of trust which are internal and external validation. In the spectrum of internal validity, mathematical proofing (stability analysis) and automatic verification (temporal trace language or TTL) will be used. Stability analysis aims to verify the equilibrium points for the proposed model as indicators that the model has been developed accordingly whereas automated verification will be established to ensure the correctness of the model in explaining possible cases existed in the literature.

3.5 Summary

This section summarizes the processes that will be used in research methodology to accomplish the needed objectives. Table 3.1 shows the detailed processes involved to achieve this study.

Table 3.1: Summary of the Research Phases

Phases	Outcome	Methods	Research Objective
Phase One (Identifying Important Features)	Trust Factors in Human Robot Interaction	Literature Review	Objective One
Phase Two (Formalization)	Cognitive Model of Trust	Differential Equations	Objective Two
Phase Three (Simulation)	Simulated tTraces	Simulator development using numerical programming language	Objective Two
Phase Four (Evaluation)	Evaluated Model	Verification (Automated verification and mathematical analysis)	Objective Three

CHAPTER FOUR

MODEL DEVELOPMENT AND SIMULATION RESULT

4.1 Factors in the Model

Based on literature review, theories and empirical evidences, a set of eighteen important factors were identified. The identified factors were classified into three categories depending on their relations among the factors. The classification includes external, instantaneous and temporal relationship as shows in Table 4.1, 4.2 and 4.3. The factor relationship concepts in this phase were used to formalize the model while Figure 4.1 shows the model design concept.

Table 4.1: External Factor Relationship

No	External Factor	Formal Representation	Description
1	Personality	*Ps*	This is the characteristic traits of the user
2	Physical Appearance	*Pa*	This refers to the robot interface and human perception about the interface
3	Believable Behavior	*Ba*	The capability of the robot to display expertise and professionalism on a behavior
4	Behavioral Cues	*Bc*	The ability of the robot to trigger action, which means robot has the capability to behave actions in similar behavior with humans (verbal and non-verbal cues)
5	Level of Automation	*La*	Prompt and timely action on the behavior
6	Positive Experience	*Pe*	Previously good feeling or encounter with the robot

Table 4.2: Instantaneous Factor Relationship

No	Instantaneous Factor	Formal Representation	Description
1	Transparency	*Tr*	Having a clear understanding about the action or behavior of the robot.
2	Short Term Trust	*Sr*	Temporary confident in the relationship with the robot
3	Short Term Distrust	*Sd*	Temporary non-confident in the relationship with the robot.
4	Short Term Perceived Risk	*Sp*	Temporary foreseen threat in the relationship with the robot to achieve a particular task.
5	Competency	*Cy*	To be knowledgably about the action or behavior. Reducing error rate while performing a critical task.
6	Positive Deception	*Pd*	Good manipulative intention, the robot has the ability to convince people with the outcome in a positive manner.
7	Reliable Behavior	*Rb*	Confident in the positive outcome of the behavior. The robot will be reliable in implementing a task
8	Perception	*Pc*	Views, awareness and knowledge about the behavior of the robot

Table 4.3: Temporal Factor Relationship

No	Temporal Factor	Formal Representation	Description
1	Long Term Perceived Risk	*Lp*	Permanent foreseen threat in the relationship with the robot, accumulative effect of foreseeing risk yields from the robot's performance.
2	Long Term Trust	*Lr*	Permanent confident in the relationship with the robot
3	Long Term Distrust	*Ld*	Permanent non-confident in the relationship with the robot
4	Long Term Positive Experience	*Le*	Permanent good feeling or encounter with the robot

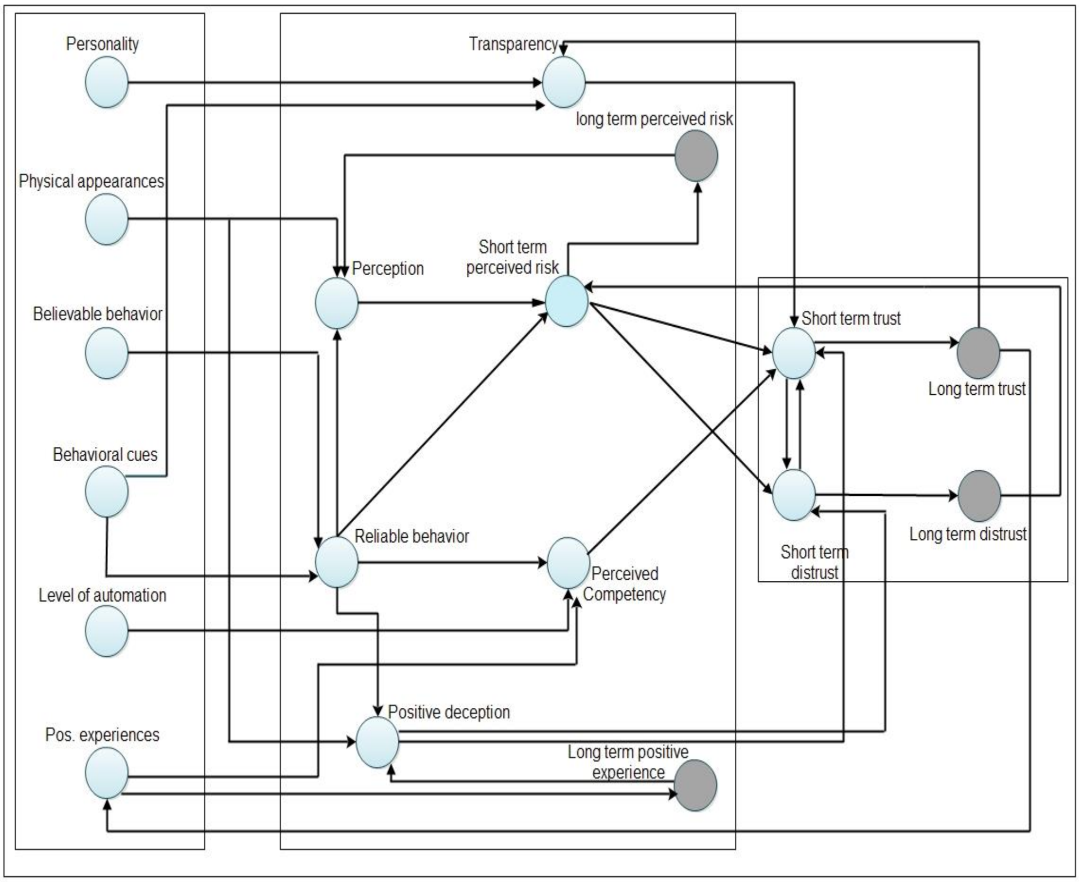

Figure 4.1: Conceptual Model of Trust in HRI

4.2 Formalization of the Model

From Phase one, the identified factors and underlying concepts were formulated into formal representations. The main aim of this phase is to obtain an executable model which was run in a simulation environment for further interpretation of the formal model. Differential equation technique was used to represent the formalization of identified factors and its relationships. Also, a set of parameters were used in order to regulate or control the obtained value from the equations ($0 \leq x \leq 1$)). The following subsections give details on the formalization of the model:

Perception (*Pc*)

This refers to views, awareness and knowledge about the behavior of robot. Figure 4.2 shows the relationship concept while Table 4.4 shows the conditions; Perception is high when Long-term perceived risk (*Lp*) is low and physical appearances (*Pa*) and reliable behaviors (*Rb*) are both high.

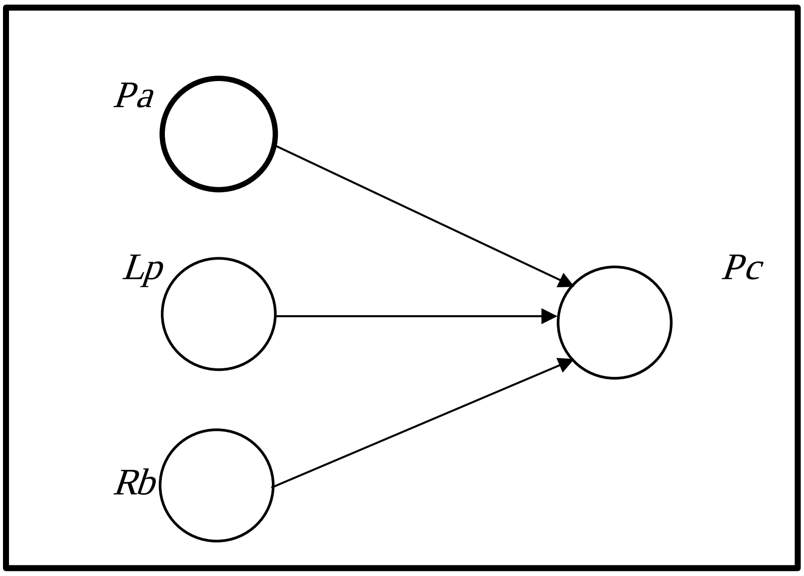

Figure 4.2: Relationships in Perception

Table 4.4: Different Condition of Perception

Condition	Factor's values	Perception value	Description
Condition 1	*Lp*=low *Pa*=high *Rb*=high	*Pc* = high	Perception is high when Long term perceived risk is low and Physical appearances and Reliable Behaviors are both high.
Condition 2	*Lp*=high *Pa*=high *Rb*=high	*Pc* = low	
Condition 3	*Lp*= low *Pa*= low *Rb*=low	*Pc* = low	

$$Pc(t) = [\alpha_{Pc} * Pa(t) + (1 - \alpha_{Pc}) * Rb(t)] * (1 - Lp(t)) \quad \textbf{(1)}$$

Reliable Behavior (*Rb*)

This refers to positive confident in outcome of the robot's behavior and Figure 4.3 show the concept while Table 4.5 shows the conditions. Reliable behavior is high when both believable behavior (*Ba*) and behavioral cues (*Bc*) are high. If one of (*Ba, Bc*) is low then *Rb* is low.

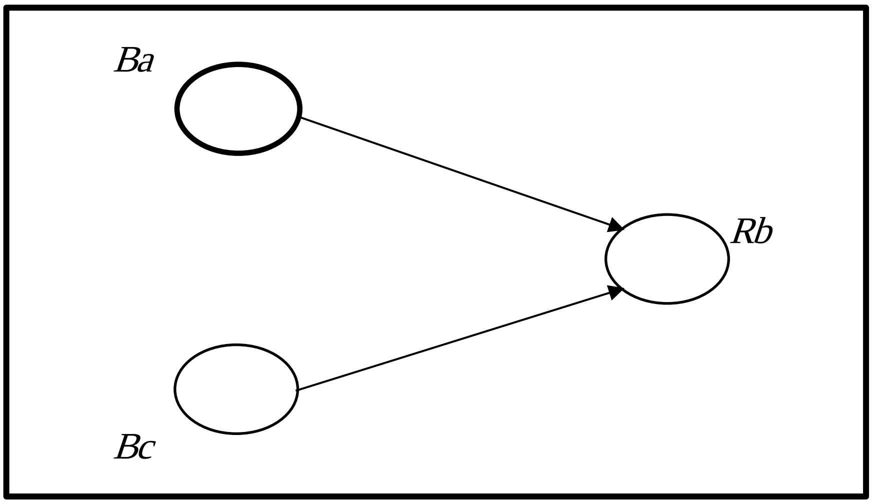

Figure 4.3: Relationships in RB

Table 4.5: Different Condition of Reliable Behavior

Condition	Factor's values	Perception value	Description
Condition 1	*Ba*=high *Bc*=high	*Rb* = high	Reliable behavior is high when both Believable behavior and behavioral cues are high. If one of (Ba, Bc) is low then Rb is low.
Condition 2	*Ba*= low *Bc*= low	*Rb* = low	

$Rb(t) = Ba(t) * Bc(t)$……………………………………………….…….**(2)**

Positive Deception (*Pd*)

This refers to the good manipulative intention to perform behavior and actions and Figure 4.4 shows the concept while Table 4.6 shows the conditions. Positive deception is high when long-term positive experience (*Le*), physical appearances (*Pa*) and reliable behavior (*Rb*) are high, otherwise, it will become low.

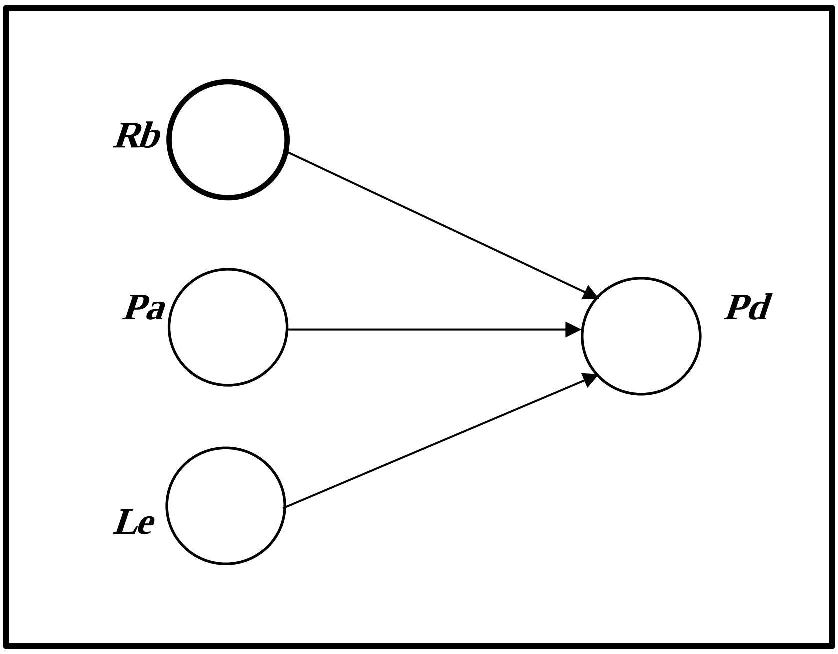

Figure 4.4: Relations in Positive deception

Table 4.6: Different Condition of Positive Deception

Condition	**Factor's values**	**Positive deception value**	**Description**
Condition 1	*Le*=low *Pa*=high *Rb*=high	Pd=low	Positive deception is high when long term positive experience, Physical appearances and Reliable behaviour are high. Otherwise, will become low.
Condition 2	*Le*=high *Pa*=high *Rb*=high	Pd=high	
Condition 3	*Le*=high *Pa*=low *Rb*=low	Pd = low	

$$Pd(\mathrm{t}) = [\lambda_{pd} * Rb(t) + (1-\lambda_{pd}) * Pa(t)] * Le(t) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(3)}$$

Competency (*Cy*)

This refers to knowledgeable ability about the robot's action or behavior. Both Figure 4.5 and Table 4.7 show the concept and conditions respectively. Competency is high when level of automation (*La*) is high and either both or one of (positive experiences (*Pe*), reliable behavior (*Rb*) is high.

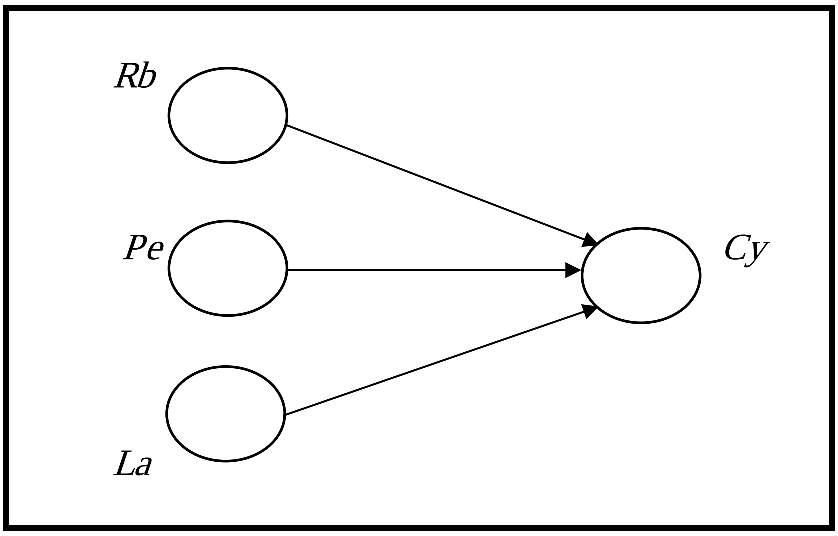

Figure 4.5:Relations in Competency

Table 4.7: Different Condition of Competency

Condition	Factor's values	Competency value	Description
Condition 1	*La*=low *Pe*=high *Rb*=high	*Cy*=low	Competency is high when level of automation is high and either both or one of (positive experiences, Reliable behavior) is high.
Condition 2	*La*=high *Pe*=high *Rb*=high	*Cy*=high	
Condition 3	*La*=high *Pe*=low *Rb*=low	*Cy* = low	

$$Cy(t) = [\varphi_{cy} * Rb(t) + (1-\varphi_{cy}) * (w_{c1} * Rb(t) + w_{c2} * Pe(t))] * La(t) \ldots\ldots \textbf{(4)}$$

Where $\sum Wc_{n=1}$. $(w_{c1} + w_{c2} = 1)$

Short Term Perceived Risk (*Sp*)

This refers to the temporary foreseen threat in the relationship on the behavior of robot. Figure 4.6 and Table 4.8 show the concept and conditions respectively. Short term perceived risk is high when long-term distrust (*Ld*) is high and both perception (*Pc*) and reliable behavior (*Rb*) are low.

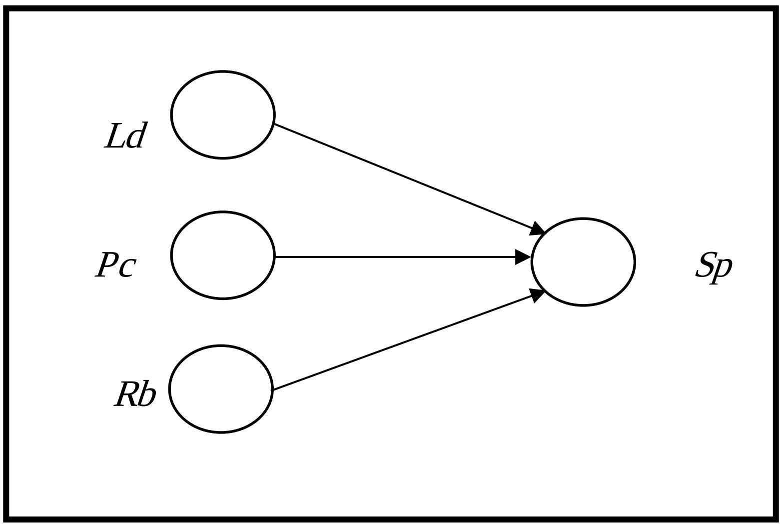

Figure 4.6: Concepts in Short- Term Perceived Risk

Table 4.8: Concept and Conditions

Condition	Factor's values	Short term perceived risk value	Description
Condition 1	*Ld*=high *Pc*=low *Rb*=low	*Sp*=high	Short term perceived risk is high when long term distrust is high and both Perception and Reliable behaviour are low.
Condition 2	*Ld*=high *Pc*=high *Rb*=high	*Sp*=low	
Condition 3	*Ld*=low *Pc*=high *Rb*=high	*Sp* = low	

$$Sp(t) = Ld(t) * [(1-Pc(t)) * (1-Rb(t))] \quad \textbf{(5)}$$

Short Term Distrust (*Sd*)

This refers to the temporary non-confident in behaviors of robot and Figure 4.6 and Table 4.8 show concept and condition respectively. Short term perceived risk is high when long term distrust is high and both Perception and Reliable behavior are low.

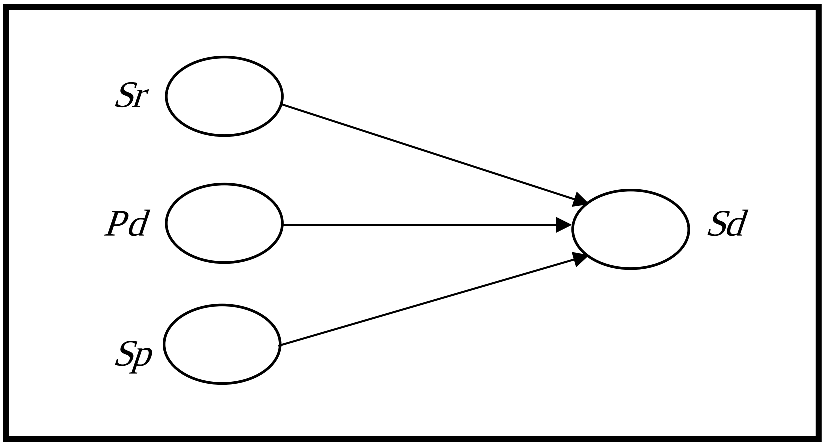

Figure 4.7: Short Term Distrust Concept

Table 4.9: Different Condition of Short term Distrust

Condition	**Factor's values**	**Short term distrust value**	**Description**
Condition 1	*Pd*=low *Sr*=low *Sp*=high	*Sd*=high	Short term distrust is high when positive deception and Short term trust are low.
Condition 2	*Pd*=high *Sr*=high *Sp*=low	*Sd*=low	

$$Sd(t) = [\beta_{sd} * Sp(t) + (1- \beta_{sd}) * (1-Pd(t))] * (1-Sr(t)) \ldots\ldots\ldots\ldots\ldots \textbf{(6)}$$

Short Term Trust (*Sr*)

This refers to temporary confident in the outcome of robot's behavior or implementation and Figure 4.7 and Table 4.9 show the concept and conditions respectively. Short term trust is high when positive deception (*Pd*) is high and competency of robot (*Cy*) and its transparency (*Tr*) are high also short-term distrust (*Sd*) must be (low).

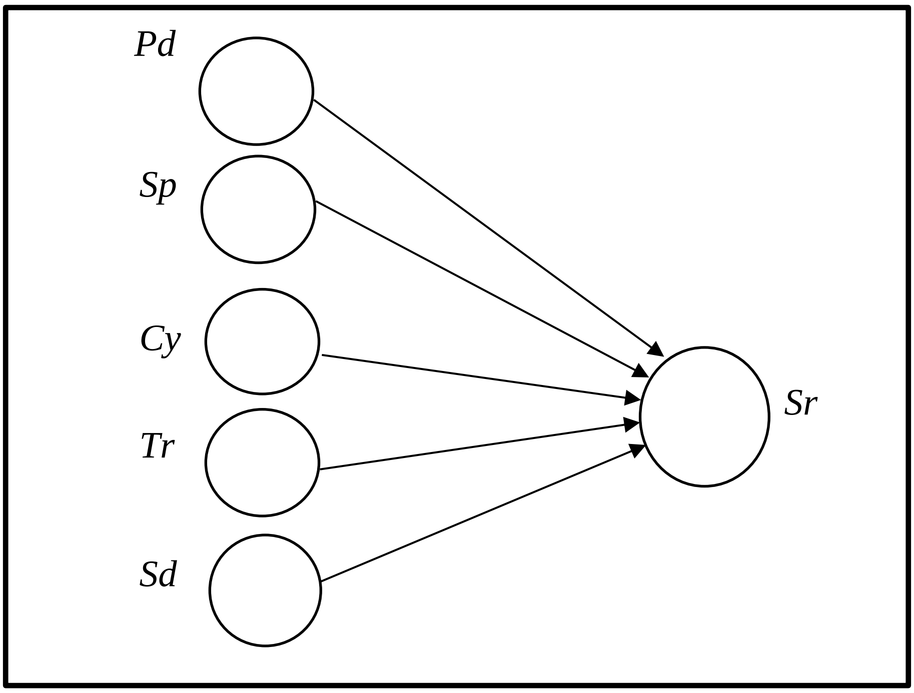

Figure 4.8: Concepts in Short Term Trust

Table 4.10: Different Condition of Short term Trust

Condition	Factor's values	Short term perceived risk value	Description
Condition 1	*Tr*=low *Cy*=low *Pd*=low *Sp*=high *Sd*=high	*Sr=low*	Short term trust is high when positive deception is high and competency of robot and it's transparency are high also short term distrust must be (low).
Condition 2	*Tr*=low *Cy*=low *Pd*=low *Sp*=low *Sd*=high	*Sr=high*	
Condition 3	*Tr*=high *Cy*=low *Pd*=high *Sp*=low *Sd*=high	*Sr = high*	
Condition 4	*Tr*=high *Cy*=low *Pd*=high *Sp*=low *Sd*=low	*Sr = high*	

$Sr(t) = [\beta_{sr} * (\alpha * Tr(t) + (1-\alpha) * Cy(t)] + (1-\beta_{sr}) * [Pd(t) * (1-Sd(t)) * (1-Sp(t))]$……..***(7)***

Transparency (*Tr*)

This implies having a clear understanding about the action or behavior of robot and Figure 4.8 and Table 4.10 show the concept and conditions respectively. Transparency is high when long term trust is high and Behavioral cues also high.

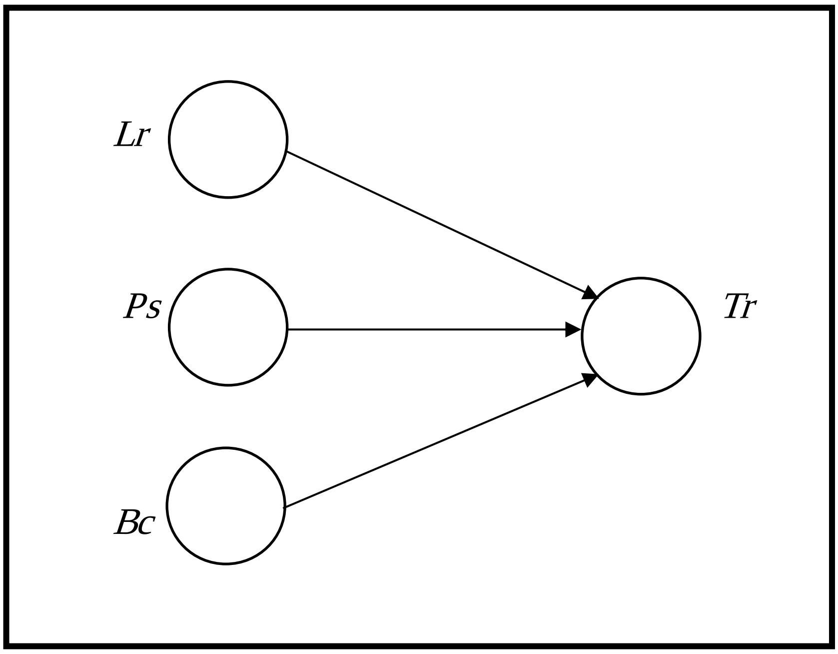

Figure 4.9: Transparency Concept

Table 4.11: Different Condition of Transparency

Condition	**Factor's values**	**Transparency value**	**Description**
Condition 1	*Lr*=high *Ps*=high *Bc*=high	*Tr*=high	Transparency is high when long term trust is high and Behavioural cues also high.
Condition 2	*Lr*=low *Ps*=low *Bc*=high	*Tr*=low	

$$Tr(t) = [\ w_{r1} * Lr(t) + w_{r2} * Ps(t) + w_{r3} * Bc(t)\ ] \ldots\ldots\ldots\ \ldots\ldots\ldots\ldots\ldots\ldots \textbf{\textit{(8)}}$$

Positive experience (*Pe*)

This implies good feeling or encounter on behavior of robot and Figure 4.9 and Table 4.11 show the concept and condition respectively. Positive experience is high when long-term trust (*Lr*) is high. Moreover, Pe_{norm} is the initial or the existing experiences with an individual.

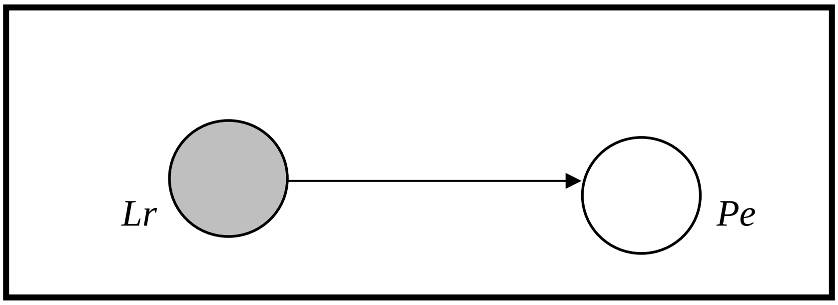

Figure 4.10: Concepts in Positive Experience

Table 4.12: Different Condition of Positive Experience

Condition	**Factor's values**	**Positive experience value**	**Description**
Condition	*Lr*=high	*Pe* =high	Positive experience is high when long term trust is high.

$$Pe(t) = \beta_{Pe\ norm} + (1-\beta_{Pe\ norm}) * Lr(t) * (1-Pe_{norm}) \quad \textbf{(9)}$$

Long Term Perceived Risk (*Lp*)

This implies permanent foreseen threat in the behavior or a relationship of robotic system and Figure 10 and Table 12 show the concept and condition respectively. Long term perceived risk is high when short term perceived risk is high and vice versa. *Lp* Increases or decreases over time depending on *Sp*. If there is considerable presence of *Sp*, the amount of *Lp* will increase. Concept in parameter and λLp is used to determine the changing rate of temporal relationship. Moreover, changing process is measured in a time interval between (*t*) and ($t+\Delta t$).

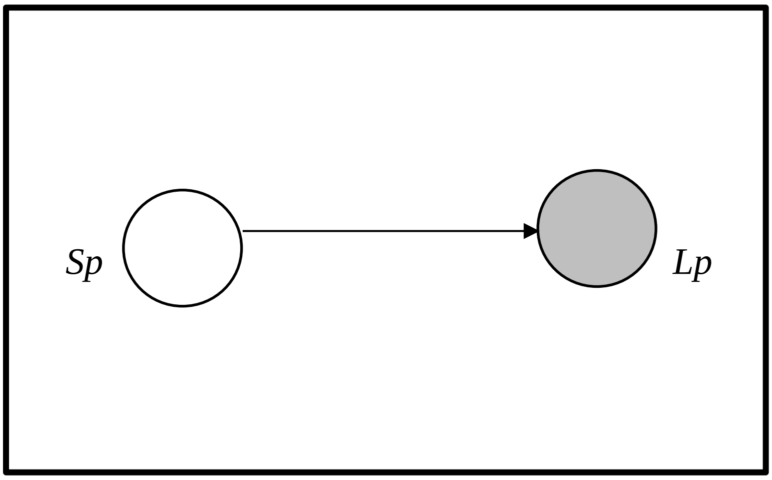

Figure 4.11: Long Term Perceived Risk (Lp)

Table 4.13: Different Condition of Long Term Perceived Risk (Lt)

Condition	**Factor's values**	**Perception value**	**Description**
Condition	*Sp*=high	*Lp* = high	Long term perceived risk is high when short term perceived risk is high and vice versa.

$$Lp(t+\Delta t) = Lp(t) + \lambda_{lp} * Lp(t) * [Sp(t) - Lp(t)] * (1-Lp(t)) * \Delta t \ldots \textbf{(10)}$$

Long Term Trust (*Lr*)

This implies permanent confident in the behaviour or relationship of robot. Figure 4.11 and Table 4.13 show the concept and condition respectively. Long term trust is high when short term trust is high and vice versa. *Lr* Increases or decreases over time depending on *Sr*. If there is considerable presence of *Sr*, the amount of *Lr* will increase. Parameter φ_{lr} is used to control the changing rate of temporal relationship. Moreover, changing process is measured in a time interval between (*t*) and (*t*+Δ*t*).

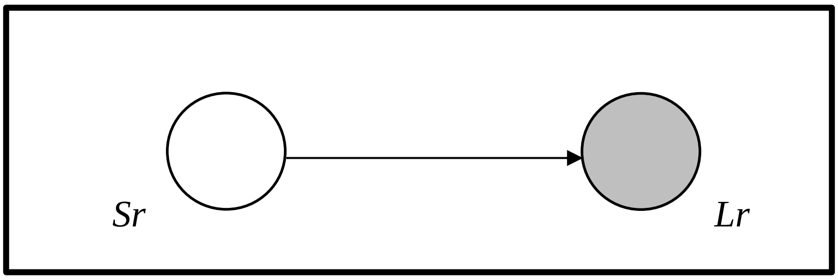

Figure 4.12: Long-termTrust Concept

Table 4.14: Different Conditions of Long term Trust

Condition	Factor's values	Long term trust value	Description
Condition	*Sp*=high	*Lt* = high	Long term trust is high when short term trust is high and vice versa.

$$Lr(t+\Delta t) = Lr(t) + \varphi_{lr} * Lr(t) * [Sr(t) - Lr(t)] * (1-Lr(t)) * \Delta t \ldots\ldots \mathbf{(11)}$$

Long Term Distrust (*Ld*)

This implies permanent non-confident in the behaviour or relationship of machine. Figure 4.12 and Table 4.14 show the concept and condition respectively. Long term distrust is high when short term distrust is high and vice versa. *Ld* increases or decreases over time depending on *Sd*. If there is considerable presence of *Sd*, the amount of *Ld* will increase. Parameter ψ_{Ld} is used to control the changing rate of temporal relationship. Moreover, changing process is measured in a time interval between (t) and ($t+\Delta t$).

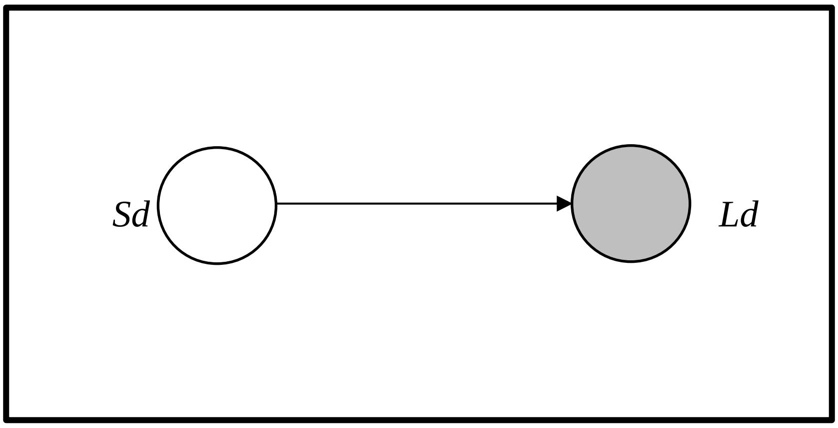

Figure 4.13: Long term Distrust Concept

Table 4.15: Different Condition of Long term Distrust

Condition	**Factor's values**	**Long term distrust value**	**Description**
Condition	*Sd*=high	*Ld* = high	Long term distrust is high when short term distrust is high and vice versa.

$$Ld(t+\Delta t) = Ld(t) + \psi_{ld} * Ld(t) * [Sd(t) - Ld(t)] * (1 - Ld(t)) * \Delta t \quad \ldots\ldots\ldots\ldots \mathbf{(12)}$$

Long Term Positive Experience (*Le*)

This implies permanent good feeling or encounter with the behavior or relationship of robot and Figure 13 and Table 15 show the concept and condition respectively. Long term positive experience is high when positive experience (*Pe*) is high and vice versa. *Le* Increases or decreases over time depending on *Pe*. If there is considerable presence of *Pe*, the amount of *Le* will increase. Parameter α_{Le} is used to determine the changing rate of temporal relationship. Moreover, changing process is measured in a time interval between (t) and ($t+\Delta t$).

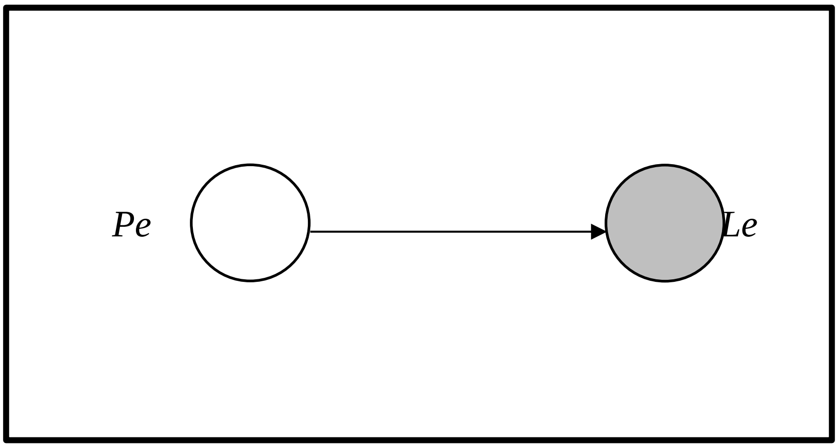

Figure 4.14: Long Term Positive Experience Concept

Table 4.16: Different Conditions of Long term Positive Experience

Condition	**Factor's values**	**Long term positive experience value**	**Description**
Condition	*Pe* =high	*Le* = high	Long term positive experience is high when positive experience is high and vice versa.

$$Le(t+\Delta t) = Le(t) + \alpha_{le} * [Pe(t) - Le(t)] * (1-Le(t)) * \Delta t \quad \text{……….……..}\mathbf{(13)}$$

4.3 Simulation Results

In this phase, the formal model was simulated in order to generate simulation traces. Moreover, simulating the computational model provides an insight in the sequences of events over time in specific instances of the process. During this phase, a simulator has designed and developed using Matlab programming language. This code was divided into four sections. First part, several parameters were initialized which used to regulate the equations. Second part, declared all factors that used in the model and initialized with specific value in order to pass these values to the third part which involved all equations and its operations. After giving initial values to all instantaneous and temporal factors, next was implemented by plotting the values to generate simulation traces. The simulation results will be discussed in following Subsections 4.3.1, 4.3.2, and 4.3.3.

4.3.1 Case #1 (High Level of Trust in HRI)

In this case, normally people initiate high level of trust during human robot interaction. The key reason of initiate high level of trust is the high values of the external factors in the formal model. However, it means that humans' characteristics are positive and also the robot has a good appearances, Believable cues, reliable behavior and high level of automation. Moreover, humans have a good experience in dealing with robot. Figure 4.15 shows the simulation results with values provided in Table 4.17.

Table 4.17: Values of Case #1

Factor	Given value
Personality (*Ps*)	0.8
Physical appearances (*Pa*)	1
Believable behavior (*Ba*)	0.9
Believable cues (*Bc*)	0.8
Level of automation (*La*)	0.9
Pos. experiences (*Pe*)	0.9
(Pe_{norm})	1

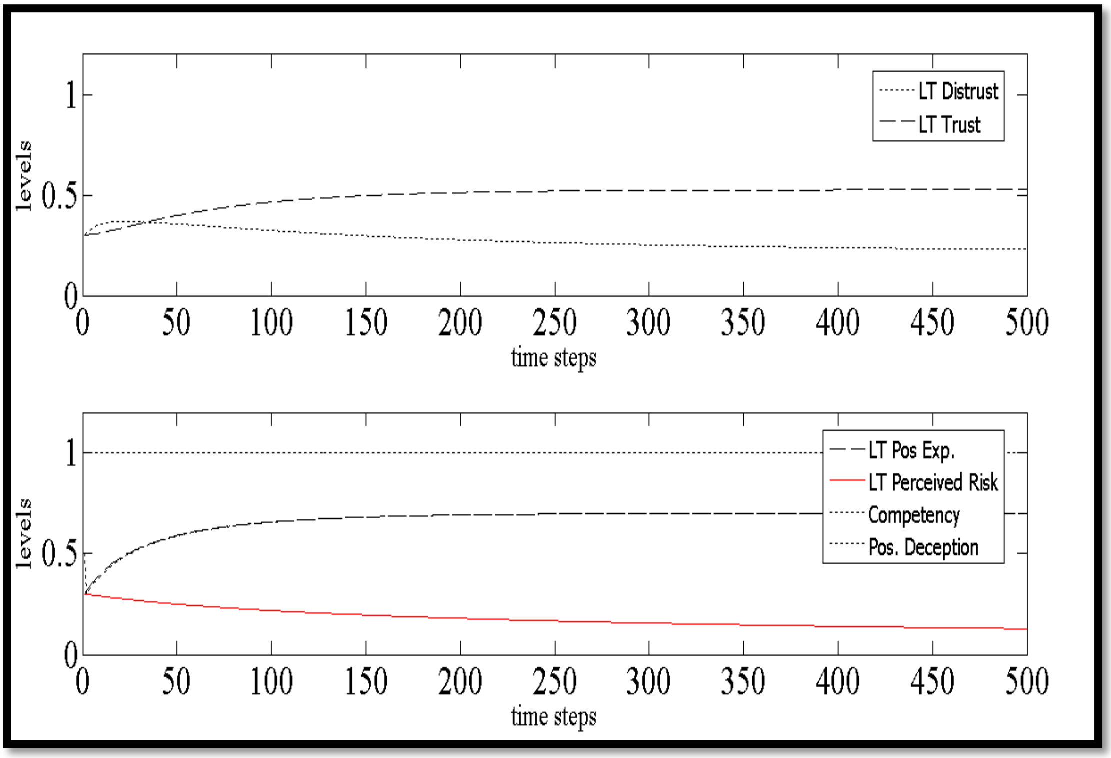

Figure 4.15 Simulation Result for Case #1

4.3.2 Case #2 (Moderate High Level of Trust in HRI)

In this case, humans are tending to evaluate some of the external factors relatively less than the values given in high level of trust during human robot interaction. The critical reason of initiate moderate high level of trust is the moderate high values of the external factors in the formal model. In addition, it means that humans' personality are positive toward a robot and also the robot has a good physical embodiment, Believable cues, reliable behavior and moderate level of automation. Moreover, humans have a good experience in dealing with robot. Figure 4.16 shows the simulation results with values provided in Table 4.18.

Table 4.18: Values of Case #2

Factor	Given value
Personality (*Ps*)	0.7
Physical appearances (*Pa*)	1
Believable behavior (*Ba*)	0.8
Believable cues (*Bc*)	0.8
Level of automation (*La*)	0.8
Pos. experiences (*Pe*)	0.9
(Pe_{norm})	1

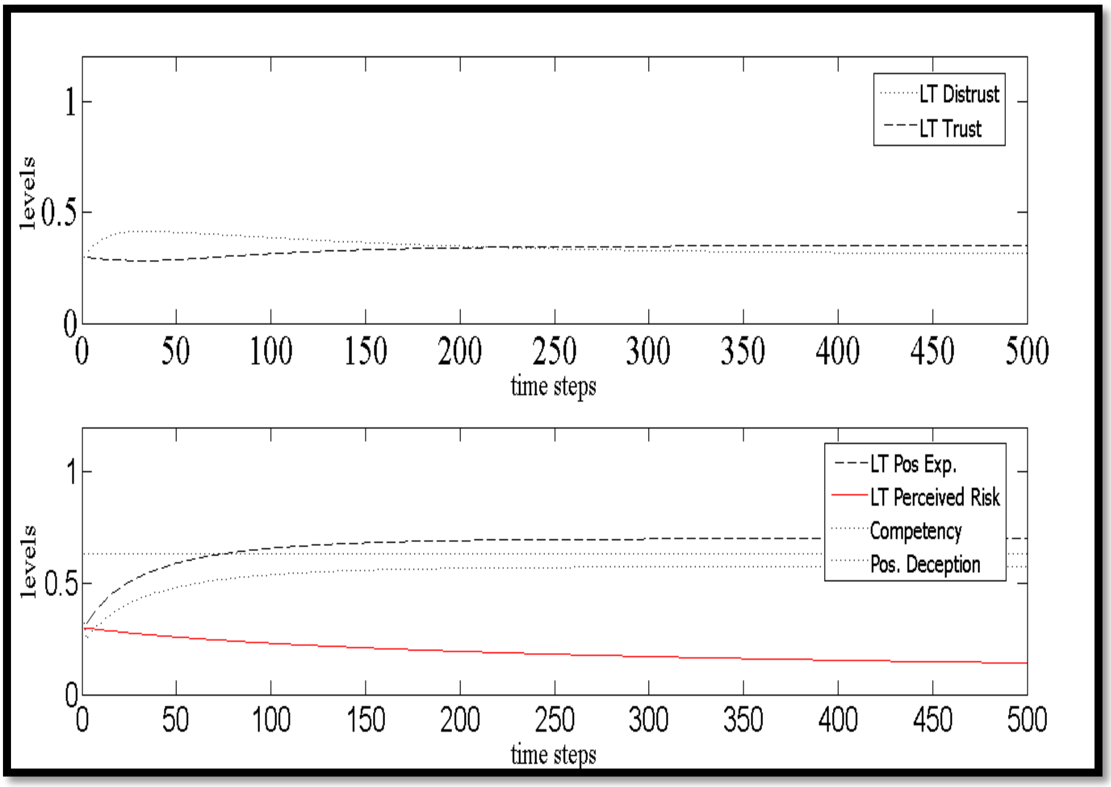

Figure 4.15: Simulation Result of Case #2

4.3.3 Case #3 (Moderate level of Trust in HRI)

In this case, the given values were altered to determine the moderate level of trust during human robot interaction. It means all external factors are still good but some of them have less power in initiating trustworthy interaction. However, Table 4.19 showed the given values in simulating moderate level of trust in HRI.

Table 4.19: Values of Case #3

Factor	Given value
Personality (*Ps*)	0.7
Physical appearances (*Pa*)	1
Believable behavior (*Ba*)	0.7
Believable cues (*Bc*)	1
Level of automation (*La*)	0.7
Pos. experiences (*Pe*)	0.7
(Pe_{norm})	1

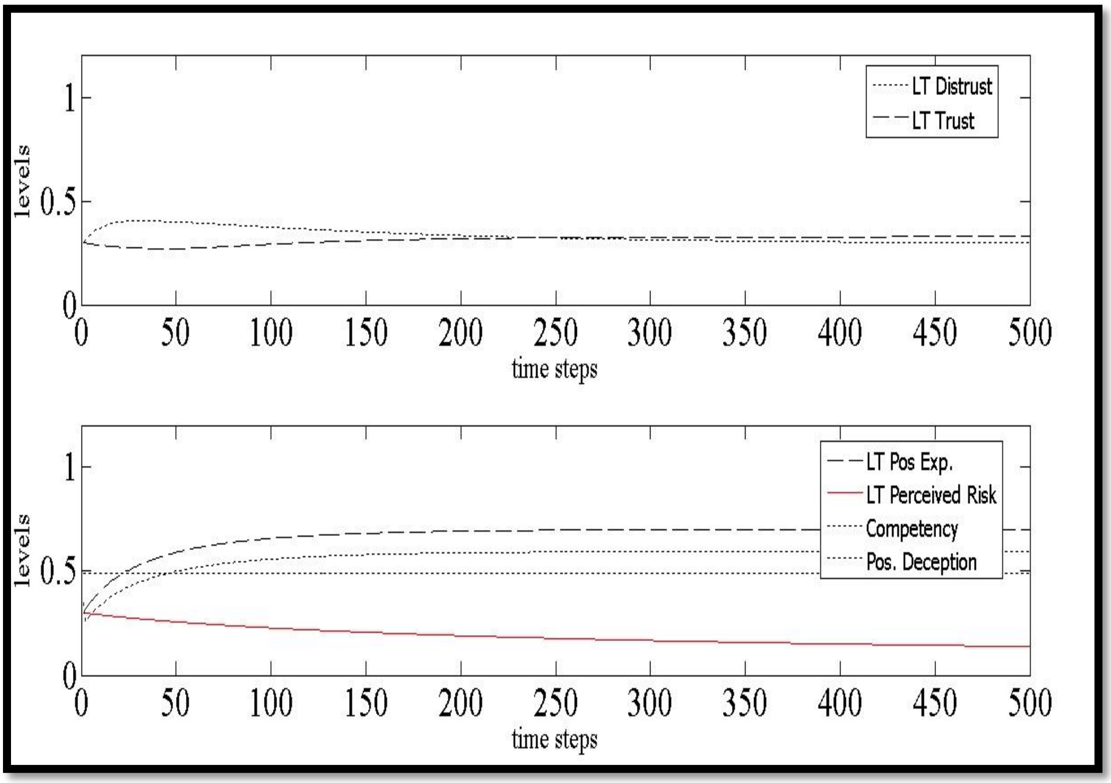

Figure 4.16: Simulation Result of Case #3

4.3.4 Case #4 (Moderate Low level of Trust in HRI)

This case showed the moderate poor trust and its influence in the interaction between human and robot. All the external factors have negative effects (low values) to initiate trust. Normally, human are less willingness to rely upon robot (negative traits) to achieve a particular task as well as robot characteristics has less influencing (poor appearances, unconvincing performance for the mission, simple or has no social cues). Table 4.20 shows all the given values that answer why human are not developing trust in their robot. Figure 4.18 is showing the simulation result of low trust in HRI.

Table 4.20: Values of Case #4

Factor	Given value
Personality (*Ps*)	0.2
Physical appearances (*Pa*)	0.6
Believable behavior (*Ba*)	0.3
Believable cues (*Bc*)	0.4
Level of automation (*La*)	0.3
Pos. experiences (*Pe*)	0.1
(Pe_{norm})	0.1

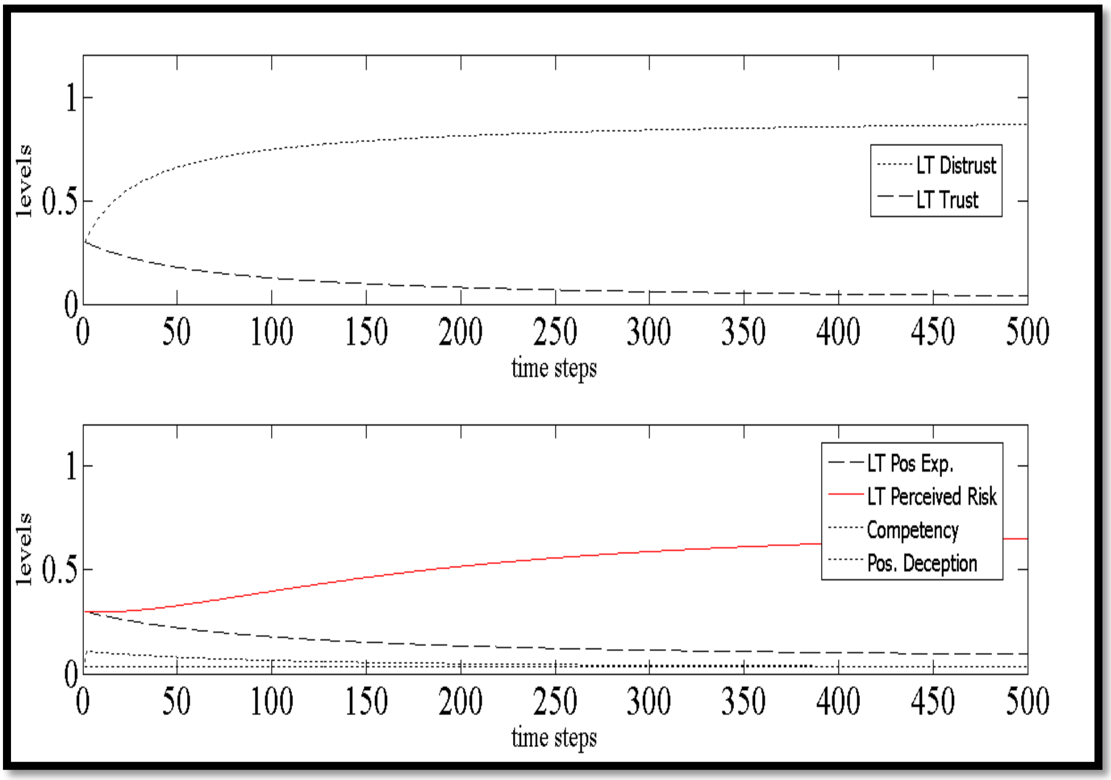

Figure 4.18: Simulation Result of Case #4

4.3.5 Case #5 (Low level of Trust in HRI)

This case showed the poor trust between human and robot interaction. All the external factors have negative effects to initiate trust (zero for each external factor). Normally, human are less willingness to rely upon robot (negative personalities) in order to accomplish a particular role as well as robot characteristics has poor impact (poor appearances, not reliable in performing the task, has no social cues). Table 4.21 shows all the given values that answer why human are not developing trust in their robot. Figure 4.19 is showing the simulation result of low trust in HRI.

Table 4.21: Values of Case #5

Factor	Given value
Personality (*Ps*)	0
Physical appearances (*Pa*)	0
Believable behavior (*Ba*)	0
Believable cues (*Bc*)	0
Level of automation (*La*)	0
Pos. experiences (*Pe*)	0
(Pe_{norm})	0

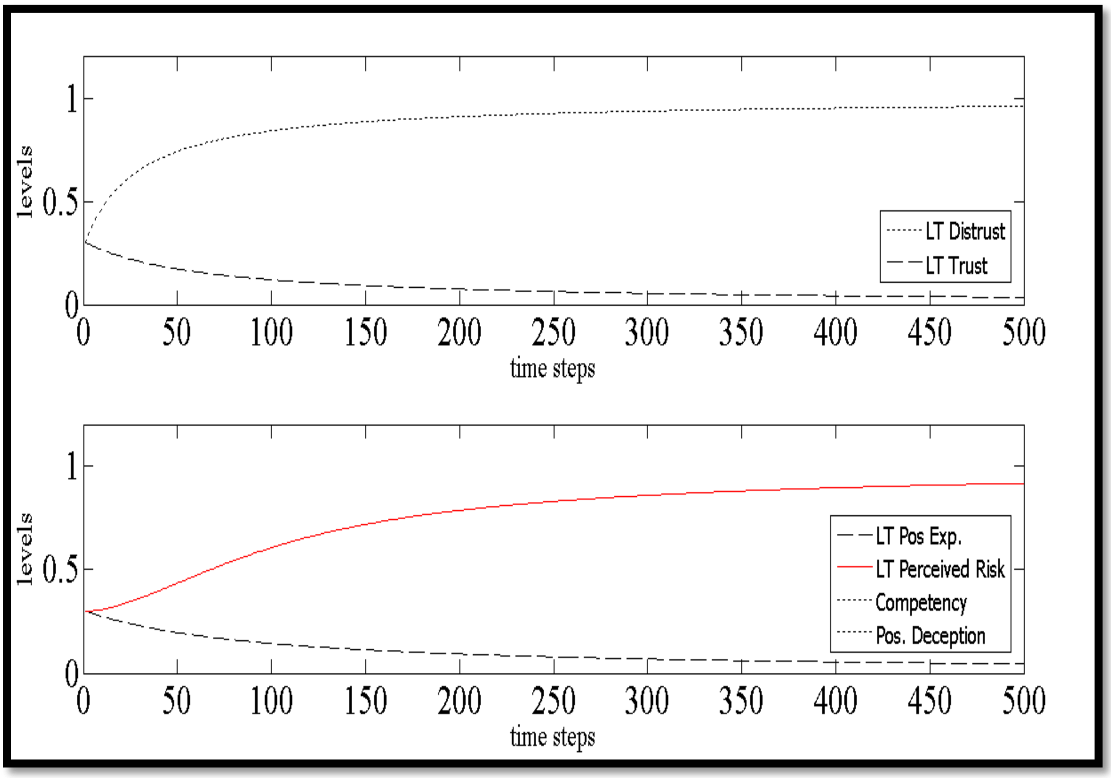

Figure 4.19: Simulation Result of Case #5

CHAPTER FIVE
EVALUATION

In this phase, the process of evaluation against simulation traces was specified. There were two mechanisms of evaluation used to verify the validity of the model of trust which were verification and validation.

5.1 Verification

In the spectrum of internal validity, stability analysis (mathematical) and automatic analysis (temporal trace language) were used. Stability analysis aimed to verify the equilibrium points for the model which indicated that the model has been developed accordingly. Whereas automated verification established that the correctness of the model is ensured.

5.1.1 Mathematical Analysis

In this section the equilibria are analyzed that may occur under certain conditions. The equilibria describe situations in which a stable situation has been reached. Those equilibria are interesting as it should be possible to explain them using the knowledge of the domain that is modelled. As such, the existence of reasonable equilibria is an indication for the correctness of the model. One important note that an equillibria condition(s) is considered stable if the system always returns to it after small disturbances. These equillibria conditions are interesting to be explored, as it is possible to explain them using the knowledge from the theory or problem that is modelled. As such, the existence of reasonable equilibria is also an indication for the correctness of the model. To analyze the equilibria, the available temporal and instantaneous equations are filled with values for the model variables such that the derivatives or differences between time point t and $t + \Delta t$ are all 0 (in particular for perceived risk, trust, distrust and positive experience). For example, using this autonomous equation,

$$dy/dx = f(y)$$

The equilibria or constant solutions of this differential equation are the roots of the equation

$$f(y) = 0$$

Then in all of the equations the reference to time t can be left out, and in addition the differential equations can be simplified by cancelling, for example, $Lp(t+\Delta t)$ against $Lp(t)$. This leads to the following equations.

$Pc = [\alpha_{pc} * Pa + (1-\alpha_{pc}) * Rb)] * (1-Lp)$	**(14)**
$Rb = Ba * Bc$	**(15)**
$Pd = [\lambda_{Pd} * Rb + (1-\lambda_{pd}) * Pa] * Le$	**(16)**
$Cy = [\varphi_{cy} * Rb + (1-\varphi_{cy}) * (w_{c1} * Rb + w_{c2} * Pe)] * La$	**(17)**
$Sp = Ld * [(1-Pc) * (1-Rb))]$	**(18)**
$Sd = [\beta_{sd} * Sp + (1-\beta_{sd}) * (1-Pd)] * (1-Sr)$	**(19)**
$Sr = [\beta_{sr} * (\alpha * Tr + (1-\alpha) * Cy] + (1-\beta_{sr}) * [Pd * (1-Sd) * (1-Sp)]$	**(20)**
$Tr = [w_{r1} * Lr + w_{r2} * Ps + w_{r3} * Bc]$	**(21)**
$Pe = \beta_p . Pe_{norm} + (1-\beta_p) * Lr * (1-Pe_{norm})$	**(22)**
$\lambda_{Lp} * Lp * [Sp - Lp] * (1-Lp) = 0$	**(23)**
$\varphi_{Lr} * Lr * [Sr - Lr] * (1-Lr) = 0$	**(24)**
$\psi_{Ld} * Ld * [Sd - Ld] * (1-Ld) = 0$	**(25)**
$\alpha_{Le} * Le * [Pe - Le] * (1-Le) = 0$	**(26)**

Assuming the parameters λ_{Lp}, φ_{Lr}, ψ_{Ld}, α_{Le} nonzero, from the equations (14) to (26) , for any possible simulation results, the following cases can be distinguished:

$$\begin{array}{lll} Lp = 0, & Sp = Lp, & Lp = 1 \\ Lr = 0, & Sr = Lr, & Lr = 1 \\ Ld = 0, & Sd = Ld, & Ld = 1 \\ Le = 0, & Pe = Le, & Le = 1 \end{array}$$

These possible conditions will result up to $3^4 = 81$ possible equilibria. Also given the other equations with a large number of input variables, and the number of possible cases involved, this makes it hard to come up with a complete classification of equilibria. However, for some typical cases the analysis can be pursued further.

Case #1: *Lp* =0

For this case, by equation (14), it follows that this is equivalent to

$$Pc = [\alpha_{pc} * Pa + (1-\alpha_{pc}) * Rb)]$$

Assuming α_{pc}=0, therefore $Pc = Rb$

Case #2: *Sp = Lp*

From equation (19), it follows that

$$Sd = [\beta_{sd} * Lp + (1-\beta_{sd}) * (1-Pd)] * (1-Sr)$$

This can only occur in the following subcases;

$$[\beta_{sd} * Lp + (1-\beta_{sd}) * (1-Pd)] = 0 \text{ or } Sr=1$$

Assuming $0 < \beta_{sd} < 1$, this is equivalent to:

$$Lp = -[(1-\beta_{sd}) * (1-Pd)] / \beta_{sd}$$

By equation (20), it follows that;

$$Sr = [\beta_{sr} * (\alpha * Tr + (1-\alpha) * Cy] + (1-\beta_{sr}) * [Pd * (1-Sd) * (1-Lp)]$$

Moreover, it can be described as;

$$[\beta_{sr} * (\alpha * Tr + (1-\alpha) * Cy] + (1-\beta_{sr}) * [Pd * (1-Sd) * (1-Lp)] = 0$$

which later can be analyzed as;

$$[\beta_{sr} * (\alpha * Tr + (1-\alpha) *Cy] = -(1-\beta_{sr}) * [Pd * (1-Sd) * (1-Lp)]$$

Assuming $0 < \beta_{sr} < 1$, and $0 < \alpha < 1$, this is equivalent to:

$$Pd = [\beta_{sr} * (\alpha * Tr + (1-\alpha) *Cy] * [(1-Sd) * (1-Lp)]/(1-\beta_{sr})$$

From equation (14), and assuming $\alpha_{pc} \neq 0$, this is equivalent to

$$Pc = [\alpha_{pc} * Pa + (1-\alpha_{pc}) * Rb)] * (1-Sp))$$

Case #3: *Pe = Le*

For this case, by equation (17), it follows that

$$Cy = [\varphi_{cy} * Rb + (1-\varphi_{cy}) * (w_{c1} * Rb + w_{c2} * Le)] * La$$

Next is the verification result from the automated logical verification process.

5.1.2 Automated Logical Verification

In order to verify whether the model indeed generates results that adherence to psychological literatures, a set of properties have been identified from related literatures. Therefore, these properties will answer whether the model produces results that are coherent with the literature and appropriate to help people with cognitive vulnerability and coping problems. To allow the verification process to take place, these properties have been specified in a language called Temporal Trace Language (TTL).

TTL is built on atoms referring to states of the world, time points, and traces. This relationship can be presented as a *state* $(\gamma, t, output(R))|=p$, means that state property *p* is true at the output of role *R* in the state of trace γ at time point *t*. It is also comparable to the *Holds*-predicate in the Situation Calculus. Based on that concept, dynamic properties can be formulated using a sorted predicate logic approach, by manipulating quantifiers over time and traces and first-order logical connectives such as ¬, ∧, ∨, ⇒, ∀, and ∃. A number of simulations including the ones described in previous sub-section have been used as basis for the verification and were confirmed.

VP1: Physical Appearances Will Improve Trust in Human Robot Interaction

If the robot has physical appearances that can be understood by human, it will increase the level of trust during human-robot interaction (Bainbridge, Hart, Kim, & Scassellati, 2008).

VP1 ≡ ∀γ: TRACE, ∀t1, t2:TIME, ∀R1,R2,D1,D2:REAL
[state(γ,t1)|= has_value(physical_appearances, R1) &
state(γ,t2)|= has_value(physical_appearances, R2) &
state(γ,t1)|= has_value(long_term_trust, D1) &
state(γ,t2)|= has_value(long_term_trust, D2) &
t1 < t2 & R2 > R1] ⇒ D2 ≥ D1

VP2: High-risk Situation Reduces Trust

Humans normally intend to less trust in robot when the situation is risky which means less trust will be developed in human robot interaction (Robinette, Wagner, & Howard, 2013).

VP2 ≡ ∀γ: TRACE, ∀t1, t2:TIME, ∀G1,G2,H1,H2,J1,J2:REAL
[state(γ,t1)|= has_value(level_automation, G1) &
state(γ,t2)|= has_value(level_automation, G2) &
state(γ,t1)|= has_value(believable_cues, H1) &
state(γ,t2)|= has_value(believable_cues, H2) &
state(γ,t1)|= has_value(long_term_trust, J1) &
state(γ,t2)|= has_value(long_term_trust, J2) &
t1 < t2 & G1 > G2 & H1 > H2] ⇒ J1 ≥ J2

VP3: Highly Competent Autonomous Robot Improves Trust

When artificial intelligent such robots have high competency in performing a particular tasks, people intend to interact and develop a trust in their interaction (Flanagin).

VP3≡ ∀γ:TRACE, ∀t1, t2:TIME ∀M1, M2, D:REAL [state(γ, t1)|= has_value(level_automation, M1) &

state(γ, t2)|= has_value(long_term_trust, M2) &

M1 $\geq$ 0.5 & t2= t1+D] $\Rightarrow$ M2 $\geq$ 0.5

VP4: Individuals with Positive Personality Tend to Trust Autonomous Robots

Human with positive personality like openness, agreeableness, and extroversion are likely to trust the autonomous robots (Tapus, 2014).

VP4≡ $\forall\gamma$:TRACE, $\forall$t1, t2:TIME $\forall$Q1, Q2, D:REAL

[state(γ, t1)|= has_value(personality, Q1) &

state(γ, t2)|= has_value(long_term_trust, Q2) &

Q1 $\geq$ 0.8 &

t2= t1+D] $\Rightarrow$ Q2 $\geq$ 0.5

VP5: Verbal and non-Verbal Cues Play Important Roles to Regulate Trust

Normally, in human robot interaction, humans are developing a certain trust with a sociable robot. Sociable cues (verbal and non-verbal) are playing major role in developing trust in HRI (Yagoda & Gillan, 2012).

VP5≡ $\forall\gamma$:TRACE, $\forall$t1, t2:TIME $\forall$G1,H1,J1,K1,D:REAL

[state(γ,t1)|= has_value(believable_cues, G1) &

state(γ,t1)|= has_value(believable_behaviours, H1) &

state(γ,t2)|= has_value(physical_appearances, J1) &

state(γ,t2)|= has_value(long_term_trust, K1) &

G1 $\geq$ 0.9 & H1 $\geq$ 0.9 & J1 $\geq$ 0.9 &

t2= t1+ D] $\Rightarrow$ K1 $\geq$ 0.5

CHAPTER SIX
CONCLUSION

6.1 Introduction

This study has shown the importance of trust in human robot interaction whereby humans are more willingness to interact with physical robot on the contrary to virtual character or written paper. This chapter concludes the outcomes of this study and shed more light on the limitation and future work that might be done later.

6.2 Conclusion

The basic concept of trust has widely explored the term of trust in human robot interaction and yielded the answer of why people trust physically designed intelligent robots more than other intelligent objects such as virtual avatar. As presented in Chapter One, three core objectives were provided; the identification the important factors of dynamics trust in human robot in short term interaction, the development a formal model of dynamic factors and its relationships obtained from previous objective and the evaluation the developed model. These three objectives were achieved (refer to Chapter Four and Five).

Based on literature review, a set of major factors were identified that determine or instigate trust during human robot interaction especially while a short term interaction. In details, 18 factors were identified that include; personality, physical appearances, believable behaviour, behaviour cues, level of automation, positive experiences, transparency, perception, long term perceive risk, short term perceive risk, reliable behavior, perceive competency, positive deception, long term positive experiences, short term trust, short term distrust, long term trust, long term distrust. All these factors have achieved first objective.

The second objective was achieved and as a result, the formal model (computational model) was developed by formalizing all identified factors from previous objective. Later a set of differential equations was formulated to make all the identified factors in executable features to be codded in a numerical simulation tool (Matlab). Next, simulation experiments were implemented to get further insights on the interactions among all identified factors to determine level of trust. To achieve this, the model was simulated with five different scenarios, namely; 1) high level of trust in HRI, 2) moderate high level of trust in HRI, 3) moderate level of trust in HRI, 4) moderate low level of trust in HRI and 5) low level of trust in HRI that present various circumstances based on external factors as presented in Sections 4.3, 4.3.1, 4.3.2, and 4.3.3.

Moreover, evaluation processes were implemented correctly whereby third objective was achieved. Verification (stability and logical verification) techniques have been used to assert the correctness and the internal validity of the developed model. In case of mathematical analysis a number of cases from equillibria points were used to show the stability as mentioned in Section 5. Meanwhile, five different real cases have been identified from the past literature. These cases were formalized in Temporal Trace Language (TTL) in order to execute an automated verification analysis as reported in Section 5. Both techniques have proved the correctness of the developed model.

6.3 Limitation and Future work

This study presented a cognitive model of dynamics trust during human robot interaction based on past cognitive and psychological theoretical studies. However, due to time constraint, the process of analysis during external validation has not been implemented yet. Therefore, this provides a future prospect to evaluate the external validity of trust model. Moreover, the model needs to integrate more factors related to humans such as biological factors to cover all aspects that might determine the level of trust in human robot interaction.

Furthermore, trust model needs to be integrated with the robotic technology; thereby the developers will be able to evaluate the developed model of trust in real time for short and long term interaction. Finally, this model is evaluated specifically in term of human -robot interaction during short term interaction. Thus, it offers great insights to investigate the applicability of this model in terms of human-robot teamwork.

REFERENCES

Aljazzaf, Z. M., Perry, M., & Capretz, L. (2010). *Online trust: Definition and principles.* Paper presented at the Computing in the Global Information Technology (ICCGI), 2010 Fifth International Multi-Conference on.

Armstrong Oma, K. (2010). Between trust and domination: social contracts between humans and animals. *World archaeology, 42*(2), 175-187.

Asada, M. (2013). Cognitive Developmental Robotics: from Physical Interaction to Social One *Autonomous Control Systems and Vehicles* (pp. 37-51): Springer.

Atkinson, D., Hancock, P., Hoffman, R. R., Lee, J. D., Rovira, E., Stokes, C., & Wagner, A. R. (2012). *Trust in Computers and Robots: The Uses and Boundaries of the Analogy to Interpersonal Trust.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Atkinson, D. J., & Clark, M. H. (2014). Attitudes and Personality in Trust of Intelligent, Autonomous Agents. *Manuscript in review*.

Bainbridge, Wilma, Hart, Justin, Kim, Elizabeth S, & Scassellati, Brian. (2008). The effect of presence on human-robot interaction. Paper presented at the Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on.

Barbalet, J. (2009). A characterization of trust, and its consequences. *Theory and society, 38*(4), 367-382.

Barnes, M., & Jentsch, F. (2010). *Human-robot interactions in future military operations*: Ashgate Publishing Company.

Beck, H., Dzindolet, M., & Pierce, L. (2002). Applying a decision-making model to understand misuse, disuse and appropriate automation use. *Advances in human factors and cognitive engineering, 2*.

Beckles, B., Welch, V., & Basney, J. (2005). Mechanisms for increasing the usability of grid security. *International Journal of Human-Computer Studies, 63*(1), 74-101.

Billings, D., Schaefer, K., Llorens, N., & Hancock, P. (2012). *What Is Trust? Defining the Construct Across Domains.* Paper presented at the Poster presented at the American Psychological Association Conference, Division.

Billings, D. R., Schaefer, K. E., Chen, J. Y., & Hancock, P. A. (2012). *Human-robot interaction: developing trust in robots.* Paper presented at the Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction.

Billings, D. R., Schaefer, K. E., Chen, J. Y., Kocsis, V., Barrera, M., Cook, Hancock, P. A. (2012). Human-animal trust as an analog for human-robot trust: A review of current evidence: DTIC Document.

Bonchek-Dokow, E., & Kaminka, G. A. (2013). Towards computational models of intention detection and intention prediction. *Cognitive systems research*.

Bosse, T., Both, F., Duell, R., Hoogendoorn, M., Klein, M. C., Lambalgen, Treur, J. (2013). An ambient agent system assisting humans in complex tasks by analysis of a human's state and performance. *International Journal of Intelligent Information and Database Systems, 7*(1), 3-33.

Bosse, T., Pontier, M., & Treur, J. (2010). A computational model based on gross' emotion regulation theory. *Cognitive systems research, 11*(3), 211-230.

Both, F., Hoogendoorn, M., Klein, M. C., & Treur, J. (2010). Computational modeling and analysis of therapeutical interventions for depression *Brain Informatics* (pp. 274-287): Springer.

Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D., & Blumberg, B. (2005). Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life, 11*(1-2), 31-62.

Casiddu, N., Cavallo, F., Divano, A., Mannari, I., Micheli, E., Porfirione, Dario, P. (2014). *Robot Interface Design of Domestic and Condominium Robot for Ageing Population.* Paper presented at the Ambient Assisted Living: Italian Forum 2013.

Castelfranchi, C., & Falcone, R. (2010). *Trust theory: A socio-cognitive and computational model* (Vol. 18): John Wiley & Sons.

Coeckelbergh, M. (2011). Humans, animals, and robots: A phenomenological approach to human-robot relations. *International Journal of Social Robotics, 3*(2), 197-204.

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance. *Journal of applied psychology, 92*(4), 909.

Cramer, H., Goddijn, J., Wielinga, B., & Evers, V. (2010). *Effects of (in) accurate empathy and situational valence on attitudes towards robots.* Paper presented at the Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction.

Daellenbach, U. S., & Davenport, S. J. (2004). Establishing trust during the formation of technology alliances. *The Journal of technology transfer, 29*(2), 187-202.

Dautenhahn, K. (2007a). Methodology and themes of human-robot interaction: a growing research field. *International Journal of Advanced Robotic Systems*.

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

De Santis, A., Siciliano, B., De Luca, A., & Bicchi, A. (2008). An atlas of physical human–robot interaction. *Mechanism and Machine Theory, 43*(3), 253-270.

Desai, M. (2012). Modeling trust to improve human-robot interaction.

Desai, M., Stubbs, K., Steinfeld, A., & Yanco, H. (2009). Creating trustworthy robots: Lessons and inspirations from automated systems.

Dimitrakos, T. (2001). System models, e-risks and e-trust *Towards the E-Society* (pp. 45-58): Springer.

Dzindolet, M. T., Beck, H. P., Pierce, L. G., & Dawe, L. A. (2001). A framework of automation use: DTIC Document.

Ellner, S. P., & Guckenheimer, J. (2011). *Dynamic models in biology*: Princeton University Press.

Farrell, S., & Lewandowsky, S. (2010). Computational models as aids to better reasoning in psychology. *Current Directions in Psychological Science, 19*(5), 329-335.

Fasola, J., & Mataric, M. (2013). A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction, 2*(2), 3-32.

Feil-Seifer, D., & Mataric, M. J. (2005). *A multi-modal approach to selective interaction in assistive domains.* Paper presented at the Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on.

Feltovich, P. J., Bradshaw, J. M., Clancey, W. J., & Johnson, M. (2007). Toward an ontology of regulation: Socially-based support for coordination in human and machine joint activity *Engineering Societies in the Agents World VII* (pp. 175-192): Springer.

Finkel, M. (2012). The Cold Patrol. *National Geographic, 221*(1), 82-95.

Fischer, K. (2011). *Interpersonal variation in understanding robots as social actors.* Paper presented at the Proceedings of the 6th international conference on Human-robot interaction.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology:* Human perception and performance, 3(4), 552.

Flanagin, Cory. A Survey of Robotics Systems and Performance Analysis.

Fong, T., Thorpe, C., & Baur, C. (2003). Collaboration, dialogue, human-robot interaction *Robotics Research* (pp. 255-266): Springer.

Freedy, E., DeVisser, E., Weltman, G., & Coeyman, N. (2007). *Measurement of trust in human-robot collaboration.* Paper presented at the Collaborative Technologies and Systems, 2007. CTS 2007. International Symposium on.

Freeman, C., Exell, T., Meadmore, K., Hallewell, E., & Hughes, A.-M. (2014). Computational models of upper limb motion during functional reaching tasks for application in FES based stroke rehabilitation. *Biomedical Engineering Journal*.

Goetz, J., Kiesler, S., & Powers, A. (2003). *Matching robot appearance and behavior to tasks to improve human-robot cooperation.* Paper presented at the Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on.

Goodrich, M. A., & Schultz, A. C. (2007). Human-robot interaction: a survey. *Foundations and trends in human-computer interaction, 1*(3), 203-275.

Grodzinsky, F., Miller, K., & Wolf, M. (2012). Moral responsibility for computing artifacts: the rules and issues of trust. *ACM SIGCAS Computers and Society, 42*(2), 15-25.

Groom, V., & Nass, C. (2007). Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies, 8*(3), 483-500.

Hancock, P. A., Billings, D. R., Oleson, K. E., Chen, J. Y., De Visser, E., & Parasuraman, R. (2011). A meta-analysis of factors influencing the development of human-robot trust: DTIC Document.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 53*(5), 517-527.

Hardin, R. (2001). Conceptions and explanations of trust. *Trust in society, 2*, 3-40.

Haring, K. S., Matsumoto, Y., & Watanabe, K. (2013). *How do people perceive and trust a lifelike robot.* Paper presented at the International conference on intelligent automation and robotics (ICIAR'13).

Haring, K. S., Watanabe, K., & Mougenot, C. (2013). *The influence of robot appearance on assessment.* Paper presented at the Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction.

Helton, W. S. (2009). *Canine ergonomics: the science of working dogs*: CRC Press.

Holm, A., Lukander, K., Korpela, J., & Sallinen, M. (2009). Estimating Brain Load from the EEG. *TheScientific World JOURNAL, 9*, 639-651.

Jianxin, Z., & Bond, M. H. (1992). Target-based interpersonal trust: Cross-cultural comparison and its cognitive model. *Acta Psychologica Sinica, 25*(2), 164-172.

Jøsang, A., Ismail, R., & Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decision support systems, 43*(2), 618-644.

Keaveney, S. M. (2008). Equines and their human companions. *Journal of business research, 61*(5), 444-454.

Khan, M. U. G., Bashir, S., Nasir, A., Shah, A. A., & Mehmood, S. (2014). Towards Computational Model of Human Brain Memory. *Journal of Mathematics (ISSN 1016-2526), 46*, 2.

Kim, J., & Moon, J. Y. (1998). Designing towards emotional usability in customer interfaces—trustworthiness of cyber-banking system interfaces. *Interacting with computers, 10*(1), 1-29.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 46*(1), 50-80.

Lee, J. J., Knox, W. B., Wormwood, J. B., Breazeal, C., & DeSteno, D. (2013). Computationally modeling interpersonal trust. *Frontiers in psychology, 4*.

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: a survey. *International Journal of Social Robotics, 5*(2), 291-308.

Lin, P., Bekey, G., & Abney, K. (2008). Autonomous military robotics: Risk, ethics, and design: DTIC Document.

Lu, G. P., Wang, Q. H., & Liu, S. B. (2014). Study on the Coordinated Control Mechanism of Orchestic Robotic Vehicle Based on Educational Robot. *Applied Mechanics and Materials, 536*, 974-978.

Lyons, J. B. (2013). *Being transparent about transparency: A model for human-robot interaction.* Paper presented at the 2013 AAAI Spring Symposium Series.

Madsen, M., & Gregor, S. (2000). *Measuring human-computer trust.* Paper presented at the Proceedings of Eleventh Australasian Conference on Information Systems.

Marble, J. L., Bruemmer, D. J., Few, D. A., & Dudenhoeffer, D. D. (2004). *Evaluation of supervisory vs. peer-peer interaction with human-robot teams.* Paper presented at the System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on.

Marsella, S., Gratch, J., & Petta, P. (2010). Computational models of emotion. *A Blueprint for Affective Computing-A sourcebook and manual*, 21-46.

Masthoff, J. (2007). Computationally modelling trust: an exploration. *University of Aberdeen, Aberdeen, Scotland, UK*.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review, 20*(3), 709-734.

McKnight, D. H., & Chervany, N. L. (1996). The meanings of trust.

McKnight, D. H., & Chervany, N. L. (2000). What is trust? A conceptual analysis and an interdisciplinary model. *AMCIS 2000 Proceedings*, 382.

McKnight, D. H., & Chervany, N. L. (2001). Trust and distrust definitions: One bite at a time *Trust in Cyber-societies* (pp. 27-54): Springer.

Mui, L., Mohtashemi, M., & Halberstadt, A. (2002). *A computational model of trust and reputation.* Paper presented at the System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37*(11), 1905-1922.

Nickerson, J. V., & Reilly, R. R. (2004). *A model for investigating the effects of machine autonomy on human behavior.* Paper presented at the System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on.

Ogreten, S., Lackey, S., & Nicholson, D. (2010). *Recommended roles for uninhabited team members within mixed-initiative combat teams.* Paper presented at the Collaborative Technologies and Systems (CTS), 2010 International Symposium on.

Oleson, K. E., Billings, D., Kocsis, V., Chen, J. Y., & Hancock, P. (2011). *Antecedents of trust in human-robot collaborations.* Paper presented at the Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2011 IEEE First International Multi-Disciplinary Conference on.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 39*(2), 230-253.

Righetti, F., & Finkenauer, C. (2011). If you are able to control yourself, I will trust you: the role of perceived self-control in interpersonal trust. *Journal of personality and social psychology, 100*(5), 874.

Robinette, Paul, Wagner, Alan R, & Howard, Ayanna M. (2013). *Building and Maintaining Trust Between Humans and Guidance Robots in an Emergency.* Paper presented at the AAAI Spring Symposium: Trust and Autonomous Systems.

Robinette, P., Wagner, A. R., & Howard, A. M. (2014). *Modeling Human-Robot Trust in Emergencies.* Paper presented at the 2014 AAAI Spring Symposium Series.

Rodić, A., & Jovanović, M. (2014). *How to Make Robots Feel and Social as Humans.* Paper presented at the COGNITIVE 2014, The Sixth International Conference on Advanced Cognitive Technologies and Applications.

Sanders, T., Oleson, K. E., Billings, D., Chen, J. Y., & Hancock, P. (2011). *A Model of Human-Robot Trust Theoretical Model Development.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Sanders, T. L., Wixon, T., Schafer, K. E., Chen, J. Y., & Hancock, P. (2014). *The influence of modality and transparency on trust in human-robot interaction.* Paper presented at the Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2014 IEEE International Inter-Disciplinary Conference on.

Schaefer, K. E. (2013). *The Perception and Measurement of Human-robot Trust.* University of Central Florida Orlando, Florida.

Schilling, R. (2013). Fundamentals of robotics.

Scholtz, J. (2003). *Theory and evaluation of human robot interactions.* Paper presented at the System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on.

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of management review, 32*(2), 344-354.

Schwarz, M., Stückler, J., & Behnke, S. (2014). *Mobile teleoperation interfaces with adjustable autonomy for personal service robots.* Paper presented at the Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction.

Sheppard, B. H., & Sherman, D. M. (1998). The grammars of trust: A model and general implications. *Academy of management review, 23*(3), 422-437.

Shiflet, A. B., & Shiflet, G. W. (2014). *Introduction to computational science: modeling and simulation for the sciences*: Princeton University Press.

Sidner, C. L., Kidd, C. D., Lee, C., & Lesh, N. (2004). *Where to look: a study of human-robot engagement.* Paper presented at the Proceedings of the 9th international conference on Intelligent user interfaces.

Simpson, J. A. (2007). Foundations of interpersonal trust.

Sun, R. (2006). *Cognition and multi-agent interaction: From cognitive modeling to social simulation*: Cambridge University Press.

Taddeo, M. (2010). Modelling trust in artificial agents, a first step toward the analysis of e-trust. *Minds and Machines, 20*(2), 243-257.

Takayama, L., & Pantofaru, C. (2009). *Influences on proxemic behaviors in human-robot interaction.* Paper presented at the Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on.

Tapus, Adriana. (2014). Towards personality-based assistance in human-machine interaction. Paper presented at the Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on.

Thilakarathne, D. J., & Treur, J. (2013). A Computational Cognitive Model for Intentional Inhibition of Actions. *Procedia-Social and Behavioral Sciences, 97*, 63-72.

Trang, S., Zander, S., & Kolbe, L. M. (2014). Dimensions Of Trust In The Acceptance Of Inter-Organizational Information Systems In Networks: Towards A Socio-Technical Perspective.

Urbano, J., Rocha, A. P., & Oliveira, E. (2013). A socio-cognitive perspective of trust *Agreement Technologies* (pp. 419-429): Springer.

van den Brule, R., Dotsch, R., Bijlstra, G., Wigboldus, D. H., & Haselager, P. (2014a). Do Robot Performance and Behavioral Style affect Human Trust? *International Journal of Social Robotics, 6*(4), 519-531.

van den Brule, R., Dotsch, R., Bijlstra, G., Wigboldus, D. H., & Haselager, P. (2014b). Do Robot Performance and Behavioral Style affect Human Trust? *International Journal of Social Robotics*, 1-13.

Wainer, J., Feil-Seifer, D. J., Shell, D. A., & Mataric, M. J. (2006). *The role of physical embodiment in human-robot interaction.* Paper presented at the Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on.

Wells, P., & Deguire, D. (2005). *TALON: A universal unmanned ground vehicle platform, enabling the mission to be the focus.* Paper presented at the Defense and Security.

Wilson, C. C. (1994). A conceptual framework for human-animal interaction research: The challenge revisited. *Anthrozoos: A Multidisciplinary Journal of The Interactions of People & Animals, 7*(1), 4-24.

Yagoda, R. (2011). WHAT! You Want Me to Trust a ROBOT? The Development of a Human Robot Interaction (HRI) Trust Scale.

Yagoda, Rosemarie E, & Gillan, Douglas J. (2012). You want me to trust a ROBOT? The development of a human–robot interaction trust scale. International Journal of Social Robotics, 4(3), 235-248.

Zeller, & Smith. (2014). Hitchhiking robot thumbs its way across Canada,. Retrieved 2 August, 2014

Zhu, Y., & Zhang, W. (2008). *The Cognitive Model of Interfirm Trust Dynamics: How Trust Is Evolved?* Paper presented at the 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing.