INITIAL TRUST IN AI AGENT: COMMUNICATING FACIAL ANTHROPOMORPHIC TRUSTWORTHINESS FOR SOCIAL ROBOT DESIGN

Yao SONG

Ph.D.

The Hong Kong Polytechnic University

2021

The Hong Kong Polytechnic University

School of Design

Initial Trust in AI Agent: Communicating Facial Anthropomorphic Trustworthiness for Social Robot Design

Yao SONG

A thesis submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

January 2021

Certification of Originality

I hereby declare that this thesis is my work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

_____________________ (Signed)

Yao SONG (Name of student)

ABSTRACT

As a technical application in artificial intelligence (AI), a social robot is one of the branches of robotic studies, which emphasizes socially communicating and interacting with human beings. Distinct from other humanoid and industrial robots, which usually have limited human-like features, the state-of-art (SOTA) social robots are usually equipped with a screen-based head, interfaced with an animated or human-like face, to respond and meet the need of human being.

Similar to interpersonal interaction, trustworthiness towards a social robot is crucial in human-robot interaction since the social robot, in daily lives, usually acts as a "listener and responder", providing not only practical assistance but also emotional support for humans. In this view, it is important that it should be considered as a trustworthy "partner or friend". However, to communicate facial trustworthiness for a social robot is still a challenging problem where few studies have tried to address this issue in a comprehensive way. This thesis tries to fulfill this research gap by deliberately exploring the concept, structurally identifying significant features, examining, modeling the effects for statistic and dynamic features, and providing practical guidelines.

Different from the constructs from human trustworthiness, facial anthropomorphic trustworthiness could have four distinct dimensions: capability, anthropomorphism, positive affect, and ethics concern. Accordingly, facial anthropomorphic trustworthiness could be initially summarized as impression-based trustworthiness where people rely on facial cues of social robots to evaluate their capability, ethics

concern, anthropomorphism, and positive affect. As for the essential features communicating facial anthropomorphic trustworthiness, four facial categories emerged accordingly: internal features, external features, combinations of features, and emotional expressions.

For static features, internal features (eye shape and mouth shape), external features (fWHR and face shape), and feature combination (feature sizes and positions) were analyzed. In general, eye shape and mouth shape have a significant impact on facial anthropomorphic trustworthiness where a robotic face with round eyes (vs. narrow eyes) and upturned or neutral mouth (vs. downturned mouth) enjoyed a higher level of facial anthropomorphic trustworthiness. fWHR has a significant impact on facial anthropomorphic trustworthiness where a robotic face with high fWHR (vs. low fWHR) enjoyed a higher level of facial anthropomorphic trustworthiness. Eye size and positions of eyes and mouth have a significant impact on facial anthropomorphic trustworthiness for social robot where a robotic face with large eyes (vs. small eyes), medium horizontal and vertical position of eye and mouth (vs. high or low horizontal and vertical positions) enjoyed a higher level of facial anthropomorphic trustworthiness. Moreover, a model for facial anthropomorphic trustworthiness in static features was developed and a validation study via variations in planned and random stimuli was conducted to confirm the reliability of the current model.

For dynamic features, facial valence and arousal, and their interaction with regulatory-focused contexts were analyzed. In general, facial valence and its interaction with regulatory contexts have a significant impact on facial anthropomorphic trustworthiness where a robotic face with positive (vs. negative expressions) enjoyed a higher level of facial anthropomorphic trustworthiness while positive expressions were compatible with the promotion-focused context and negative expressions were compatible with prevention-focused context, in signaling facial anthropomorphic trustworthiness. Similarly, a model for facial anthropomorphic trustworthiness in dynamic features was developed.

This research would advance the theory of human-robot interaction. This study, for the first time, set out to explore and develop the meaning of facial anthropomorphic trustworthiness, providing a relatively holistic picture when evaluating the trustworthiness of a social robot at first sight. In addition, through a series of experiments, this research tried to adopt both subject and physiological measures to examine the difference and similarities between human facial features and robot facial features in signaling perceived trustworthiness and further model and discuss the mechanism leading to such phenomenon.

As for practical implications, although the current market has various social robots and some of them even won some honors, detailed and specific guidelines are still missing to some extent that companies design a social robot primarily relying on their own understandings and intuitions. Accordingly, this research advanced the design guidelines of a social robot, which could work as preliminary instructions in helping robot designers to make a trustworthy robot and increasing purchase intentions for the consumers.

PUBLICATIONS ARISING FROM THE THESIS

Song, Y.; Luximon, Y. The Face of Trust: The Effect of Robot Face Ratio on Consumer Preference. *Computers in Human Behavior*, 2021, 116: 106620. (SSCI Q1)

Song, Y.; Luximon, A; Luximon, Y. The Effect of Facial Features on Facial Anthropomorphic Trustworthiness in Social Robots. *Applied Ergonomics*, 2021, 94, 103420. (SSCI Q1)

Song, Y.; Luximon, Y. Trust in AI Agent: A Systematic Review of Facial Anthropomorphic Trustworthiness for Social Robot Design. *Sensors* 2020, 20, 5087. (SCI Q1)

Song, Y. Building a "Deeper" Trust: Mapping the Anthropomorphic Facial Trustworthiness in Social Robot Design through Multidisciplinary Approaches. *The Design Journal* 2020, 2. (A&HCI)

Song, Y.; Luximon, Y.; Luo J. A Moderated Mediation Analysis of the Effect of Lettering Case on Trustworthiness Perception and Investment Decision in Advertising. *International Journal of Bank Marketing* 2020, 3. (SSCI Q2)

Song, Y.; Luximon, Y. Role of Trust in Value Sensitive Design, Spectatorship and Identity. ACED 2020, Philippines, 2020 (EI)

Song, Y.; Luximon, Y. Hotspots and Trends in Creativity and Design Research: A Knowledge Graph Analysis. AHFE2020, San Diego, U.S.A, 2020 (EI)

Song, Y.; Luximon, Y. BD Leong, Z Qin. The E-Commerce Performance of Internet of Things (IoT) in Disruptive Innovation: Case of XiaoMi. ICSEB2019, Tokyo, Japan, 2019 (EI)

Song, Y.; Luximon, Y. Exploring the Effect of Wordmark in E-Commerce Website Interface Design. ACHI2019, Athens, Greece, 2019

ACKNOWLEDGEMENTS

In the past three years, I have imagined thousands of times writing acknowledgment when finishing my Ph.D. thesis. I guessed the moment when I began to write up this part would be joyful and emotional. You know what? It is true: a feeling of a dream coming true.

Throughout my exploration in the academics at PolyU, I have to say that I have received a great deal of assistance and support from various parties.

First of all, I would like to thank my chief supervisor, Dr. Luximon, with my deepest gratitude. I do appreciate her conscious and insightful consideration and guidance during my whole Ph.D. period. With her help and advice, I am tutored not only the academic expertise and latest trends but also the creativity and rigorous requirements for the greater good, which would definitely nurture me for the rest of my life.

In addition, I would like to thank my co-supervisor, Benny Ding Leung, for his mentorship and support. Also, I would like to express my deep gratitude to Prof. Lee and Prof. Siu, Dean and Chair Professor of School of Design, PolyU. Without their extraordinary leadership, our school cannot keep holding a leading position in the world. I would also like to acknowledge The Hong Kong Polytechnic University for providing me with a UGC fellowship to support my Ph.D. study.

Moreover, many thanks are given to Carrie Lam, Hong Kong Government, Shen Zhen Government, Apartment One of China Merchants Shekou, and related civil servants in Hong Kong and Shen Zhen. It is the Hong Kong government that deliberately sends a plane to take me back to Hong Kong during the COVID lockdown. It is the Shen Zhen government that provides a cozy shelter for me during the social activity.

My appreciations also go to my peers and colleagues at PolyU and Shenzhen University. First of all, I would thank my best friend and partner, Ms. Zhenzhen Qin. It is she who keeps encouraging and discussing my research. This mutual learning process indeed helps me to overcome various academic difficulties. Moreover, I sincerely appreciate Ms. Jing Luo who assisted me with the experiment implementation and supported me in life. Dr. Tao Da and his graduate students at Shen Zhen University were specially thanked for their support and assistance for my research. Ms. Lai Wei, Ms. Jing Ding, Ms. Suxin Zhang, Ms. Fu Fang, Ms. Jiaxin Zhang, and Ms. Jun Zhang are also thanked for their kind help in my research and life, which makes this Ph.D. a beautiful journey.

Furthermore, I appreciate the support from my family members, mom, and dad. Without their considerations and helps, it would be impossible for me to go through this tough but fruitful journey. Also, thanks to the girlfriend who hasn't shown up yet, which lets me immerse myself in the academic ocean.

CONTENTS

List of Figures

List of Tables

CHAPTER ONE. INTRODUCTION

This chapter introduces the motivations, the relevant terms, philosophies, research objectives, research questions regarding trustworthiness, and robot design. It aims to provide a philosophical background and significance of the current study.

1.1 Motivation and Research Background

Following the evolution of technology and its application in various daily contexts (Del Río et al., 2020; González-González et al., 2020, 2019a; Torres-Carrión et al., 2019), the latest innovation is the social robot, which is an advanced artificial intelligence (AI) system to communally interact with humans (Fasola and Mataric, 2015; Fernández-Rodicio et al., 2020; González-González et al., 2019b). Distinct from established robots (e.g., mechanical-looking or industrialized robots), which have seldom human characteristics, some advanced robots (anthropomorphic robots, i.e. Pepper, NAO, ASIMO, Kaspa, Cozmo, and Buddy robot) are featured with a screen that renders an animated anthropomorphic face to interact with human (Oh and Ju, 2020; Westlund et al., 2016). For example, Figure 1.1 illustrates the Buddy robot, which is characterized by its anthropomorphic facial features to affectionately support humans and satisfy their needs. Because social cognition of anthropomorphic objects could generalize people's understanding of human-relevant knowledge toward a social robot (Prakash and Rogers, 2015), it could be needed to have a face-like interface for facilitating interaction in human-robot relationships (McGinn, 2019; Stroessner and Benitez, 2019).

Figure 1.1 An example of a social robot – Buddy®

In social robots, the characteristics people perceive such as dominance, friendliness, and attractiveness, trustworthiness plays a crucial role in human-robot relationships based on two reasons (Mathur and Reichling, 2016). First, trustworthiness plays an essential role in interpersonal communication because it has a critical influence on persuasion (Salem et al., 2015; Song et al., 2020). Second, social robots indeed work as "close friends or even family members," providing not only physical assistance but also affectionately encouragement toward humans. Accordingly, social robots should be appropriately designed to be trusted and be relied on (Yu et al., 2015).

Gompei and Umemuro (2018) indicated that several types of factors play essential roles in influencing the perceived trustworthiness in the context of social robots: robot-related factors (i.e. the specific features and activities of the robot), human-related factors (i.e. the particular need and situational awareness), and scenario-related factors (i.e. the specific application of task, collaboration culture, and interpersonal communication). Within these factors, robot-related factors are considered as the most vital because they influence the perceived trustworthiness judgment in HRI (Hancock et al., 2011; Saunderson and Nejat, 2019). More specifically, such factors might be related to social robots' actions, reliability, predictability, level of automation, and the degree of anthropomorphism (Hancock et al., 2011). For instance, related studies have indicated that rendered anthropomorphic faces are more likable and tend to make people experience more arousal (Luo et al., 2006; McGinn, 2019; Sproull et al., 1996), eventually resulting in a higher level of perceived trustworthiness toward the social robot (compared social robots with mechanical faces) (Prakash and Rogers, 2015).

Humans examine faces or facial characteristics on humans and objects, such as robots, in an extremely short time (Landwehr et al., 2011). A prior study suggested that 100 ms was enough for individuals to examine various personality features, i.e. trustworthiness, competence, and aggressiveness evaluation (Willis and Todorov, 2006). The reason for this judgment might lie in the human him/herself, namely cephalization, meaning how the physical organs and neural systems tend to concentrate in the upper part of the body (i.e., the head) (Holliday, 1993). While examining an anthropomorphic robot, individuals might process concrete facial features or expressions in robots by comparing specific characteristics with human and evaluating the difference between a robot and human face (Maeng and Aggarwal, 2018; McGinn, 2019; Stroessner and Benitez, 2019).

When a person encounters another person or sees a robot, he or she might form a mental image of them quickly. This is called a first impression, and it is crucial in people's daily lives; it not only occurs in human perception but also the formation of an attitude toward a person or social robot (Maeng and Aggarwal, 2018; Stroessner and Benitez, 2019; Verberne et al., 2015). Moreover, first impressions occur unconsciously, difficult to recognize, and influence people's decision-making process (Bar et al., 2006; Yan et al., 2015). One study discussed positive attitudes and social perceptions toward specific appearances in the context of humans, suggesting a rule-of-thumb named the "beauty premium" or "plainness penalty" (Etcoff et al., 2011).

Specifically, facial features are strong predictors of the first impression. They work as significant indicators in the formation of a mental image and an initial evaluation of someone's attributes (Sofer et al., 2015). People with specific facial features tend to have a higher likelihood of being trusted, being liked, and ultimately obtaining more actual benefits (Etcoff, 1994; Etcoff et al., 2011). This "halo" effect of particular facial features could be interpreted by people's subconscious facial-feature processing; specific features could contribute to various positive interpretations or expectations about others (Aharon et al., 2004). For example, some facial features such as a smile could be considered a social signal of being friendlier, more capable, more confident, and more outstanding (Etcoff, 1994).

Among various social perceptions, human trustworthiness plays a crucial role in human relationships and related behavioral responses (Calvo et al., 2017; Lyons et al., 2020; Song and Luximon, 2019). Indeed, humans are skilled in evaluating facial trustworthiness based on physical traits (Hoff and Bashir, 2015). For example, Wout and Sanfey (2008) suggested that people had a higher intention to interact, cooperate, and invest with trustworthy-looking individuals in an interactive risky decision-making game.

Facial features could influence the anthropomorphic trustworthiness of artificial intelligence agents (Ghazali et al., 2018; Gompei and Umemuro, 2018; Saunderson and Nejat, 2019). Related review research on facial trustworthiness has largely focused on outlining trustworthiness and analyzing general perceptions of trustworthiness within human-computer/human-robot relationships. However, the concept “trustworthiness” is a multi-construct term that contains two examination steps, the initial examination during the first impression and the latter examination during continuous interaction (Willis and Todorov, 2006). Moreover, though research has attempted to evaluate the influence of various human facial features in signaling the perceived trustworthiness (Santos et al., 2016), it might not be directly included in HRI because of the anthropomorphic distinction of AI agents, compared with human (Prakash and Rogers, 2015).

Indeed, the literature on the facial design of AI agents is multi-disciplinary research that is seldom comprehensively examined and occasionally involved by other literature. To specify, there are at least three main disciplines that are related to the current research (Song, 2020). First, because humans and AI agents, such as social robots, might share similar facial features, i.e. nose, ear, or eyes, facial trustworthiness originated from cognitive psychology could potentially shed light and bring inspiration to the anthropomorphic trustworthiness of AI agents (Stirrat and Perrett, 2010). In addition, anthropomorphic robots in the commercial market might also be sparked from the object or product appearance in the contexts of marketing and engineering design, which has explored the way to design trustworthy-looking anthropomorphic objects or products (Creusen and Schoormans, 2005), i.e. the frontal face of a car or a plug (Miesler et al., 2011). For instance, Maeng and Aggarwal (2018) indicated that the front looking of a car with a low level of facial width-height ratio (fWHR) could have a high level of perceived trustworthiness. Although anthropomorphic robot design is sporadically relevant with product design, it could offer sparks for making a trustworthy-looking robot as well, as they all enjoyed similar facial characteristics in signaling trustworthiness. Last, though the research on AI agents has primarily concentrated on the anthropomorphic trustworthiness from a general perspective, it could still shed sparks as a basis for a better understanding of trustworthiness in HRI (Prakash and Rogers, 2015).

1.2 Research Objectives

To obtain a better understanding of what creates facial anthropomorphic trustworthiness and how facial features influence it, a structural and in-depth exploration of its specific meaning and an examination of facial features in communicating anthropomorphic trustworthiness is required. Moreover, to contribute to the detailed design guidelines for creating a trustworthy robot, a facial anthropomorphic model is desired to structurally model related features to enjoy relatively holistic insights. To fulfill these aims, the research objectives of the current study were as follows:

- **Objective 1**: To investigate the meaning of facial anthropomorphic trustworthiness and its sub-dimensions in robot evaluation and to develop and validate a scale for further empirical studies.
- **Objective 2**: To examine the effect of separate and combinations of static features on facial anthropomorphic trustworthiness and develop a model that empirically predicts the relationship between facial features and perceived trustworthiness.
- **Objective 3**: To further examine the effect of dynamic features on facial anthropomorphic trustworthiness under different contexts.
- **Objective 4**: To summarize specific guidelines for improving facial anthropomorphic trustworthiness in robot design.

1.3 Scope and Research Questions

The majority of research on facial trustworthiness has mainly been conducted on human features, such as the eyes (Zebrowitz et al., 1996), mouth (Calvo et al., 2017), and nose (Santos and Young, 2011), which is comprehensively discussed in Chapter two; however, limited research has attempted to explore the effects of different facial features on anthropomorphic trustworthiness in the context of a social robot. The current thesis attempted to fulfill this research gap by deliberately exploring the concept, structurally identifying significant features, and examining and modeling the effects for statistic and dynamic features.

Trustworthiness evaluation mainly contained two phases: the initial evaluation and evaluation in the latter stage through continuous interaction (Song and Luximon, 2020a). Although the latter stage—trustworthiness evaluation through continuous interaction—is vital in HRI (Song and Luximon, 2020a), the current technical level of robot response and action might still constrain user experience. People might still perceive the unnatural or even wired behavior of social robots, which might have influenced the accuracy and reliability when integrating continuous interaction factors for the current study. Accordingly, the scope of this thesis would primarily focus on an impression-based evaluation of facial trustworthiness at first sight, where people would rely on both static and dynamic facial features to make the evaluation.

Accordingly, this thesis works as an initial step in developing a relatively comprehensive understanding of facial anthropomorphic trustworthiness. The current exploration and examination merge two streams of knowledge in the research scope: human facial trustworthiness in different daily contexts, which considers specific facial features for signaling trustworthiness, and anthropomorphism, which considers specific robot attributes, as depicted in Figure 1.2.

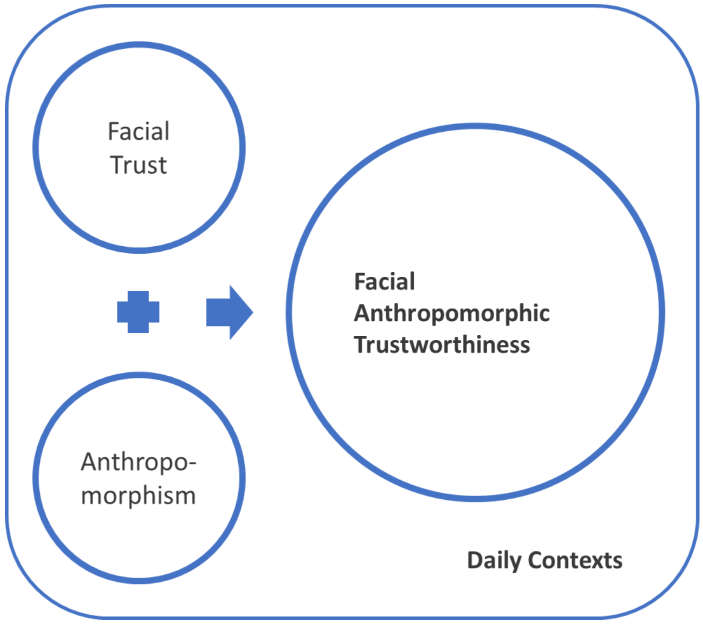


Figure 1.2 The scope of this research

The examination and modeling of HRI are indeed a sophisticated procedure that requires extensive data-driven experiments and simulation for different parameter configurations (Foo et al., 2017; Sim and Loo, 2015). In addition, although numerous human facial features exist, only a few potential features could play a significant role in communicating trustworthiness (Dotsch et al., 2017). Therefore, this study was limited to the examination and validation of specific potential features in previous studies that facilitate communicating facial trustworthiness.

The research questions (RQs) that guided the current study were as follows:

- **RQ1**: What is the facial anthropomorphic trustworthiness for a social robot?
- **RQ2**: How do static features influence facial anthropomorphic trustworthiness?
    - RQ2-a: How do the separate static features influence facial anthropomorphic trustworthiness?
    - RQ2-b: How do the combinations of static features influence facial anthropomorphic trustworthiness?
- **RQ3**: How do dynamic features influence facial anthropomorphic trustworthiness in daily contexts?

RQ1, RQ2, and RQ3 follow a hierarchical rationale where RQ1 serves as the theoretical foundation from a macro perspective while RQ2 and RQ3 further analyze the effect of different features on facial anthropomorphic trustworthiness from a micro level.

Particularly, RQ1 (study 2) mainly involves exploring the definition of facial anthropomorphic trustworthiness, which signals the directions of the impact. Indeed, the RQ1 works as an introduction and theoretical foundation for RQ2 and RQ3. It not only investigated the meaning of facial anthropomorphic trustworthiness for a social robot but also empirically assessed different dimensions of this concept. With the development of this scale, RQ1 helps to provide a toolkit and protocol to measure this concept.

Based on the results of RQ1, RQ2 and RQ3 further explore the directions of the relationship between different facial features and anthropomorphic trustworthiness. To be more specific, RQ2 (studies 2-5) focuses on the role of static features (separated and combinations) in signaling facial anthropomorphic trustworthiness while RQ3 (studies 6-7) emphasizes the role of dynamic features and contextual factors in signaling facial anthropomorphic trustworthiness.

With the exploration of RQ1-3, a relatively holistic view of facial anthropomorphic trustworthiness and its associated influential factors was determined: the meaning, sub-dimensions, and scale of facial anthropomorphic trustworthiness were initially established while various levels of factors (both static and dynamic features) were investigated and modeled for a relatively comprehensive understanding of facial anthropomorphic trustworthiness.

1.4 Significance of the Current Study

This research would advance the theory of human-robot interaction. Firstly, this study, for the first time, sets out to explore and develop the meaning of facial anthropomorphic trustworthiness, which is rarely discussed in prior literature. Through synthesizing the theories among interpersonal trustworthiness, the uncanny valley, and general robot trustworthiness, it could try to provide a relatively holistic picture when we evaluated the trustworthiness of a social robot at the first impression. In addition, through a series of behavioral experiments, the current study tried to validate that facial trustworthiness features could be applied to social robot design and improve people's trustworthiness and attitude toward the social robot. Moreover, via examining different facial features in communicating facial anthropomorphic trustworthiness, this research tried to adopt both subject and physiological data to examine the difference and similarities between human facial features and robot facial features in signaling perceived trustworthiness and further model and discuss the mechanism leading to such phenomenon.

As for practical implications, although the current market has various social robots and some of them even won some honors (CES, 2018), detailed and specific guidelines are still missing to some extent that companies design a social robot primarily relying on their own understandings and intuitions (Vanderborght et al., 2012). Regarding the risk in intuition-based design that might dampen user experience and potential market performance (Ulrich, 1992), this research provides preliminary instructions for designing a trustworthy social robot. This research advanced the design guidelines of a social robot, which could work as preliminary instructions in helping robot designers to make a trustworthy robot and increasing purchase intentions for the consumers.

1.5 Research Framework

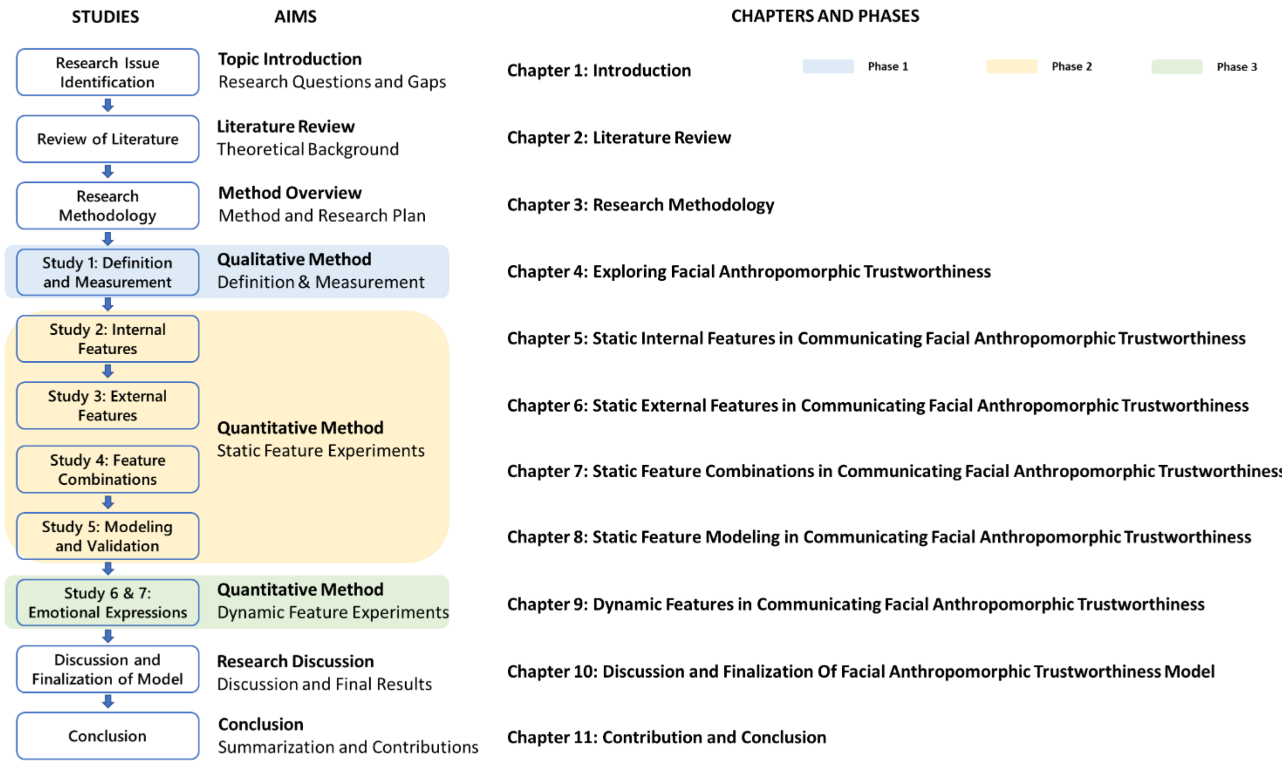


Figure 1.3 The research path diagram with three phases in this study

The research framework is shown in Figure 1.3:

Chapter 1 introduces the research background and motivations. The research objectives are proposed, and the scope and research questions are identified. The significance and value of the study are discussed.

Chapter 2 provides a review of the literature on potential features for communicating facial anthropomorphic trustworthiness. It reviews literature from three sources, robot trustworthiness, human trustworthiness, and product trustworthiness. Then it discusses and compares the feature differences in the context of humans and robots. Last, it summarizes the potential significant features like the research rationale for further chapters.

Chapter 3 describes the research methodology for this study. A mixed methodology was employed. Qualitative research provides an in-depth understanding of the concept's meaning and dimensions while quantitative research generalizes the quantified relationships between specific features and further develops anthropomorphic trustworthiness for designers.

Chapter 4 outlines the first phase of the study – a hybrid mixed method (via crowdsourcing and natural language processing) is introduced to explore people's understanding of facial anthropomorphic trustworthiness and its sub-dimensions (RQ 1). It also focuses on scale development and validation for facial anthropomorphic trustworthiness via iterative exploratory factor analysis and confirmative factor analysis. This chapter addresses objective 1 and RQ 1.

Chapters 5-8 outline the second phase of the study – a series of experiments is introduced to examine the effect of different static features (RQ 2-a), their combinations, and their modeling (RQ 2-b) on facial anthropomorphic trustworthiness. This study addresses objective 2 and RQ 2.

Chapter 9 outlines the third phase of the study – a lab experiment (via subjective and physiological measures) is introduced to examine the effect of different dynamic features on facial anthropomorphic trustworthiness under different daily contexts (RQ 3). This study addresses objective 3 and RQ 3.

Chapters 10-11 synthesize a general model (static and dynamic features) for facial anthropomorphic trustworthiness. Further, it summarizes and discusses the conclusions, limitations, and future studies of this research. Last it addresses the major finding and contributions for the theory and practice. This study addresses objective 4.

CHAPTER TWO. LITERATURE REVIEW

This chapter illustrates a literature review of the relevant concepts, knowledge, and theories regarding facial trustworthiness and robot design. Through structurally reviewing literature from human trustworthiness, product trustworthiness, and robot trustworthiness, this chapter aims to provide a holistic picture of the theoretical background of the current study.

2.1 Trustworthiness for Human, Product, and Robot

Although human trustworthiness, product trustworthiness, and robot trustworthiness are correlated and interwoven with each other, where trustworthiness refers to the ability to be relied on as honest or truthful (Frith, 2009; Guido and Peluso, 2009; Song and Luximon, 2020a; Xu, 2019), there are still difference among those concepts. While product trust focused on the assured expectations of a product's reliability and inclination in scenarios involving risk of its user (Delgado-Ballester, 2004; Munuera-Aleman et al., 2003), human trustworthiness is a mental interpretation of the possibility or the confidence that other social actors would act in line with his or her words (Ames et al., 2010; Walker et al., 2011).

Prior research has tried to explore and analyze the sub-dimensions of human trustworthiness: ability, benevolence, and integrity. To specify, ability in interpersonal trustworthiness refers to the individual's evaluation whether others' related capability and knowledge in the given task (Mayer et al., 1995). This evaluation is the initial expectation of others' experience or endorsement (for example, the academic certificate of a doctor). When it comes to HRI, it might refer to the belief that a social robot has the functions to accomplish its task (Thatcher et al., 2011). As for integrity, it refers to the individual's evaluation of whether others would obey a set of social rules in the interpersonal interaction (Mayer et al., 1995). This evaluation is the confidence initial expectation of the perceived risk and confidence in others' behavior. With regard to HRI, people might show various ethical concerns about robots. On the one hand, people would concern more about the integrity of the robot creator or designer since they are "the man behind the curtain" (Fosch-Villaronga et al., 2019). On the other hand, people might be anxious about the self-awareness of such emerging technology and even doubt they will stick to their program (Chanseau et al., 2016). As to benevolence, it refers to the degree that people tend to do well to themselves, even beyond their profit motivation (Mayer et al., 1995). This evaluation is the effective expectation of the kindness of other people, either in physical or psychological interaction. Regarding its role in HRI, people might expect the personality of a social robot should be altruistic and emotionally approachable (Stuck and Rogers, 2018).

Indeed, human and product trustworthiness has different objects: one considers its partner as a human where their conducts would generally follow social rules while the other treats it as a nonhuman item where the existence of an object is just to fulfill human needs (Ames et al., 2010; Bart et al., 2005; Kocsor and Bereczkei, 2017; Song and Luximon, 2020a; van 't Wout and Sanfey, 2008). Considering robot trustworthiness, its special difference between human trustworthiness and product trustworthiness lies in its anthropomorphic nature (Rosenthal-von der Pütten et al., 2019; Y. Zhang et al., 2019). On one hand, social robots are taking various kinds of social roles and actively engaging and bonding with humans, which makes them friends or companions for humans (Hoorn, 2018). On the other hand, they were initially designed as a humanlike machine that aims to help people get works done (Al-Qaderi et al., 2018; Förster et al., 2019). Therefore, robot trustworthiness stands in the middle between human trustworthiness and product trustworthiness, which would be influenced by not only the affective reactions, i.e. human trustworthiness, but also the logical reasoning, i.e. product trustworthiness.

2.2 Anatomy of Facial Features of Trustworthiness

The nature of this Ph.D. research decides its multiple disciplinary directions, which requires various areas to respond to the issues. Particularly, the current research involves theories from human trustworthiness, anthropomorphism, and robot design. Accordingly, the review of literature on particular features of humans, products, or robots in communicating trustworthiness can be categorized into two main groups: static features and dynamic features.

Both static features and dynamic features rely on specific states of features in communicating facial trustworthiness. For example, the eye and mouth regions are considered the most salient characteristics for both static and dynamic features (Calvo et al., 2019; Riegelsberger et al., 2004), though both of them have their distinct features, such as fWHR for static features and regulatory-focus for dynamic features. While static attributes of the eye and mouth regions, such as shape and size of eye and mouth, play a significant role in facial trustworthiness (Ferstl et al., 2016), dynamic attributes of the eye and mouth regions, such as smile or anger, also have a critical impact on trustworthiness evaluation (Centorrino et al., 2015).

Based on an extensive literature review, this research focuses on not only the static attributes of specific features, which constitute internal features, external features, combinations of features, but also discusses their dynamic attributes, such as the emotional expressions (see Table 2.1 and Appendix A). To specify, internal features contained static attributes of eye and mouth shape; external features contained features like fWHR and face shape; combinations of features involve a group of features that enjoy certain typicality, such as baby schema; and dynamic features mainly involves dynamic attributes of eye and mouth regions for emotional expressions, such happiness or depression (Cowell and Stanney, 2005).

Table 2.1 Summarized facial features on trustworthiness

Static Features			**Dynamic Features**	**Emotions**
Internal Features	**External Features**	**Combinations**		
Eye size	fWHR	Babyface	Eye movement	Anger
Eye color	Brow-nose-chin	Masculine/feminine	Mouth movement	Sadness
Eye shape	Forehead-sellion-nose	Symmetry	Smile (Authentic/Fake)	Fear
Eye gaze	Hair	Look similar	Other movements	Happiness
Eyebrow	Forehead	Look typical		Disgust
Nose	Ears			
Mouth	Beard			
Lips	Chin			
Teeth	Glasses			
Cheek	Tattoo			
Color cue	Age			
Luminance contrast	Ethnicity			

2.3 Anthropomorphic Facial Features

2.3.1 Internal Facial Features

The eye region is believed to be one of the most crucial fields which could have an essential impact on the evaluation of facial trustworthiness, both for humans and robots (Ichikawa et al., 2011; Kaisler and Leder, 2016; Kleisner et al., 2013; Landwehr et al., 2011; Santos and Young, 2011; Windhager et al., 2010; Zebrowitz et al., 1996). Several specific features, such as eye size, shape, gaze, and color, as well as the eyebrows, could signal facial trustworthiness (Santos and Young, 2011; Todorov et al., 2008a). Research indicated that individuals could be believed to have a high level of trustworthiness with round eyes (compared to narrow eyes) (Ferstl et al., 2017; Masip et al., 2004) and larger eyes (compared with smaller eyes) (Kleisner et al., 2013; Linke et al., 2016) because those features both experience the evolution advantage from baby-appearance (Brownlow, 1992; Haselhuhn et al., 2013; Maoz, 2012). During the human physical growing process, maturation procedure and the force of gravity make a typical adult face, which is significantly different from a baby's appearance (Zebrowitz et al., 2013). Generally, a baby was considered as more vulnerable, innocent, and reliant than an adult (Montepare and Zebrowitz, 1998). It needs to have apparent signals, such as specific cues in the eye region, which could be instantly and automatically recognized and processed by adults to support their needs and survival (Masip et al., 2004). Specifically, compared with the growth process of other facial features, the eye region does not evolve significantly (eye size and eye shape grow slowly) (Montepare and Zebrowitz, 1998). Accordingly, a baby tends to have relatively round and big eyes, which becomes one of the most apparent symbols for baby-appearance looking. For example, large and round eyes could promote the perceived honesty, kindness, and trustworthiness (Berry, 1991) regardless of people's true age (Montepare and Zebrowitz, 1998). Moreover, both eye gaze and eyebrows might result in promoting one's trustworthiness since eye gaze and its associated eyebrows play a critical role in attracting attention for social recognition and interaction. To be more specific, numerous studies have proposed that a direct-gaze (compared with averted-gaze) looking while thin (compared with thick) and up-turned (vs. down-turned) inner ridge eyebrows could be expected to both trustworthy and attractive (Kaisler and Leder, 2016; Kompatsiari et al., 2019; Ma et al., 2015; Masip et al., 2004; Santos and Young, 2011; Todorov et al., 2008a; Xu et al., 2012). However, there exists a nuanced relationship between eye gaze and trustworthiness evaluation in the context of a social robot. For example, Stanton and Stevens (2017) indicated that a constant gaze (vs. an averted gaze) could stimulate signaling perceived dominance instead of trustworthiness. Furthermore, eye color is not a solitary characteristic but relevant to typical ethnic groups, which is originated from specific cultures (Stanley et al., 2011). For example, Kleisner and his colleagues inferred that, compared with blue eyes, brown eyes are believed to have a high level of trustworthiness (2013). Nevertheless, they additionally cleared that the observed difference of perceived trustworthiness could be associated with other implicit bias and misinterpretation.

Moreover, the area of nose and mouth is also believed to be an essential area that influences individuals' facial trustworthiness. Literature suggested the central area of the face, where nose and mouth lie in, works as a positive predictor for attracting one's attention and forming perceived trustworthiness (Okubo et al., 2013). Concerning the shape of the mouth, related research indicated its categorization in three kinds: an up-turned mouth, a down-turned mouth, and a neutral mouth (Landwehr et al., 2011). A noticeable distinction among these three kinds in signaling trustworthiness lie in, a face, whether human or robot, with an up-turned or neutral mouth (compared with a down-turned mouth), was perceived to have a high level of trustworthiness, warmness, and attractiveness (Arminjon et al., 2015; Kleisner et al., 2013; Landwehr et al., 2011; Maeng and Aggarwal, 2018). In addition, cheek, lips, and teeth might have a significant impact on facial trustworthiness: individuals with pronounced cheekbones, wide chins, thin lips, and complete front teeth, could be perceived to have a higher level of trustworthiness than those with shallow cheekbones, thin chins, full lips, and missing front teeth (Todorov et al., 2008a; Willis et al., 2008). As for the region of the nose, prior empirical evidence has shown the contrast effects of nose area on perceived trustworthiness: while a few works of the literature suggested that an individual with a small nose (compared with a big nose) could be believed to a low level of robustness and trustworthiness (Kleisner et al., 2013), the other research has suggested that individual with a short nose and shallow nose sellion (compared with a long nose) tended to have a high level of trustworthiness (Linke et al., 2016; Todorov et al., 2008a). In accordance with evolution psychology, individuals have a high level of intention to trust babies whose faces are usually featured with a pug nose (Masip et al., 2004; Xu et al., 2012).

Kleisner and his colleagues (2013) further explained this inconsistency in perceiving the area of the nose, suggesting individuals might tend to examine the characteristics of the nose and its relationship with other features in a whole, such as sellion, rather than examining feature separately (Xu et al., 2012).

With regard to color cues and luminance, different research has tried to explore their effect on related social judgments (Etcoff et al., 2011; Roberts et al., 2017). Indeed, a large number of pieces of literature have suggested that the perception of attractiveness is significantly associated with a different skin condition and the represented color (Fink et al., 2001). For evaluation of facial trustworthiness, scholars indicated that appropriate cosmetics (compared with plain) could promote complexion, brightness, and luminance, which in turn promoted social judgments, such as likability, trustworthiness (Etcoff et al., 2011), and health status (Roberts et al., 2017).

2.3.2 External Facial Features

External features, such as face shape and fWHR, work as an essential role in signaling facial trustworthiness. Within these features, fWHR might be the most obvious secondary sexual characteristics of humankind, which are also the most frequently explored its influence in social judgments in related research (Maeng and

Aggarwal, 2018). To specify, in the context of human perception, face with a large fWHR, compared with a small fWHR, is believed to have a high level of dominance, aggressiveness, unattractiveness, and untrustworthiness (Ferstl et al., 2017; Linke et al., 2016; Stirrat and Perrett, 2010). Nevertheless, in the context of human-robot interaction, the most counter-intuitive phenomenon might be a robotic face with a large fWHR that could enjoy more popularity because it plays a critical role in signaling one's social status, i.e. dominance (Maeng and Aggarwal, 2018). Correspondingly, other facial ratios, such as the brow-nose-chin and forehead-sellion-nose ratios, are proved to be negatively associated with perceived trustworthiness (Ma et al., 2015). Nevertheless, the association of facial ratios and facial anthropomorphic trustworthiness could differ in different scenarios. For instance, the ratio of brow-nose-chin did not have a significant effect on samples of different ages or genders (Ma et al., 2015).

Furthermore, other research has explored diverse features in signaling facial trustworthiness (Masip et al., 2004; Santos and Young, 2011; Xu et al., 2012). For instance, some literature (Masip et al., 2004) implied babies, who naturally have a relatively big forehead, short chin, and small ears could enjoy a high level of trustworthiness due to evolution psychology (Kleisner et al., 2013; Maeng and Aggarwal, 2018). Furthermore, Hellström and Tekle (1994) indicated that individuals wearing glasses (compared with no glasses) or having a beard (compared with no beard) were usually believed to have a high level of helpfulness and trustworthiness. Moreover, individuals with hair (compared with no hair or bald) or no facial tattoos (compared with having facial tattoos) might also significantly lead to more favorable social judgments, such as likability, trustworthiness, and leadership (Bakmazian, 2014; Funk and Todorov, 2013). Nevertheless, this phenomenon might depend on specific professions. On the one hand, salesman, normally believed to have hair and did not wear glasses, have been considered to have relatively less favorable judgments, such as believing to be untrustworthy, unintelligent, and suspicious, which might, in turn, decreases sales; on the other hand, individuals with education, i.e. academic researchers, normally believed to wear thick glasses and bald, were considered to have more favorable social judgments, such as trustworthiness, intelligence, and helpfulness (Bonnefon et al., 2013; Hellström and Tekle, 1994).

Age and ethnicity of the human face might also play a critical role in signaling facial trustworthiness (Cowell and Stanney, 2005; Masip et al., 2004; Roberts et al., 2017; Santos and Young, 2011; Xu et al., 2012). Prior literature has mentioned there might exist a U-shape association between perceived age and facial trustworthiness. To specify, a young-aged face (baby) and an old-aged face (old men) could experience a high level of facial trustworthiness when compared with a medium-age face (adults) as a result of the baby schema effect. The baby schema describes a stereotype that kids tend to be believed to be innocent and naive (Cowell and Stanney, 2005; Masip et al., 2004). Moreover, through the human evolution could unconsciously and continuously adapt individuals' visual features to strengthen or less heritable cues to affect social judgments, diverse ethnical groups (e.g., Asian vs. Caucasian (Etcoff et al., 2011) or Caucasian vs. African vs. Asian (Birkás et al., 2014)) indeed developed similar evaluation strategies to evaluate facial trustworthiness with no significant difference. Nevertheless, particular ethnical groups, such as Hungarians (Birkás et al., 2014; Cowell and Stanney, 2005), could be biased and have a more favorable attitude toward facial features with the characteristics of their own ethnicity (Stanley et al., 2011).

With regard to the nonhuman object, the face shape of an AI agent could have an impact on facial trustworthiness as well. From a broader perspective, there might generally exist two types of object shape: rounded shape and rectangular shape (Westerman et al., 2012). Though prior research has tried to examine the association between product shape and the associated judgments, there might still exist inconsistency on this relationship. For one point, based on the history of the robotic form, the rectangular form of the robotic face was believed to be a typical component in robotic design (Meeden and Blank, 2006). Furthermore, typicality in object form might assist individuals to have a better allocation and understanding of certain categories (Loken and Ward, 1990) and enhance its associated judgments (Blijlevens et al., 2012). For the other point, individuals tend to have an overall preference and positive attitude for rounded designs (Westerman et al., 2012). Thus, it would be theoretically interesting to explore, in the context of robot design, which form might have a better evaluation of trustworthiness.

2.3.3 Combinations of Facial Features

Among different combination facial features, a baby schema might play a key role in eliciting facial trustworthiness (Glocker et al., 2009a). To specify, baby schema (also called 'Kindchenschema') describes a phenomenon that a certain set of infantile physical characteristics, such as relatively large head, cute facial features, short extremities, and plump body, could induce spontaneous positive evaluations and related behavioral reactions in the human, such as perceived cuteness and associated caretaking or nurturing behavior for babies (Lorenz, 1943). Accordingly to the evolution theory, such a phenomenon could provide adaptive benefits for humans, eliciting their caring behavior and the survival rate of offspring (Glocker et al., 2009a). Thus, spontaneous reactions to baby schema might work as the basic foundations of human social cognition.

The latest neuroscientific studies have further explained this altruistic instinct by illustrating the crucial role of the brain reward system when exposing to babies or even cute faces (Venturoso et al., 2019). Specifically, based on the observation from functional magnetic resonance imaging (fMRI), Glocker and his colleagues (2009b) suggested baby schema could trigger the essential area for reward processing, the nucleus accumbens, in nulliparous women. In addition, similar to fixation patterns when processing adult faces, people also could recognize and differentiate baby faces from adults' faces in a limited amount of time, confirming the automatic and immediate facial processing pattern in infants (Brosch et al., 2007).

Regarding the facial features in the baby schema, two basic regions, eye and mouth, are playing a significant role in attracting attention and communicating trustworthiness (Penton-Voak et al., 2001). Facial babyishness typically tended to have large eyes, a high brow ridge, a small chin, a pug nose, short ears, and thin lips (Ma et al., 2015; Maeng and Aggarwal, 2018; Masip et al., 2004). As illustrated in Figure 2.1, the typical face proportion contains dimensions in three perspectives: the upper face length (forehead length), inter-pupillary distance (eye distance), and lower face length (chin length) (Hildebrandt and Fitzgerald, 1979).

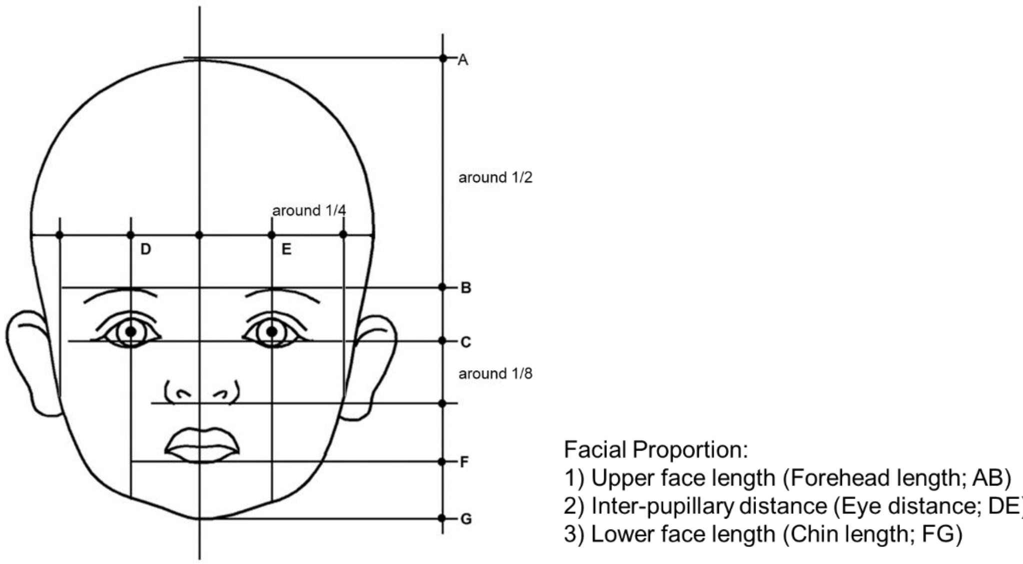


Figure 2.1 An illustration of the measurement parameters in an infant facial proportion

To be more specific, as for the size of the eye and mouth, infants (normally under one-year-old by definition) typically have round and large eyes (Ferstl et al., 2017; Masip et al., 2004). That is to say, individuals with round and large eyes would enjoy a high level of trustworthiness. However, people tend to have a nuanced preference toward mouth size: while a baby usually has a small mouth (Glocker et al., 2009b), people with big mouths might also be considered as more trustworthy and capable (Re and Rule, 2016). With regard to the position of eye and mouth, infant facial features tend to have a centralization (inward) positioning tendency (Lee, 2013; Miesler, 2011): babies normally have a higher forehead (namely, the lower vertical position of eyes; larger AC or AB) (Gorn et al., 2008; Miesler, 2011;

Miesler et al., 2011), closer pupils (namely, closer horizontal position of eyes; smaller DE) (Venturoso et al., 2019), and a more centralized chin (namely, higher vertical position of mouth; larger FG) (Zebrowitz and Montepare, 2008).

Similar to the facial babyishness, a feminine face (compared with a masculine face) could also be believed to be a high level of trustworthiness due to their relatively larger eyes while smaller eye distance (Buckingham et al., 2006; Todorov et al., 2015). Accordingly, individuals usually consider a male face to have a high level of dominance and a low level of cooperation and honesty, compared with a female face (Dzhelyova et al., 2012; Kleisner et al., 2013).

Moreover, individuals tend to have a high evaluation of facial trustworthiness for the ones who shared similar facial features with the evaluator, the ones who have cultural typicality in facial features, the ones whom the evaluator met before, and the ones who a high degree of facial symmetries (Farmer et al., 2014; Kocsor and Bereczkei, 2017; Okubo et al., 2013; Sofer et al., 2015; Verberne et al., 2015; Zebrowitz et al., 1996). The explanations accounting for people's preference for similarity and typicality might lie in the role of familiarity, which, in turn, have a more favorable trustworthiness perception (Farmer et al., 2014; Okubo et al., 2013; Sofer et al., 2015). In addition, interacting with relevant social actors might have an impact on our expectation of the general facial model, assisting the unfamiliar facial signal processing, which is usually based on the experience of daily lives (Kocsor and Bereczkei, 2017). Moreover, previous literature also discussed the relationship between facial symmetry and perceived trustworthiness from the perspective of evolution: Facial symmetries might work as a significant signal for communicating physically attractive and healthy (Zebrowitz et al., 1996).

2.3.4 Dynamic Features, Emotional Expressions, and Associated Contexts

In order to have a more comprehensive understanding of emotional expressions, emotions and their associated facial regions/features should be emphasized. Russell (1980) has identified two fundamental constructs, valence and arousal, and further proposed the 'Circumplex Model of Affect' (CMA) to explain the involved mechanism and process. To specify, CMA works as a framework to categorize, position, and summarize emotion terms in orthogonal axes of a plane. While valence serves as the horizontal axis (negative emotions position at the left and positive emotions position at the right), arousal serves as the vertical axis (inactive emotions position at the bottom and active emotions position at the top). Different emotion terms, such as exited or depressed, were mapped in a circumplex shape while the center of the plane indicates a neutral emotion with a medium level of valence and arousal. Accordingly, given emotional states could be depicted with specific configurations of valence and arousal (see Figure 2.2).

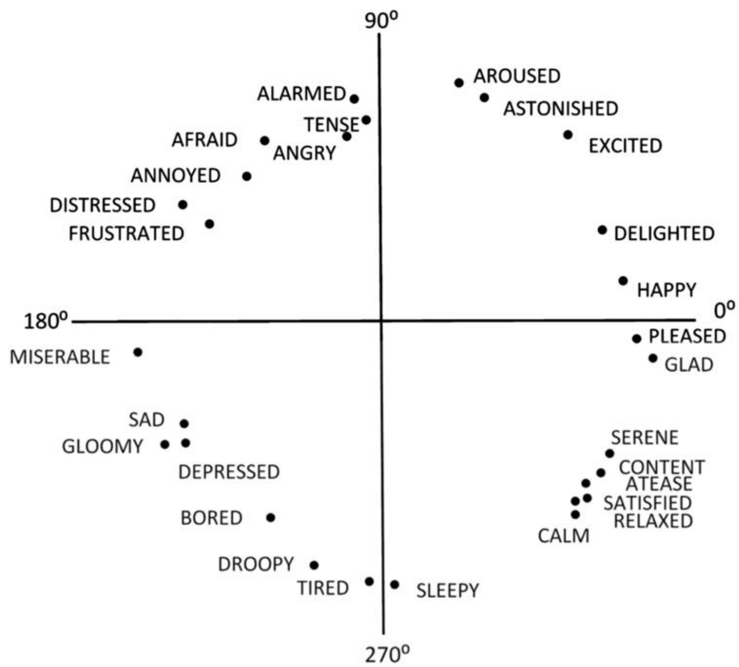


Figure 2.2 Circumplex model of affects (J. A. Russell, 1980)

A previous study has indicated four typical emotions positioned at 45 degrees of each quarter in the model (Jaeger et al., 2019): excited emotion locates at the top-right of the circumplex model with a high level of valence and arousal; afraid emotion locates at the top-left with a low level of valence and a high level of arousal; depressed emotion locates at the bottom-left with a low level of valence and arousal; relaxed emotion locates at the bottom-right of the model with a high level of valence and a low level of arousal (Aguado et al., 2011). Regarding dynamic features signaling these emotional expressions, they mainly concentrate on two facial areas:

the eye/brow and mouth region (Calvo et al., 2019, 2017; Gutiérrez-García and Calvo, 2016a). To specify, positive emotions, excited or relaxed expressions, is usually featured with an upturned U-shaped mouth while lip corners and eyebrows are raised while negative emotions, depressed, or afraid expressions, are characterized by an inverted U-shaped mouth while lip corners and eyebrows are lowered (Calvo et al., 2017; Gill et al., 2014; Johnston et al., 2010; Ma et al., 2015).

Indeed, different emotional expressions have particular social-functional values which could assist human to communicate information in interpersonal interactions (Oosterhof and Todorov, 2008). Numerous traditional studies tried to analyze the specific meaning and effect of facial expressions (Galinsky et al., 2020; Kocsor et al., 2019; Landwehr et al., 2011; Liu et al., 2013; Vesker et al., 2018; Vrticka et al., 2012). Among those research, Darwin (1872) was the first to introduce and summarize the adaptation and evolution of facial expressions over time. To be more specific, humans tend to, to some extent, share universality across different cultures, ages, and gender in emotional detection and recognition, which works as the basis for instant social inferences toward others (Steinel et al., 2008). A recent neuroscientific study has also provided evidence that brain areas that are mainly involved with impression and judgment formation, such as the amygdala, medial prefrontal cortex (mPFC), and superior temporal sulcus (STS), would experience a significant activation and stimulation when exposing to emotional expressions (Ames et al., 2010). For example, positive expressions, such as smiles, which are featured by the use of the orbicularis oculi (enlargement of the pupil and the muscle surrounding the eyes) (Q. Liu et al., 2019) in combination with the zygomatic major (raising the corners of the mouth) (Calvo et al., 2019), are considered to be evolved to facilitate partnership and collaboration among people (Calvo et al., 2017).

A large number of prior research has tried to unveil the magic of smiles: smiles indeed assists to signal good intent and to build a friendly relationship, which could facilitate interpersonal communication, cooperation, and interpersonal rapport (Engell et al., 2010; Hall et al., 2010; Oosterhof and Todorov, 2009; Sofer et al., 2017; Todorov et al., 2015, 2008a, 2008b, 2005; Willis and Todorov, 2006). Though the positive emotions are equally shared and immediately recognized by individuals regardless of their races, cultures, and ethnicity, the subtle difference in various kinds of positive representations still could be perceived and noticed. For example, people might find it easy to identify and differentiate various kinds of smiles, such as authentic smiles and fake smiles while individuals tend to have a significantly favorable attitude to collaborate with others with authentic smiles (Johnston et al., 2010; Krumhuber et al., 2007). Additionally, positive expressions are strongly associated with perceived trustworthiness and enjoy more visual attention (Calvo et al., 2017; Dunn and Schweitzer, 2005; Pavlov et al., 2015; Sanchez et al., 2014). For example, Calvo and his colleagues (2019) mentioned, “facial happiness (i.e., an expresser’s smiling face) is significantly related to the perception of trustworthiness by observers. People showing happy expressions are judged as more trustworthy than those with non-happy faces (while facial anger is perceived as untrustworthy)" while Sancheza and Vazquez (2014) used experiments to indicate, "both the emotional and cognitive components of subjective well-being were related to a general bias to attend to happy faces and avoid sad faces". However, few studies have explicitly explored the effect of facial arousal and found a nuanced relation between facial trustworthiness and visual attention. On the one hand, the majority of research on facial expressions treated facial emotions either separately as different types, such as smile and anger (Chiller-Glaus et al., 2011; Karbauskaite et al., 2020; Sanchez and Vazquez, 2014; Weiβ et al., 2019), or within different facial valences, such as positive, neutral, and negative (Calvo et al., 2019; Gantiva et al., 2019; Lane et al., 1999). Considering different types of emotions might have their specific coordinates in the circumplex model of affects (J. A. Russell, 1980), it might be probably to infer the effect of arousal on facial trustworthiness and visual attention. For instance, facial sadness and anger might be paired negative emotions with different arousal where sadness might be believed to be a negative emotion with inactive arousal and anger might be considered as a negative emotion with active arousal by definitions (Marriam-Webster, 2002). However, prior research has suggested no significant difference between these two emotions in communicating trustworthiness and attracting visual attention (Okubo et al., 2018; Sanchez and Vazquez, 2014). On the other hand, the intensity of facial expressions might influence visual attention. For example, Ngan and Yu (2019) conducted an eye-tracking experiment to illustrate the longest fixations occurred on smiles with teeth showing (a more intense smile), compared with a smile without teeth showing (a less intense smile). Thus, it might be possible that facial arousal could play a significant role in signaling facial trustworthiness and visual attention.

Moreover, dynamic emotional expressions did not exist alone but were embedded in daily contexts with different regulatory focus (Bargh and Shalev, 2012). According to regulatory focus theory (RFT), it divides contexts into two categories: promotion-focused or prevention-focused to guide people's behavioral regulation (Aaker and Lee, 2006). To specify, daily contexts with different regulatory focus differ in their strategic tendency for obtaining inclined states. When exposing to promotion-focused contexts (promotion foci), people are activated to be inclined toward the need for improvement and achievement and seek goals and dreams by promptly striving towards them. The promotion system triggered by a promotion focus could be considered as continuous striving to be successful through taking as many opportunities as possible to exploit action for positive outcomes. When exposing to prevention-focused (prevention foci), people are activated to be inclined toward the need for security and safety and pursue the fulfillment of responsibilities by strategically avoiding possible failure or risks in goal attainment. The prevention system elicited by a prevention focus goes for success by acting vigilantly to avoid errors of commission (Ewe et al., 2018). Regarding both foci vary situationally as well as chronically between individuals, a large number of studies have confirmed the regulatory focus theory: the contextual cues, such as the framing of a reward-pursuing or risk-prevention priming of contextual messages, could influence people's situational regulatory focus (Avnet and Higgins, 2006; Crowe and Higgins, 1997; Higgins, 2000)

Further, recent neuroscientific evidence from fMRI measures has shown that the overlapping of the human amygdala to facilitate the processing of facial valence with gains and losses during decision-making where positive facial expression is associated with gains and negative facial expression is associated with losses (Gupta et al., 2016). Thus, in daily contexts, individuals might have a more favorable attitude and behavioral intention when they are exposed to an emotional expression and a message that are congruent with their regulatory focus (Ludolph and Schulz, 2015; Nabi et al., 2020). This situational combination, an appropriate match between the people's situational regulatory state and the emotional expressions, makes a regulatory fit (Aaker and Lee, 2006; Avnet and Higgins, 2006).

When there is a fit between regulatory focus and emotional expression, people could feel appropriate and right, which makes intuitive judgments based on the feeling, and have a more positive attitude toward the information sender (Crowe and Higgins, 1997; Higgins, 2000, 1998). That is, the information sender might be considered as more trustworthy and the contextual message might be considered more persuasive when a promotion-focused (prevention-focused) contextual message is presented in positive (negative) facial expressions. For example, Jin (2010) indicated the regulatory fit could enhance people's perceived trustworthiness toward the information sender, which, in turn, increases the perceived informational and educational values in the context of health messages spread. Considering different regulatory focuses could influence people's processing of nonverbal cues, such as emotional expressions (Sassenrath et al., 2014), the regulatory fit between nonverbal cues, such as emotional expressions, and an individual's self-regulatory orientation could increase the effectiveness of the message in a given context (Bosmans and Baumgartner, 2005). Therefore, a regulatory fit plays a crucial role in enhancing the trustworthiness of the information sender in a certain situation.

2.4 Summary and Research Gap

Drawing intuitions of particular features in the contexts of the human face, appearance for the product, and facial anthropomorphic features for a robot, chapter two tried to comprehensively summarize facial features from various perspectives, which provide sparks for additional design implication for a social robot. However, the problem may still exist considering the meaning of facial anthropomorphic trustworthiness for social robots and the way to merge various facial features to design an appropriate face for a social robot.

Regarding the nature of abstract illustration of a rendered presentational face for the social robot (Dehn and Van Mulken, 2000), merely perfecting individual feature then merging them as a whole cannot guarantee the overall trustworthiness evaluation, which might take risk of creating a “Frankenstein-like” appearance for the robot (Jentsch, 1997; Spotts, et al., 2004). In order to address this problem, salient facial features, regions, and balance should be emphasized for social robot design (Engell et al., 2010). Together with the current technological levels in the rendered face (Kalegina et al., 2018) and universality in robot design (Kiesler et al., 2018; Walters et al., 2008), the most promising facial features are given priority while other factors are controlled in this thesis. For example, following the ethical guidelines of the robot and AI (Torresen, 2018; Winfield, 2019), robot design should avoid any potential racism and ethnical related issues. Accordingly, eye color, color cue, luminance contrast, tattoo, and ethnicity were controlled in the thesis. In addition, facial features should also follow the guidelines of on-screen animation characters (Chen et al., 2010; Dehn and Van Mulken, 2000; Kalegina et al., 2018; Luo et al., 2006), nose, eye gaze, eyebrow, and other features were not appropriate for animation representation (Luo et al., 2006) and rarely adopted in actual practice (Blue Frog Robotics, 2018; CES, 2018). Accordingly, based on the literature mentioned before, four essential research gaps are identified:

- A clear definition of facial anthropomorphic trustworthiness for social robot and its sub-dimensions are still missing in the current literature (Objective 1 and RQ 1).

Considering the significant role of trustworthiness in HRI, facial anthropomorphic trustworthiness indeed plays a crucial role in building the initial credibility and the approaching intention at a later stage (Hoorn, 2015). Within the domain of social robots, it lacks a valid scale to measure the construct of interest: facial anthropomorphic trustworthiness. A reliable and valid measurement to assess facial anthropomorphic trustworthiness is necessary for the context of social robots, assisting empirical studies in HRI.

- A clear causal relationship between separate and combination of essential static features (internal features, external features, and their combinations) and facial anthropomorphic trustworthiness has not been clarified (Objective 2 and RQ 2). Particularly,

For the internal features, the shape of eye and mouth for social robots could be the most essential features (Kaisler and Leder, 2016; Kompatsiari et al., 2019; Ma et al., 2015; Masip et al., 2004; Santos and Young, 2011; Todorov et al., 2008a; Xu et al., 2012) because the eyes might work as a salient facial factor to catch one's attention while its shape is significantly related to human evolution (Haselhuhn et al., 2013; Maoz, 2012). Similarly, as one of the central facial properties, the shape of mouth (upturned, neutral, and downturned) is also perceived to be a significant feature that has an impact on people's evaluation of trustworthiness (Landwehr et al., 2011).

With regard to external features, face ratios (e.g., fWHR) are the most obvious and essential secondary sexual characteristics in human evaluation (Maeng and Aggarwal, 2018) and one of the most noticeable characteristics during human-robot interaction (Kramer, 2015); moreover, robots' face shape would also significantly influence their anthropomorphic trustworthiness, especially considering people's ambiguous attitude toward typicality and rounded shape preference (Blijlevens et al., 2012). For example, literature regarding the influence of fWHR in signaling facial trustworthiness have indicated that humans featured with a high level of fWHR (compared with a low level of fWHR) are believed to have a low level of trustworthiness and attractiveness (Ferstl et al., 2017; Stirrat and Perrett, 2010). Nevertheless, Maeng and Aggarwal implied that social robots featured with a high level of fWHR (compared with a low level of fWHR) are indeed enjoying more popularity and likability. Thus, it is necessary to further explore the effect of both fWHR and face shape on facial andromorphic trustworthiness.

For a combination of static features, the baby schema, which enjoys the advantage of the evolution in interpersonal interaction, is featured by the belief of youthfulness and naiveness. It might play a critical role in signaling facial trustworthiness (Miesler et al., 2011). Therefore, the combination of feature size and feature position might have a considerable impact on facial anthropomorphic trustworthiness. For instance, Borgi and his colleagues (2014) suggested that robots with babyish features might enjoy higher popularity and that people indeed are inclined to wish to see such robot designs.

However, in the context of a social robot, empirical evidence on the relationship between internal/external/combinations of facial features and anthropomorphic trustworthiness has not been addressed before, which is a research gap in the current study.

- A clear causal relationship between essential dynamic features and facial anthropomorphic trustworthiness in different contexts remains ambiguous (Objective 3 and RQ 3).

Dynamic facial features, which mainly involve the movements of the eye and mouth region, are the most obvious characteristics in emotional expressions (Menne and Schwab, 2018). Considering that emotions can be categorized into two dimensions, namely valence and arousal (Zhu et al., 2018), its effect on facial anthropomorphic trustworthiness is still unclear which needs empirical analysis. Further, emotional expressions did not exist alone but were embedded in daily contexts with different regulatory focus (promotion-focused or prevention-focused) (Ludolph and Schulz, 2015). According to regulatory fit theory (Aaker and Lee, 2006), positive emotion fits the promotion-focused context while negative emotion fits the prevention-focused context. However, the current study still lacks empirical evidence to assess this relationship in the context of the social robot.

- Detailed practical design guidelines for making a trustworthy-looking robot are still missing in the current literature (Objective 4).

In order to fill the research gaps, a series of studies was developed to address the research questions. This research explored the meaning of facial anthropomorphic trustworthiness and its associated sub-dimensions. Then, three behavioral experiments were introduced to empirically investigate the effect of static features on facial anthropomorphic trustworthiness. Based on the results of experiments, a static features model was developed and validated. Similarly, another lab experiment was conducted to inspect the effect of dynamic features on facial anthropomorphic trustworthiness under different regulatory-focused contexts. Based on this result, a model for dynamic features was also developed. Last, detailed design guidelines were summarized according to the results above. The methodology to address each research question is discussed in the next chapter.

CHAPTER THREE. RESEARCH METHODOLOGY

This chapter discusses the chosen research methodology, to be more specific, describing the mixed methods used in different studies.

3.1 General Methodology Description

To achieve this study's objectives, we had to use scientific design methods to narrow down the RQs. Accordingly, mixed-method research was conducted to address the RQs.

Since the current research involves theories from human trustworthiness, anthropomorphism, and robot design. Each theory secures the mainstream approach for analysis. For example, human facial trustworthiness has mainly been examined using advanced physical apparatus (e.g., eye tracker and electrodermal activity) in experiments to examine the effect of particular facial features on individuals' evaluation (Samal and Iyengar, 1992; Tranel and Damasio, 1985; Warrington and James, 1967). Anthropomorphism research has mainly relied on the theory about related social science, examining RQs from the perspective of human psychology (Kurniawati, 2017; van Rompay et al., 2009; Van Rompay and Pruyn, 2011). Robot design literature separates into two fields: the first stream might base on the engineering theory to analyze the structure and function of robots from the perspective of ergonomics (Dul et al., 2012; Karwowski, 2005; Stanton et al., 2006), whereas the second stream has been more interested in user reaction (Bellotti and Bly, 1996; Falck and Rosenqvist, 2012; Green et al., 1981). In this way, regarding the multidisciplinary nature of facial anthropomorphic trustworthiness, a mixed methodology, which utilizes both quantitative and qualitative methods, is introduced to address the RQs.

Moreover, the academic orientation of research might be in accordance with the particular paradigm which fits the related RQs (Johnson and Onwuegbuzie, 2004). With regard to the influential factors of facial anthropomorphic trustworthiness, a pragmatic method might be required due to its minimal requirements of speculations. As a result, it might give the researchers a large degree of freedom to conduct a particular investigation that attracts their interest. Furthermore, pragmatism naturally explores the research question like “what works”, to examine particular features signaling facial anthropomorphic trustworthiness (Collins, 1990; Johnson and Onwuegbuzie, 2007). Furthermore, data triangulation might work as a critical approach for exploring and validating the same scenario from various disciplines (Lougen, 2009). Thus, seven studies in three phases were planned to address different research objectives, of which the details are explained later (see chapters 4-9).

3.1.1 Qualitative Method: Diary Study via Crowdsourcing

Before launching a new robot in the market, user experience teams seek to utilize traditional qualitative methods, such as semi-structured interviews or focus groups, to explore insights into a user experience that reflect their opinions toward the given robot or service (Lee and Lee, 2009). While semi-structured interviews can be conducted to obtain descriptive messages on people's consideration, recognition, and behavior, as well as its underlying mechanism, focus groups are instructive for gathering information from the pre-decided sample, who are invited to a discussion planned on a given subject (Lederman, 1990).

Based on the collected data, trained researchers screen the transcripts, remove redundancies, and form the initial corpus for further review. Through manual coding and clustering, the statements of the corpus would be grouped toward certain themes. Sometimes, the generated themes can even have a hierarchical relationship (Jiao et al., 2006). Finally, based on the extracted themes, researchers can identify and summarize the user experience.

Although the traditional method has enjoyed sufficient reliability and validity in exploring user experience, it still might face some problems, such as having a relatively small sample (Bhattacherjee, 2002) and being time-consuming and expensive (Ratislavová and Ratislav, 2014). To address these problems, the latest qualitative research has attempted to utilize a diary study to collect qualitative data.

Compared with interviews and focus groups, a diary study has several advantages. First of all, a diary study could gather deep insight into the subject at hand easily and provides more flexibility for people to show their ideas (Frey and Fontana, 1991). Facial anthropomorphic trustworthiness is indeed an evaluation of the specific object based on their own experience, and thus, we must obtain a thoughtful insight into the opinions they really hold (Lederman, 1990). However, the duration of the interview or focus group for one participant could be extended longer than the time spent for others, which could make the result less orientated and clear, leading to bias in the results (Frey and Fontana, 1991). Furthermore, interviews and focus groups have a higher probability of facing moderator bias. When the moderator conducts an interview or focuses group to gather qualitative inputs, this process might have an impact on the quality of data in an unconscious manner (Frey and Fontana, 1991). For instance, the attributes of the moderator, such as age, gender, voice, overall appearance, wearing, tone, or even speed could lead to systematic influences, thus resulting in moderator bias.

Furthermore, the latest diary research has attempted to collect data from online sources, such as user-generated content (UGC) and crowdsourcing platforms (Boynton and Richman, 2014; Stone et al., 2020). Although UGC is a data source, its quality varies widely, and it mainly depends on a specific website and cannot reach particular populations outside the given theme on the website (Johnson et al., 2012). Different from UGC, qualitative data can be actively collected from crowdsourcing platforms, which might enjoy similar quality compared with traditional methods in various disciplines (Goucher-Lambert and Cagan, 2019). Specifically, crowdsourcing, a combination of the words “crowd” and “outsourcing,” refers to the distribution of microtasks to others online (Shank, 2016). Based on personalized services from workers, crowdsourcing is very helpful since it can provide diverse, innovative, and numerous solutions (Lovett et al., 2018).

Among all the crowdsourcing platforms, Amazon Mechanical Turk (hereinafter “MTurk”) is the largest and most popular, having more than 800,000 registered workers (commonly called “mturkers”) across more than 200 countries with more than 600,000 microtasks (so-called Human Intelligence Tasks [HITs]) a day (AMT, 2021). Considering the number, age, diversified locations, and affiliations of registered MTurkers, MTurk has become an efficient tool and a reliable data collection source for scientific research, ranging from recruiting behavioral experiment participants (Song and Luximon, 2019) and labeling specific image or video (Khare et al., 2015) to retrieving consumer insight (Lutz and Newlands, 2018). Much research has suggested the data on MTurk exhibit adequate quality and reliability even compared with physical lab experiments (Lovett et al., 2018).

3.1.2 Quantitative Method: Experiment

The experiment is a frequently used quantitative approach that serves as a mature analytical tool to examine a particular hypothesis (Kruglanski and Higgins, 2008). It is especially beneficial to evaluate the causality association between various factors via controlling other factors for the same (Fox and Denzin, 1979). Accordingly, it is natural that the experiment fits to examine the particular influence of specific facial features on trustworthiness evaluation. For instance, the area of the eyes is a typical characteristic in facial trustworthiness evaluation. When adding or modifying eye-related features in robot design, other relevant cofounding factors must be kept the same to exclude their influences. Thus, the causal association between eye-related features and their influence on facial anthropomorphic trustworthiness can clearly be evaluated and determined.

Experiments could have different categories based on their various attributes. For example, online experiments, lab experiments, and field experiments are the most frequently used methods in the context of social science. Among the three kinds, online experiments and lab experiments might be appropriate solutions to control the confounding effects while maintaining internal validity. To be more specific, an online experiment might be more could be generalizable since its sampling could reach a large population. Furthermore, with help from the particular crowdsourcing platform, i.e. Amazon Mechanical Turk (MTurk), the online experiment might recruit participants to a large distribution while checking their attention during the experiment (Babbie, 2008). MTurk is a web-based platform that can help to recruit registered workers to finish given work for compensation (Mortensen and Hughes, 2018). Numerous research studies in various disciplines have been conducted using this platform because it is reliable (Deal et al., 2016), accurate (Song et al., 2020), effective (Harber and Leroy, 2015), and diverse (Mortensen and Hughes, 2018). Thus, we considered that it might be appropriate and reliable for recruiting subjects to obtain an enhanced understanding of trustworthiness and attitudes toward the social robot (Song and Luximon, 2021). In addition, a lab experiment would be especially suitable for conducting data triangulation when examining dynamic features in the communication of trustworthiness (Hedt and Pagano, 2011; Kam et al., 2007). By making related dynamic expressions and introducing them in a prototype, a lab experiment could help to measure both subjective rating and objective measures, such as eye-tracking technique and electrodermal activity, in a controlled environment (Hedt and Pagano, 2011; Kam et al., 2007).

3.2 Summary of the Methodology Framework

To comprehend the meaning of facial anthropomorphic trustworthiness and examine the effect of different facial features on anthropomorphic trustworthiness, a methodology framework was developed, which is presented in Figure 3.1.

Following a mixed-methods design (qualitative exploration first followed by quantitative examination), this methodology framework contained three phases to address RQs 1–3. Specifically, in phase 1, study 1 was conducted to address RQ1 for a conceptual exploration of facial anthropomorphic trustworthiness, which worked as a foundation for phases 2 and 3. In phase 2, studies 2–5 were conducted to empirically examine and model the effect of static features on facial anthropomorphic trustworthiness. In phase 3, studies 6 and 7 were conducted to empirically analyze the effect of dynamic features and their interaction with different daily contexts on facial anthropomorphic trustworthiness.

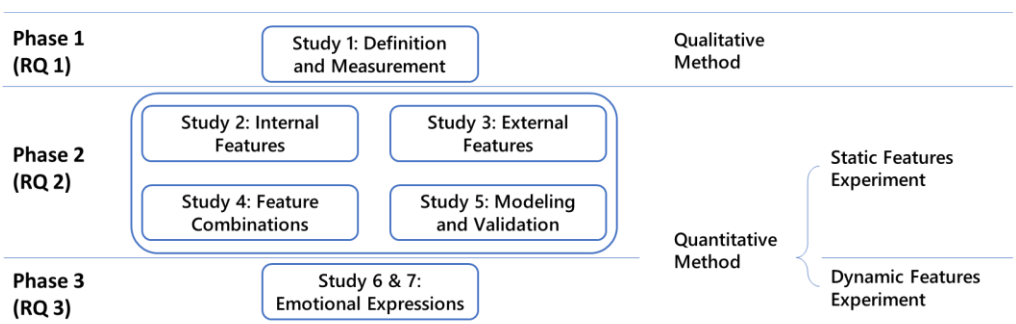


Figure 3.1 Overview of the methodology framework

3.3 Analysis Technique

3.3.1 Content Analysis and Natural Language Processing

Qualitative data were analyzed through content analysis and natural language processing. Content analysis is a structural and reliable reduction approach for summarizing qualitative data from the bulk of a corpus (Neuendorf, 2020), whereas natural language processing is an advanced computational technique for extracting and analyzing qualitative data (Deng, 2014). Although content analysis works as an efficient method of compiling main ideas in a structural manner, it might be time-consuming and the sample size might be relatively small (Frey and Fontana, 1991). Considering the efficiency and efficacy of processing qualitative data, natural language processing is becoming increasingly popular in the latest research (Timoshenko and Hauser, 2019). Accordingly, it might be beneficial to consolidate these two methods to obtain an enhanced understanding of the meaning of facial anthropomorphic trustworthiness.

Semi-structured interviews were adopted as a pilot study in study 1 to explore people's preliminary understanding of facial anthropomorphic trustworthiness when exposed to a social robot. Then, a diary study through crowdsourcing collected numerous qualitative data to probe people's comprehension of facial anthropomorphic trustworthiness toward a social robot. Natural language processing was then utilized to analyze, categorize, and summarize the rich corpus regarding anthropomorphic trustworthiness. Therefore, significant dimensions of this concept were revealed and organized into related constructs and topics.

Through other quantitative validations, relatively thorough insights into this concept were illustrated.

3.3.2 Experimental Analysis and Behavioral Modeling

The experiment results were measured according to subjective ratings on a trustworthiness scale. Participants were required to indicate their agreement with statements from strongly disagree to strongly agree (Likert Scale: 1–9) when exposed to the given robot stimuli.

All quantitative data were analyzed in SPSS. An alpha level of .05 was used for the statistical analysis. Descriptive data analysis was conducted for the factors of user characteristics, such as gender (male or female), age (in years), robot use experience (never, 0–1 year [1 year not included], 1–2 years [2 years not included], 2+ years), and educational level (high school graduate or lower, some college education, college graduate or above), including mean, median, standard deviation (SD), and frequency. Pearson and Spearman correlation analyses were used to analyze the possible relationships between user characteristics and navigation behavior. In addition, analysis of variance (ANOVA; between-subject, within-subject, and mixed design) was used to evaluate the interaction effects of the variables on dependent variables. Furthermore, a generalized linear regression analysis was used to model the relationship between different facial features and anthropomorphic trustworthiness.

Lastly, reliability, which reflects the degree of repeatability or consistency of the findings by other literature, was also examined in the subjective ratings. For example, when conducting a survey or distributing questionnaires, Cronbach's alpha works as a significant coefficient for reliability, calculating the correlation of the items within a particular dimension. As suggested in prior literature (Johnson and Onwuegbuzie, 2004), if the Cronbach's alpha value is larger than 0.7, it could be considered as having statistically adequate reliability for evaluating a given dimension.

3.4 Summary of Chapter Three

To sum up, regarding the research questions and the context of facial anthropomorphic trustworthiness, the mixed method works as a good solution to address the given questions. Under the guidance of pragmatism, the data triangulation, from qualitative and quantitative data, helps to confirm the effect of facial anthropomorphic trustworthiness.

To specify, phases 1-3 are according to address different research questions and related research objectives. Phase 1 mainly focuses to analyze and understand the meaning of facial anthropomorphic trustworthiness via a hybrid method while phase 2-3, with the related finding from study 1, would tend to analyze and model the effect of static and dynamic features on anthropomorphic trustworthiness evaluation.

Within different phases, appropriate protocols and advanced analysis methods would be developed and performed for studies 1-6. Besides, internal validity, external validity, and reliability would also be concerned and checked during the analysis since they would help to control confounding variables and replicability.

CHAPTER FOUR. EXPLORING FACIAL ANTHROPOMORPHIC TRUSTWORTHINESS

This chapter explores the meaning and sub-dimensions of facial anthropomorphic trustworthiness. Through a hybrid mixed method, this chapter innovatively collects qualitative data from a crowdsourcing platform, applies the state-of-art natural language processing technique to form the scale of facial anthropomorphic trustworthiness, and performs iterative factor analysis to validate the scale, contributing to a better understanding of facial anthropomorphic trustworthiness.

4.1 Introduction

When considering the relationship between human and robot trustworthiness, it is natural that facial anthropomorphic might share some common dimensions of interpersonal trustworthiness (Atkinson et al., 2012). Mayer and his colleagues (1995) suggested that interpersonal trustworthiness could have three dimensions: ability, benevolence, and integrity. These dimensions are theoretically distinct from each other because they could be classified into the cognitive and affective attributions of trustworthiness (Bhattacherjee, 2002; Song et al., 2020). Although these dimensions could provide a theoretical basis for us to get a relatively comprehensive understanding of trustworthiness evaluation, they are parsimonious to some extent that other dimensions of trustworthiness literature could be reconciled and complemented to the original dimensions, especially in the social robot research.

In addition to the potential dimension of facial trustworthiness of anthropomorphic robots, there are specific and significant characteristics that are unique for robot design: the uncanny valley (Appel et al., 2020). It suggests to a situation the relationship between people's emotional reaction towards a robot and the degree of a robot resemblance to humans does not follow a linear pattern: people would feel unsettling when meeting a robot with imperfect human traits, especially in facial features (Ho and MacDorman, 2017). Accordingly, it is possible that people might prefer and trust a robot with appropriate facial features, which are neither too robotic nor too humanoid. Nevertheless, its related evidence is missing to some degree which needs further exploration.

4.2 Method and Results

To develop the scale of facial anthropomorphic trustworthiness for social robots, we used content analysis, data collection via crowdsourcing, natural language processing, exploratory factor analysis (EFA), and confirmative factor analysis

(CFA) to address our research aim. Specifically, based on the literature review, we firstly ran a pilot study (interview) to provide preliminary evidence of various constructs of facial anthropomorphic trustworthiness and further confirmed the validity of the questions used in the second step. Then, we distributed the questionnaire via a crowdsourcing platform to collect a relatively large sample of qualitative data. With the help of the state-of-art (SOTA) natural language processing technique, the qualitative data were clustered into several clusters. Together with the items retrieved from the literature, we formed the initial item pool. Last, we followed the standard scale development procedure to form the items pool and further examined and validated the validity of our scale via EFA, CFA, etc (see Figure 4.1). This Ph.D. study was approved by the Research Committee of Hong Kong Polytechnic University (see Appendix. B).

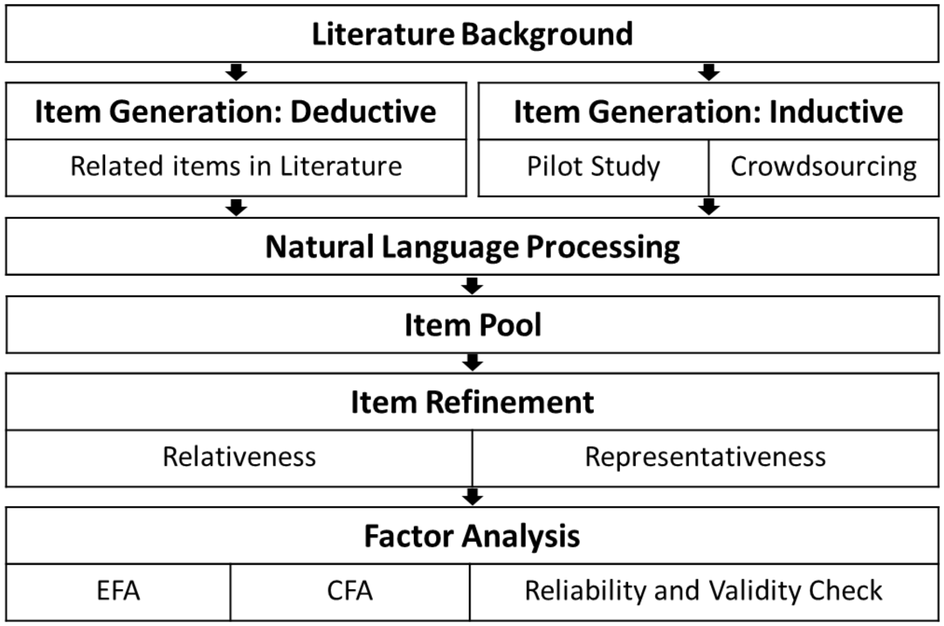


Figure 4.1 Overview of the research model of study 1

4.2.1 Item Generation via a Hybrid Method

Before the main study, we initially conducted a pilot study via a convenience sample of ten participants (mean age = 31.5; 5 Chinese and 5 non-Chinese; 2 males and 8 females). Participants were first informed about the aim and were instructed to see a set of 80 robot faces (Mathur and Reichling, 2016). The interviewer explored their experience toward the set of robot face and the reasons for a robot to look trustworthy or untrustworthy. After fully transcribing and manually coding, the content analysis showed: 1) participants agreed that they have an unconsciously facial trustworthiness evaluation on a social robot at first sight; 2) four themes ("capability", "positive affect", "ethics concern", and "anthropomorphism") emerged and were partially consistent with previous literature (Mayer et al., 1995). Thus, a deep-learning-based theme generation was adopted to a larger sample to confirm the finding in the pilot study.

For the main study, the scale development process starts by generating an item pool for further exploration. This process could be conducted by two approaches: a deductive approach (e.g. retrieving items from previous literature) and an inductive approach (e.g. retrieving items from an interview).

With a deductive approach, two researchers familiar with the literature on HRI, human perception, and cognitive psychology conducted an extensive literature review on facial anthropomorphic trustworthiness. They discussed, theorized, and summarized all the related items from prior studies, then collected and removed the replicated items, last agreed on main themes (see Table 4.1).

With the inductive approach, a questionnaire was distributed via MTurk. 200 participants were recruited to give their opinions on facial anthropomorphic trustworthiness towards social robots. Specifically, they have initially informed the general information of this study. After contending to participant in this research, they were asked to report their demographic information and prior robot interaction experience (mean age = 36.31, SD = 10.47; 112 males and 88 females; 19 with high school graduate or lower, 60 with some college, 122 with college graduate or above; 128 with never use a robot, 39 with 0-1 year use experience, 19 with 1-2 years use experience, 14 with more than 2 years use experience). Then, they were randomly exposed to a robot face from a robot face dataset (Mathur and Reichling, 2016), which structurally summarized eighty typical robot faces in the current market. Last, they were instructed to reflect on their reasons why the robot face showed looks trustworthily/untrustworthy.

Next, we utilized NLP techniques to process the collected responses in the four steps: preprocess the content, identify informative content, clustering, and meaning extracting (see Figure 4.2).

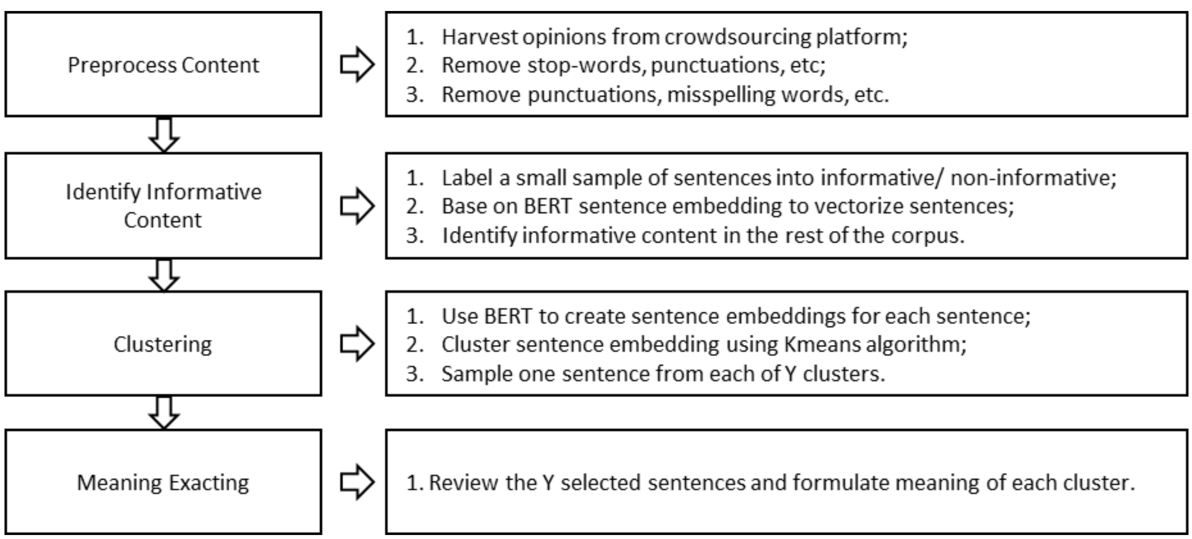
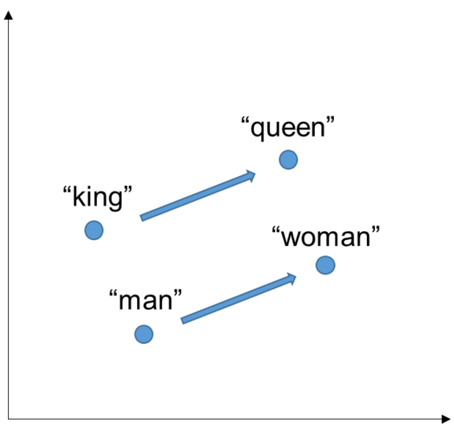


Figure 4.2 The architecture of the item pooling process in study 1

4.2.2 Natural Language Processing (NLP) Approach

Natural language processing (NLP) is a multidisciplinary field of artificial intelligence, languages, information extracting and representation, and data science, which aims to explore the associations between mathematical representations and natural languages, especially in the field of processing and analyzing a large-scale human language via coding (Deng, 2014). There are generally two subfields of NLP techniques that are related to this study: word2vec (word-to-vector) embedding, and Bidirectional Encoder Representations from Transformers (BERT). Semantic words or sentences are trained to be mathematically represented (vectorized) as real-valued mappings (around 20-400 dimensions). In this way, similar words or sentences are mathematically close to each other in the vector space (Goldberg and Levy, 2014). This technique indicates the mechanism in the word2vec or sentence embedding (either average, sum, or contacting a set of word vectors to produce sentence embedding): words or sentences which co-occurred in a similar context would share similar linguistic connotations (Goldberg and Levy, 2014). Moreover, after training a large number of a specific corpus, word2vec embedding could reveal not only the extent of similarity between words or sentences but also the semantic relationships between words and sentences (Mikolov et al., 2013). For example, after training on the corpus of Google News, word2vec embedding could reflect the following relationship: the vector distance between the word "king" and "queen" is similar to the vector distance between the word "man" and "woman" (Mikolov et al., 2013) (Figure 4.3).



$$v(\text{"king"}) - v(\text{"queen"}) \approx v(\text{"man"}) - v(\text{"woman"})$$
$$v(\text{"France"}) - v(\text{"Paris"}) \approx v(\text{"Britain"}) - v(\text{"London"})$$
$$v(\text{"strongest"}) - v(\text{"strong"}) \approx v(\text{"quickest"}) - v(\text{"quick"})$$

Figure 4.3 Word representation example in vector space in phase 1

However, traditional word2vec has two main drawbacks: it cannot solve neither word polysemy problem nor the complex characteristics of a sentence. Word2vec starts from the distributed hypothesis of word meaning (the meaning of a word is given by words that frequently appear in its context), and the result is a look-up table, where each word is associated with a unique dense vector (Goldberg and Levy, 2014). Indeed, each word in different contexts may have different meanings: its numerical values should not be a fixed vector (Verhelst and Moons, 2017). However, the word representation generated by word2vec is static, regardless of context. In other words, a look-up-styled word2vec embedding is difficult to adapt and perform well to all downstream tasks, thus a variety of adapted models are introduced for different tasks, which are basically generated by adding their own inductive biases for each task (Lan et al., 2019).

In order to address those drawbacks of word2vec, in 2018, Jacob Devlin and the research team (Devlin et al., 2018) from Google Co. introduced BERT as a state-of-art (SOTA) technique for NLP. Generally speaking, BERT is a method of pre-training language representations, namely, a general "language comprehension" model trained through numerous corpora, such as Wikipedia (around 2,500 million words) and a book corpus (around 800 million words), then is utilized for downstream language tasks (after fine-tuning), such as classification and autonomous conversions (Devlin et al., 2018). The major innovation of BERT is the proposed pre-train method. To specify, it relied on the mask language modeling (MLM) and the next sentence prediction (NSP) to capture the representation of text and sentence level, respectively. Compared with traditional word2vec embedding, BERT has two significant advantages: on one hand, BERT uses a transformer (encoder) as a feature extractor to solve the polysemy problem through representing each word as a function of the whole sentence and extracting context information for each word in the forward and backward directions (Chen et al., 2017). Cooperating with de-noising targets such as the masked language model (MLM) on large-scale corpora, the generated representations are constructive for downstream tasks, such as classification. Therefore, compared with the word embedding method represented by word2vec, BERT has a more noticeable improvement which is more dynamic and can model the phenomenon of polysemy. On the other hand, pre-trained models are designed to include different levels of language features at different network layers since different tasks rely on different levels of features differently (Devlin et al., 2018): some tasks might rely on more abstract information, while others focus more on grammatical information. In this way, BERT can selectively use the information at all levels, which could reflect different levels of features on different network layers due to its learning in a "deep" network (Y. Liu et al., 2019). Accordingly, BERT was adopted in the following four stages.

- Preprocess Content.

Previous qualitative research has suggested that each sentence in the corpus is a natural unit that could potentially reflect user opinion or experience (Rietjens, 2015). Thus, all the qualitative responses from AMT was split into a set of sentences via an unsupervised sentence split toolkit (Kiss and Strunk, 2006). Then, we cleaned them by removing the stop-words, converting all letters to lowercase, transferring numbers into number signs, removing punctuations, accent marks, and other diacritics, removing white spaces, and processing abbreviations (Timoshenko and Hauser, 2019).

- Identify Informative Content.

In this stage, we labeled a relatively small set of sentences into two categories (informative/non-informative) and then applied BERT embedding to each sentence, filtering out non-informative sentences (sentences not relevant to this study) from the rest of the corpus. With BERT identification, researchers would focus more on informative sentences (Timoshenko and Hauser, 2019).

- BERT sentence embedding clustering.

Considering similar sentences should have close distance in the BERT embedding vector space, the set of sentences were then grouped into the cluster via K-Means clustering algorithm, which is a method of vector quantization commonly used in text clustering studies (Xiong et al., 2017). To specify, K-means clustering is to divide the n observations (x_1, x_2, ..., x_n) into k (≤ n) set S = {S1, S2, ..., Sk}, ensuring the minimization of the within-cluster sum of squares.

$$\operatorname{argmin} \sum_{i=1}^{k} \sum_{x \in S_i} \|\mathrm{x} - \mathrm{n_i}\|^2 = \operatorname{argmin} \sum_{i=1}^{k} |S_i| \operatorname{Var} S_i$$

where u_i is the mean of points in set S_i

To identify an optimal number of clusters, the elbow method was widely used to determine Y clusters (Marutho et al., 2018). This method relies on calculating the sum of squared distance as different clusters of k increase to choose the optimal number of k when the sum of squared distance is only reduced marginally. As shown in Figure 4.4, four might be an appropriate number of clusters for this dataset, which is also consistent with results in a pilot study and inductive approach (Syakur et al., 2018).

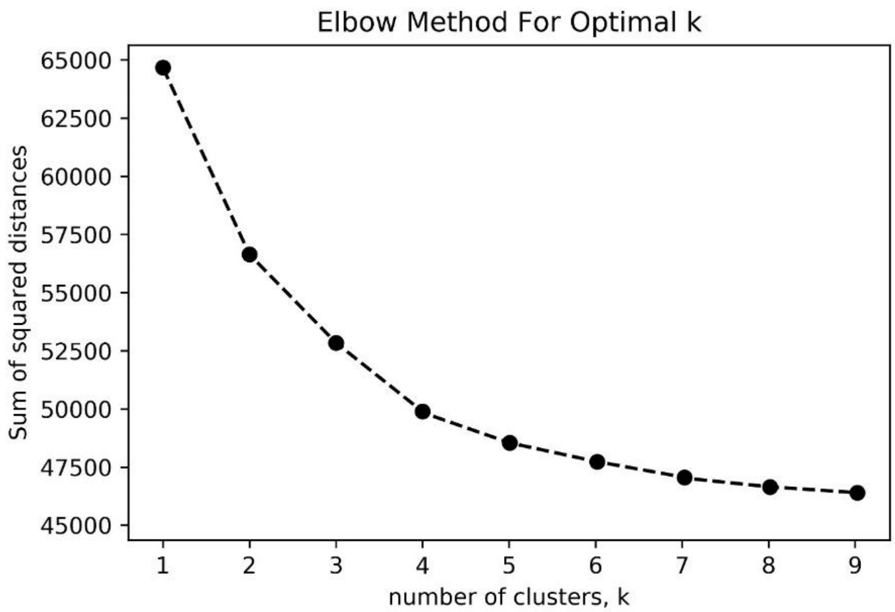


Figure 4.4 The elbow method to determine the optimal clusters in study 1

- Meaning Extracting.

In order to get an insight into the abstract opinions on facial anthropomorphic trustworthiness, we invited two experienced qualitative researchers to retrieve the relevant intuitions from the clusters. Together with the result of the pilot study and inductive approach, four themes are similarly identified and resulted in the generation of 82 items in total (see Table 4.1 and Appendix C).

Table 4.1 The source and example of the item pool in study 1

Variables	Methods	Source and Example
Ethics Concern	Deductive	Schaefer (2016), Hancock et al., (2011), Tay et al., (2014), Wheless and Grotz (1977), Colquitt and LePine (2007), Yogoda and Gillan (2012), Bhattacherjee (2002), Büttner & Göritz (Park and Mowen, 2007)
	Inductive	*"I'd trust the one learned from a compassionate creator in a safe loving environment"* *"People can write various codes and programs to make robots do evil things"*
Capability	Deductive	Schaefer (2016), Hancock et al., (2011), Tay et al., (2014), Wheless and Grotz (1977), Colquitt and LePine (2007), Yogoda and Gillan (2012), Bhattacherjee (2002), Büttner & Göritz (Park and Mowen, 2007)
	Inductive	*"They are good robots and competent enough to their programmed task"* *"I want to see them as robots that will perform their duties in an efficient manner"*
Positive Affect	Deductive	Schaefer (2016), Hancock et al., (2011), Tay et al., (2014), Wheless and Grotz (1977), Colquitt and LePine (2007), Yogoda and Gillan (2012), Bhattacherjee (2002), Büttner & Göritz (Park and Mowen, 2007)
	Inductive	*"Overall, the robot should be cute as that makes me feel protective of it and more trustful"* *"The robot looks like a robot it feels more honest and open"*
Anthropomorphism	Deductive	Ho & MacDorman (2017), Walters et al., (2008)
	Inductive	*"If it's making a poor attempt at looking humanlike I immediately distrust it and am afraid of it"* *"Trusting robots that look like classic robots is easier than a robot that is made to look like a human"*

4.2.3 Item Refinement and Polishing

Item refinement follows a two-step item refinement. To specify, a group of five professors and Ph.D. candidates in various disciplinary, such as design, sociology, business, evaluated the content validity of the items. Every researcher was informed with four dimensions associated with facial anthropomorphic trustworthiness and an example item. Next, followed the suggestions by Blijlevens et al. (2017) and Bloch et al. (2003), researchers were asked to classify each of 82 items to one of four dimensions or a "none of these" category. Items were removed when they cannot be classified to a particular category by at least four researchers. Thus, this process resulted in 45 remaining items.

Then, the 45 items of four dimensions were served as the input for a second step categorization task. Ten more researchers (different from the above researchers) with various disciplinary background have informed the definition of each dimension and then asked to rate the extent of representation of remaining items on a 5-point Likert scale (1 = "not representative of the dimension", 5 = "very representative"). The average scores of each item served as the reference for the panel discussion among ten researchers. The panel discussion was aimed to negotiate and determine whether the item was representative enough to the given dimension until reaching a consensus (Xie and DeVellis, 1992). This process resulted in 23 items. Then, two native English-speaking researchers modified the items to the given structure which served as the input for the item reduction (i.e. "this robot looks [adjective]").

4.2.4 Item Reduction and Exploratory Factor Analysis

This step consisted of item reduction and exploratory factor analysis. The process and stimuli were the same as in the pilot studies. Participants were informed that they see and evaluate a set of social robot faces. Upon exposure to the stimuli, they were requested to specify their agreement with a set of items via a 9-point Likert scale (1 = strongly disagree, 9 = strongly agree). A total of 154 people were recruited from the crowdsourcing platform, AMT. Of these 154 participants, incomplete responses or responses with only consecutive or extreme values (1-9) were screened out from the questionnaire. The EFA analysis was conducted with a total of 125 participants (mean age = 36.35, SD = 10.29; 77 male and 48 females; education level: 12 with high school graduate or lower, 48 with some college, 65 with college graduate or above; robot use experience: 102 with never use, 16 with 0-1 year use [1 year not included], 6 with 1-2 years use [2 years not included], 1 with more than 2 years use).

All items were scored via a 9-point Likert scale where higher scores suggested a higher level of facial anthropomorphic trustworthiness for a social robot. Before the exploratory factor analysis, we initially ran the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity to check whether the sample was appropriate for EFA (Kaiser, 1974). Results showed the KMO value was 0.914 and a significant Bartlett test ($p < 0.001$), suggesting the current dataset was appropriate for EFA. Following the initial item reduction of the scale development procedure (Bloch et al., 2003), the corrected item-total correlations estimation was calculated on the whole set of 23 items. Considering the threshold of the corrected item-total correlations (0.4 or above), two items were removed since they did not meet satisfactory item-total correlations. Then, a preliminary EFA via varimax rotation was conducted on the remaining set of 21 items. Three items were removed since they were conceptually irrelevant to the other items that loaded on the specific construct: "typical representation" (construct typicality) and "diverse representation" (construct variety) (Blijlevens et al., 2017), and one item was removed due to its failure to exhibit simple structure on any factors (Bloch et al., 2003). Again, we re-performed the EFA on the remaining set of 17 items. All the corrected item-total correlations were beyond the threshold; thus, the remaining 17 items were retained. Besides, the EFA indicated a 17-item scale with four dimensions (cluster relied on its eigenvalues were one and above): ethics concern, capability, affective perception, and the uncanny valley effect. Regarding Cronbach's alphas of each dimension, 0.89 was for "ethics concern"; 0.94 was for “capability”; 0.94 was for “positive affect”; 0.88 was for “anthropomorphism” (Table 4.2).

Table 4.2 The 17 items of the final scale

Ethics concern (5)	**Capability (4)**
EC1. This robot does not look evil	CAP1. This robot looks competent in its work
EC2. This robot looks as if its creator is not intending to harm humanity	CAP2. This robot looks like it can perform its duties in an efficient manner
EC3. The designer has ethically programmed this robot	CAP3. This robot looks like it can be successful in the matter it is programmed to do
EC4. This robot seems to act following its program	CAP4. This robot looks like it can provide appropriate information
EC5. This robot seems reasonable when interacting with a human	
Positive Affect (4)	**Anthropomorphism (4)**
AFF1. This robot looks kind	AN1. This robot face looks neither too living nor too inanimate
AFF2. This robot looks cute	AN2. This robot face looks neither too humanoid nor too robotic
AFF3. This robot looks considerate	AN3. This robot face looks neither too real nor too synthetic
AFF4. This robot looks like it cares about my welfare	AN4. This robot face strikes a balance between a human-like face and a machine-like face

4.2.5 Validation

The same survey procedure was conducted in section 4.2.3: the same stimuli exposure, screening, and recruitment process was conducted to examine the final set of 17 items. As a result, 282 participants were included in the validation (mean age = 36.38, SD = 10.65; 177 male and 105 females; education level: 27 with high school graduate or lower, 81 with some college, 174 with college graduate or above; robot use experience: 201 with never use, 53 with 0-1 year use [1 year not included], 19 with 1-2 years use [2 years not included], 9 with more than 2 years use).

Before the main analysis of confirmative factor analysis (CFA), we ran EFA first to confirm the validity of the four dimensions. Based on the eigenvalues (greater than one), the current result was consistent with the conclusion of section 4.2.3, suggesting the same four dimensions with the same items. The exploratory factor analysis of the final scale is depicted in Table 4.3.

Table 4.3 The exploratory factor analysis of the scale in study 1

	Factors			
	Capability	**Ethics Concern**	**Anthropomorphism**	**Positive Affect**
CAP1	0.848			
CAP2	0.794			
CAP3	0.766			

CAP4	0.752			
EC2		0.752		
EC1		0.696		
EC5		0.653		
EC3		0.635		
EC4		0.615		
AN2			0.898	
AN3			0.832	
AN1			0.725	
AN4			0.673	
AFF3				0.744
AFF2				0.743
AFF1				0.715
AFF4				0.692
Eigenvalues	8.662	2.136	1.438	1.024
% of variance	50.953	12.564	8.459	6.021

Note: Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in six iterations with loading values more than 0.5.

As for CFA, the analysis was conducted through AMOS 25 for structural equation modeling (SEM) (J. Zhang et al., 2019). Particularly, SEM was utilized to examine whether the proposed model was structurally fitted with the sample. Specifically, the identified items on the same factors from EFA should be treated as the proposed model in the validation. Thus, the four-factor model (ethics concern, capability, positive affect, and anthropomorphism) from section 4.2.3 was used to test the data obtained in the second study through SEM (Blijlevens et al., 2017).

According to the result, it was shown the model fit index achieved adequate values, confirming the general appropriateness of the model from EFA (Fornell and Larcker, 2006). To specify, the goodness of fit measure (GFI) was 0.892 (threshold: 0.9 and above); incremental fit index (IFI) was 0.957 (threshold: 0.9 and above); the normed fit index (NFI) was 0.929 (threshold: 0.9 and above); the comparative fit index (CFI) was 0.957 (threshold: 0.9 and above); the adjusted goodness of fit measure (AGFI) was 0.854 (threshold: 0.8 and above), and the root-mean-square error of approximation (RMSEA) was 0.07 (threshold 0.05-0.08). In addition, all items had significant factor loadings, varying from 0.71 to 0.91 (see Figure 4.5), and all explained variances (squared multiple correlations, SMC) of our items varied between 0.49 and 0.83.

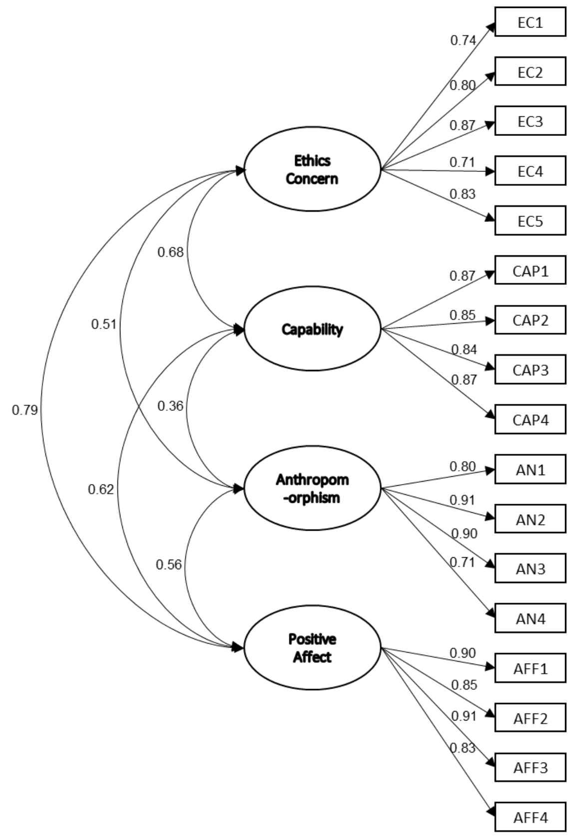


Figure 4.5 The path diagram of facial anthropomorphic trustworthiness scale in study 1

As for reliability and convergent validity, the average variance extracted (AVE) for every construct had achieved a satisfactory value (0.50 and above), suggesting the current sample had sufficient convergent validity (Fornell and Larcker, 2006). As suggested by Fornell and Larcker (2006), the composite reliability (C.R.) of each constructs all achieved adequate value (0.60 and above).

Table 4.4 The correlation matrix of different constructs in study 1

	CR	AVE	MSV	EC	CAP	AN	AFF
Ethics Concern (EC)	0.89	0.63	0.63	0.79			
Capability (CAP)	0.92	0.74	0.46	0.68***	0.86		
Anthropomorphism (AN)	0.90	0.69	0.31	0.51***	0.36***	0.83	
Positive Affect (AFF)	0.93	0.76	0.63	0.79***	0.62***	0.56***	0.87

Note: * p<0.1; ** p<0.05; *** p<0.01

Regarding the discriminant validity within the model. The discriminant validity between the four constructs was measured by whether the square root of the AVE (diagonal values in Table 4) is larger than the rest of the inter-construct correlations and the maximum shared variance (MSV) (Fornell and Larcker, 2006). Results suggested all four constructs have reached adequate discriminant validity.

4.3 Summary and Discussions

With increased technology and equipment applied in social robots, it is becoming a medium or a communication partner between human and digital data in our daily lives, supporting us in physical and emotional ways (Hoorn, 2015). Considering the significant role of ethics evaluation at the initial step of HRI, facial anthropomorphic trustworthiness indeed plays a crucial role in building the initial credibility and the approaching intention at a later stage (Hoorn, 2015). Within the domain of social robots, it lacks a valid scale to measure the construct of interest: facial anthropomorphic trustworthiness. In order to address the research gap above, this thesis tries to develop a reliable and valid measurement to assess facial anthropomorphic trustworthiness in the context of social robots. Thus, the current scale could be utilized in future empirical studies in HRI that aim to assess the determinants influencing trustworthiness, especially at first sight. As a result, four dimensions were illustrated that could be applied to assess facial anthropomorphic trustworthiness: ethics concern, capability, positive affect, and anthropomorphism. This conclusion was consistent with the theory of robot communication (TORC). According to TORC, there are generally two parallel paths for humans to interact with a social robot: robot mediated communication (a slow and reflective route; RMC) and human-robot communication (a quick and affective route; HRC) (Hoorn, 2015). People might have these two routes at the same time though the one might dominate the other or constantly switch with the other, which were consistent with dimensions found in the current study: the affective route (ethics concern and positive affect) and reflective route (capability, and anthropomorphism).

There are several points worth noting. Different cultures might have different preferences in processing facial trustworthiness features. For example, Japanese/Israeli people would consider their own-culture typical faces to be more trustworthy than other-culture typical faces (Sofer et al., 2017). It is also interesting to explore the effect of cultural issues on perceived trustworthiness when encountering and interacting with the same social robot. Thus, a future study is needed to check whether cultural factors could influence the dimensions of facial anthropomorphic trustworthiness.

CHAPTER FIVE. STATIC INTERNAL FEATURES IN COMMUNICATING FACIAL ANTHROPOMORPHIC TRUSTWORTHINESS

This chapter explores the effect of internal static features on facial anthropomorphic trustworthiness. Through examining internal facial features (eye shape and mouth shape), this chapter empirically investigates their influence on perceived trustworthiness, contributing to our knowledge of internal features in influencing facial anthropomorphic trustworthiness.

5.1 Introduction

As indicated in Chapter Two, the eye shape and mouth shape are considered as the most significant features that could influence people's evaluation of trustworthiness, both for human and robot (Ichikawa et al., 2011; Kaisler and Leder, 2016; Kleisner et al., 2013; Landwehr et al., 2011; Santos and Young, 2011; Windhager et al., 2010; Zebrowitz et al., 1996). Although prior research has discussed the effect of facial biological features, such as eye shape, mouth shape, positioning, and movement, on trustworthiness evaluation, the majority of research concentrated in the context of human facial perception. Limited research has addressed the relationship between specific facial features of the robot and its trustworthiness evaluation. Similar to human facial features, the facial features of a social robot also refer to the size, shape, positioning, and movement of a facial organism (Liu et al., 2013). McGinn (2019) suggested social robot would be ideally equipped with facial features that could provide social interaction by attention attraction, and, eventually, be perceived to be like a real human or a partner in human-robot interaction. However, to date, it is unclear whether there is the "halo" effect in social robots' perception and whether the rules in human facial perceptions could be applied and still work as significant drivers for trustworthiness evaluation for social robots. Thus, it is both theoretical and practically important to explore the effect of specific facial features, such as eye and mouth shape, on a robot's trustworthiness evaluation. Based on the literature above, we might have the theoretical model and following hypotheses (see Figure 5.1):

H 5.1: People tend to have a higher trustworthiness perception toward a robot with round eyes design (vs. narrow eyes design).

H 5.2: People tend to have a higher trustworthiness perception toward a robot with an upturned or a neutral mouth design (vs. downturned mouth design).

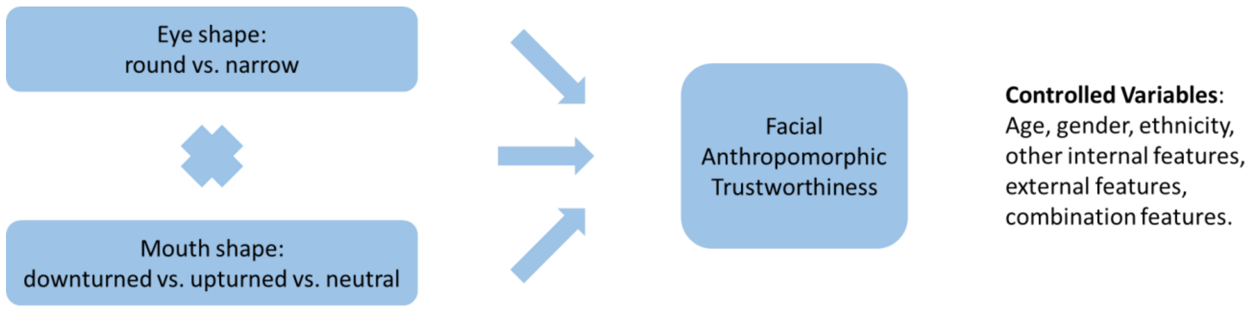


Figure 5.1 Overview of research model of study 2

5.2 Method

5.2.1 Stimuli and Experiment Design

Figure 5.1 shows the research model of study 2. We conducted a 2 (eye shape) * 3 (mouth shape) between-subject experiment design which included six scenarios: two eye shape (round vs. narrow) and three mouth shape (upturned vs. neutral vs. downturned). A robot designer made all the robot stimuli (Figure 5.2). Cooperated with a designer, we carefully controlled the potential confounders to avoid influences from an existing social robot or other related fields. Besides, we kept other robotic features, such as body height, width, positioning, posture, color, and background, unchanged.

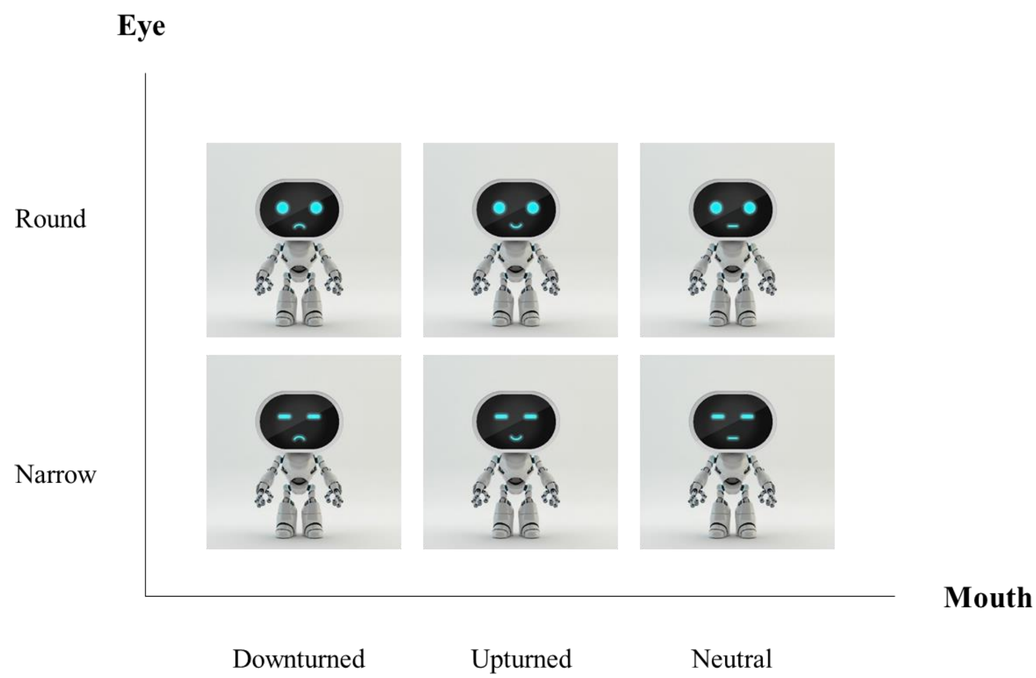


Figure 5.2 Mouth and eye shape interaction on trustworthiness evaluation in study 2

5.2.2 Participants and Experiment Procedure

211 participants took part in this experiment study (mean age = 35.74, SD = 10.7; 113 males and 98 females; see Table 5.1). After consenting to participate, they were randomly selected to expose to one of six stimuli and were asked to fill in the questionnaire.

Table 5.1 Demographic characteristics of participants in study 2

	Gender		Age	
Scenarios	**Male**	**Female**	**Mean**	**SD**
Round + Neutral	19	15	35.41	10.30
Round + Upturned	22	14	37.01	11.74
Round + downturned	19	16	35.43	9.24
Narrow + Neutral	17	20	35.42	12.83
Narrow + Upturned	20	12	36.70	10.21
Narrow + downturned	16	21	34.62	9.97

As for the measurement items, one thing that worth noticing is that the facial trustworthiness in studies 2-5 was estimated by five items on a nine-point Likert scale (I think this robot looks credible/ I think this robot looks sincere/ I think this robot looks honest/ I think this robot looks believable/ I think this robot looks convincing) (Gorn et al., 2008). The reason why I did not use the developed scale of facial anthropomorphic trustworthiness study 1 lies in my learning schedule of related analytical techniques. During this Ph.D. study, to address the issue of subjective interpretation of qualitative data, a more objective and state-of-art technique, natural language processing, in particular, was necessary to be learned and contribute to the current research. Nevertheless, this SOTA technique keeps evolving and requires numerous pre-knowledge as the basis for practical implementation, such as knowledge of machine learning. Considering this self-learning process might take an uncertain amount of time and might meet unpredictable difficulties, I began to learn this technique for study 1 while explored the effect of specific features on facial anthropomorphic trustworthiness based on the existing scale for studies 2-5 at the same time.

As discussed in section 2.3, the prior exploration of trustworthiness might focus on the overall perception of trustworthiness. For example, the scale developed by Gorn et al. (2008) mainly examines the trustworthiness perception from a general perspective (i.e. credible/ sincere/ honest/ believable/ convincing are all synonyms of trustworthiness according to the definitions of Merriam-Webster) (Marriam-Webster, 2002). Numerous studies have discussed the difference between general scale (i.e. single item or synonym substitution item) and multiple-dimension scale (i.e. multiple items) and suggested each scale might have its own advantages and disadvantages (Bergkvist and Rossiter, 2007; Gardner et al., 1998; Loo, 2002): though a general-styled scale might enjoy fair reliability of measuring a subjective perception, a fine-grained scale with multiple dimensions might help people to understand the meaning of a concept more comprehensively.

Thus, considering the practicality, the trustworthiness scale developed by (Gorn et al., 2008) was adopted in studies 2-5 while the FATSR-17 was adopted in study 6 after the acquisition of necessary knowledge. Although study 6 have suggested both scales from Gorn (2008) and study 1 enjoyed adequate reliability and validity to measure facial anthropomorphic trustworthiness, it might still be a limitation and was discussed in detail in chapter ten.

5.3 Data Analysis and Result

A two-way ANOVA was conducted with eye shape (round vs. narrow) and mouth shape (upturned vs. neutral vs. downturned shape) as the independent variables, and perceived trustworthiness as the dependent variable. To specify, the Cronbach's alpha coefficients showed a high internal consistency of five items (0.939), indicating a high consistency of the current measurement. The results showed the main effect of mouth and eye design on trustworthiness evaluation was significant while the interaction effect was not significant. To be more specific, people in the round eyes scenario (Mean = 6.02 vs. 5.51, SD = 1.69 vs. 2.14) showed significantly higher trustworthiness than those who exposed to the robot with narrow eyes ($F(1, 205) = 3.70$, $p = 0.05$). In addition, people in the upturned and neutral mouth scenario tended to have a higher trustworthiness evaluation than those in the downturned mouth scenario (Mean = 6.16 vs. 6.15 vs. 5.02; SD = 1.80 vs. 1.67 vs. 2.11, respectively; $F(2, 205) = 8.49$, $p < 0.05$) while there is no statistically significant difference between upturned and neutral mouth scenario ($p = 1.00$). There is no interaction effect between mouth and eye on trustworthiness evaluation (F(2, 205) = 0.25, p = 0.78). Thus, H 5.1 and H 5.2 were supported (see Figure 5.3).

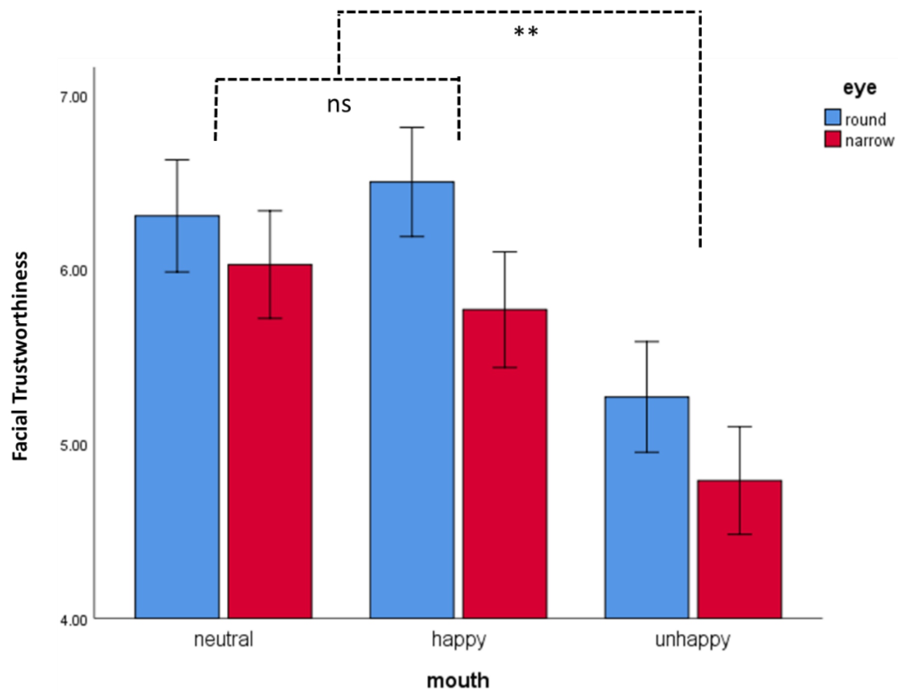


Figure 5.3 The effect of mouth and eye shape on trustworthiness toward the social robot in study 2

Note: ** means significant < 0.05; ns means non-significant

Last, additional investigation was performed to analyze the correlation relationship between demographics and facial anthropomorphic trustworthiness, results of the

Pearson test showed no significant correlation between facial anthropomorphic trustworthiness and age ($p = 0.11$) or gender ($p = 0.74$).

5.4 Summary and Discussions

Although trustworthiness is one of the most fundamental social attributions and numerous research has explored the relationship between specific facial features and trustworthiness perception (Calvo et al., 2017; Ma et al., 2015; Stirrat and Perrett, 2010), the majority of prior research has focused on the facial trustworthy features in the context of human in which facial trustworthy features for social robots have largely neglected. Regarding trustworthiness toward social robots also plays a crucial role in human-robot interaction, this research tries to address this research question by examining the effect of specific facial features, eye and mouth shape, on robot trustworthiness evaluation. According to the results, this research validated that 1) round eyes (vs. narrow eyes) and an upturned-shape mouth or neutral mouth (vs. downturned-shape mouth) for social robots could significantly improve people's trustworthiness evaluation in social robots; 2) round eyes (vs. narrow eyes) and an upturned-shape mouth or neutral mouth (vs. downturned-shape mouth) for social robots could also significantly improve people's trustworthiness evaluation in social robots; 3) there was no interaction effect between eye and mouth shape on trustworthiness evaluation.

The current research contributed to the field of human-robot interaction. Prior research on facial trustworthiness has suggested people might consider round eyes (vs. narrow eyes) are strong indicators for the baby-face appearance traits (Haselhuhn et al., 2013; Maoz, 2012), thus improving trustworthiness (Ferstl et al., 2017; Masip et al., 2004). Similarly, compared with a downturned mouth (sad mouth) and a neutral mouth (Landwehr et al., 2011), people with an upturned mouth (smiling mouth) were believed to be more trustworthy and friendlier (Arminjon et al., 2015; Kleisner et al., 2013; Landwehr et al., 2011; Maeng and Aggarwal, 2018). However, it was unclear whether the rules might work in the facial design of the social robot and influence the related social perceptions. Consistent with the previous conclusion, the current research, for the first time, provided the initial evidence to prove social robots with round eyes and an upturned mouth (or neutral mouth) could improve people's trustworthiness toward the social robot.

CHAPTER SIX. STATIC EXTERNAL FEATURES IN COMMUNICATING FACIAL ANTHROPOMORPHIC TRUSTWORTHINESS

This chapter explores the effect of external static features on facial anthropomorphic trustworthiness. Through examining external facial features (fWHR and face shape), this chapter empirically investigates their influence on perceived trustworthiness, contributing to our knowledge of external features in influencing facial anthropomorphic trustworthiness.

6.1 Introduction

fWHR is one of the most significant features in communicating social attribution, such as perceived dominance and trustworthiness, from various facial features (Ferstl et al., 2017; Gomulya et al., 2017; Re and Rule, 2016). To specify, prior research has suggested that fWHR is negatively related to facial trustworthiness and general facial attitude (likeness) while positively associated with facial dominance. Indeed, individuals could have a more favorable acceptance and attitude toward the object which could fulfill their desires (Ajzen, 2001).

Contradicted with the result that people with a high level of fWHR could experience a low level of facial attractiveness and people with a low level of fWHR could enjoy a low level of popularity, a recent neuroscientific study has indicated that objects with a high-level fWHR could rather have a high level of powerfulness and a more favorable attitude (a high rewarding status) because dominant appearance could assist individuals to experience an empowered self (Maeng and Aggarwal, 2018). Thus, it is possible that social robots with a high level of fWHR could assist individuals to experience an empowered image, during which the dopaminergic effects on the reward system make individuals have a high level of perceived trustworthiness.

Moreover, considering the crucial effect of shape in designing objects (Hsiao and Huang, 2002), the face shape of social robots could have an impact on individuals' trustworthiness as well. There are generally two shapes in objects, round shape, and rectangular shape (Westerman et al., 2012). Although many efforts have been made to examine the association between object shape and relevant attribution, there still exists controversy on the relationship between shape and related attribution. On the one hand, a rectangular face stands as a typical visual component for robot design (Meeden and Blank, 2006). The symbolized typicality could assist individuals to easily categorize one object and increase the associated evaluation (Blijlevens et al., 2012). On the other hand, individuals have indicated a more positive attitude toward round shape (vs. rectangular shape) (Westerman et al., 2012). In light of this, we might argue that, despite rectangular typicality in the robot form, preference for round shape could still promote facial anthropomorphic trustworthiness perceptions (see Figure 6.1). Namely,

H 6.1: People tend to have a high level of facial anthropomorphic trustworthiness toward a social robot with a high level of fWHR face (vs. a low level of fWHR face).

H 6.2: People tend to have a high level of facial anthropomorphic trustworthiness toward a social robot with a round face (vs. rectangular face).

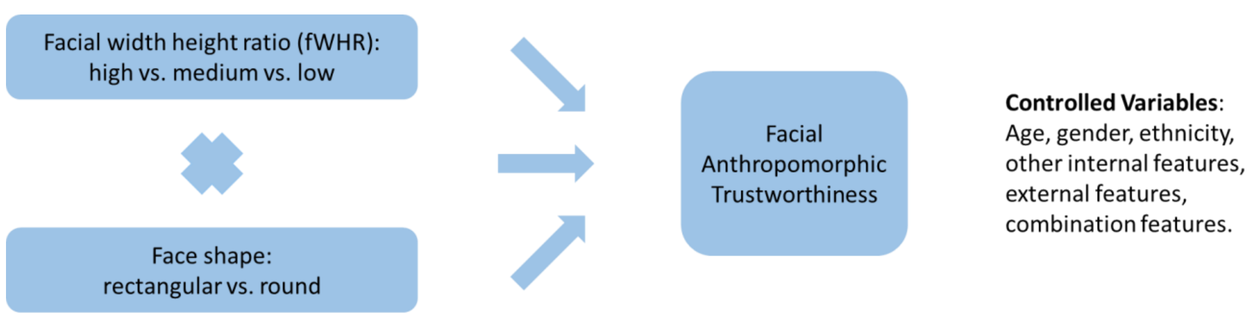


Figure 6.1 Overview of research model of study 3

6.2 Method

6.2.1 Stimuli and Experiment Design

With regard to the experiment design, a two (face shape: round vs. rectangular shape) by three (fWHR: high vs. medium vs. low) between-participants experiment was planned. In total, the trial included 6 various schemes. A professional robot designer was recruited to design all six trial stimuli (see Figure 6.2). Within the progress of robot design, the robot designer was instructed to keep the potential influential components, for example, the emotional expressions, other facial features, robot body configurations (height and width), color selection, and presented the background, staying the same.

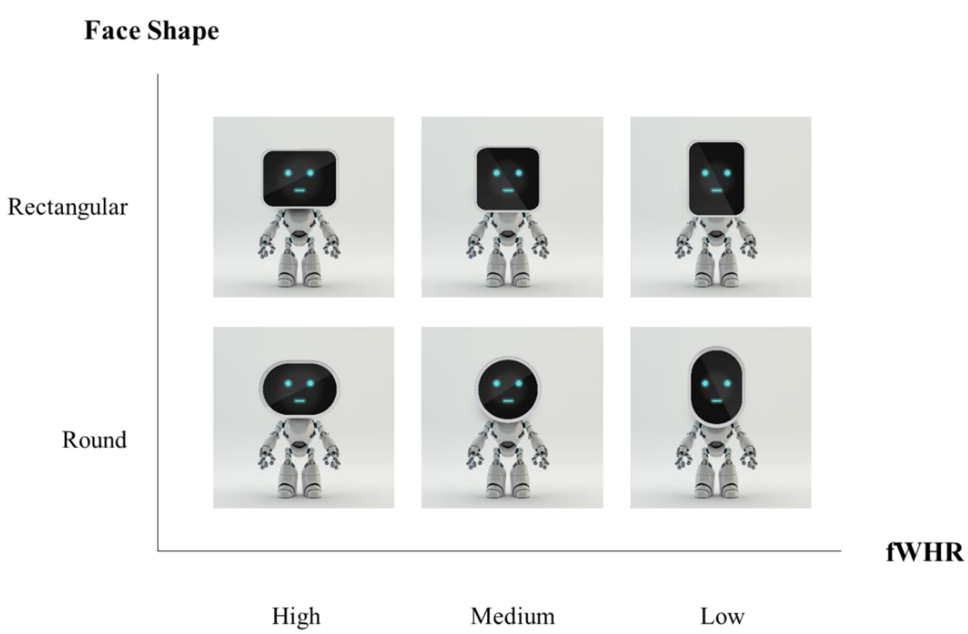


Figure 6.2 The effect of different fWHR and face shape of experimental stimuli in study 3

6.2.2 Participants and Experiment Procedure

In order to test the effect of fWHR and face shape on facial anthropomorphic trustworthiness, an online sample from Amazon Mechanical Turk (MTurk) was recruited in this study. Since MTurk is a reliable crowdsourcing platform that invites individuals to finish the assigned work for compensation (Mortensen and Hughes, 2018), much research, ranging from psychology, behavior, to HCI, have adopted this platform for scientific research. Regarding its satisfactory accuracy (Khare et al., 2015) and reliability (Deal et al., 2016) compared with physical lab experiments (Brañas-Garza et al., 2018), it was deemed as an appropriate data collection method to analyze the association between specific components of robot design and facial anthropomorphic trustworthiness.

As for the manipulation check, individuals were instructed to express the degree they agreed with two particular items on a 9-point scale (Item contained: "I think the face of this robot is wide"; "I think the shape of this robot's face is round"). Then, individuals were instructed to show their agreement of the extent of facial anthropomorphic trustworthiness toward the given robot on a 9-point scale, which works as the dependent variable in this study (Gorn et al., 2008).

As a result, a total number of 240 individuals was recruited in the study (mean age = 36.63; SD = 11.19). They were randomly and equally assigned to six schemes (each scheme contained forty individuals). Table 6.1 showed detailed demographic information in study 3.

Table 6.1 Demographic characteristics of participants in study 3

	Frequency	**Percent**		**Frequency**	**Percent**
Gender			**Education**		
Male	144	60.0%	High school graduate or lower	27	11.3%
Female	96	40.0%	Some college education	58	24.2%
			College graduate or above	155	64.5%
Age					
18–25	30	12.5%	**Robot interaction experience**		
26–30	59	24.6%	Never	169	70.4%
31–40	81	33.8%	0–1 year (1 year not included)	44	18.3%
41+	70	29.1%	1–2 years (2 years not included)	23	9.6%
			2+ years	4	1.7%

6.3 Data Analysis and Result

Prior to the main study, the manipulation check for fWHR and face shape was examined in advance. The investigation of the one-way ANOVA between fWHR and face shape suggested a significant difference between various fWHR (mean = 6.15 vs. 7.29 vs. 7.78; SD = 2.08 vs. 1.68 vs. 1.37; $F(2, 237) = 18.35$, $p < 0.01$) and face shapes (mean = 2.81 vs. 6.45; SD = 2.57 vs. 2.31; $F(1, 238) = 132.59$, $p < 0.01$). This result confirmed the successful manipulations of fWHR and face shape.

Regarding the main study, we then examined the effect of fWHR and face shape on the facial anthropomorphic trustworthiness within different schemes.

To be more specific, a two-way ANOVA, with three levels of fWHR (high vs. medium vs. low) and two levels of face shape (round vs. rectangular) as the independent variables and with facial anthropomorphic trustworthiness as the dependent variable, was conducted. The internal consistency of all measurement items was tested via Cronbach's alpha (0.951), indicating adequate reliability for facial anthropomorphic trustworthiness, (Song and Luximon, 2020b). As for the results of the two-way ANOVA, it suggested that the significant main effect of fWHR on facial anthropomorphic trustworthiness ($F(2, 234) = 6.01$, $p < 0.01$), and the insignificant effect of face shape ($F(1, 234) = 0.28$, $p = 0.60$) and the interaction ($F(2, 234) = 0.12$, $p = 0.89$) on facial anthropomorphic trustworthiness. To specify, the post-hoc results unveiled that social robots with a high level of fWHR or medium level of fWHR had a high level of facial anthropomorphic trustworthiness than those with a low level of fWHR. Nevertheless, social robots with a medium level of fWHR did not have a significant difference in facial anthropomorphic trustworthiness with robots with a high level of fWHR. Thus, H 6.1 is supported while H 6.2 is not supported (see Table 6.2, Table 6.3, and Figure 6.3).

Table 6.2 Descriptive statistics for trustworthiness in different fWHR and face shape scenarios in study 3

		fWHR (Mean ± SD)			
		Low	Medium	High	Total
Face shape	Rectangular	5.38 ± 1.41	5.92 ± 1.84	6.08 ± 1.53	5.79 ± 1.62
	Round	5.35 ± 1.79	6.05 ± 1.35	6.29 ± 1.37	5.90 ± 1.56
	Total	5.36 ± 1.60	5.99 ± 1.61	6.18 ± 1.45	5.85 ± 1.59

Table 6.3 The Post-hoc comparisons and effect sizes for trustworthiness between different fWHR scenarios

	Mean difference	**SE**	***t* statistic**	**Cohen's d**	**Effect size**	***p* (Tukey)**	**95% CI**	
							Lower bound	**Upper bound**
Low–Medium	−0.624	0.247	−2.526	−0.389	Medium	< 0.05	−1.206	−0.041
Low–High	−0.819	0.247	−3.318	−0.537	Large	< 0.01	−1.401	−0.237
Medium–High	−0.196	0.247	−0.792	−0.128	Small	0.708	−0.778	0.387

Note: Effect size classification follows Cohen's work (2013)

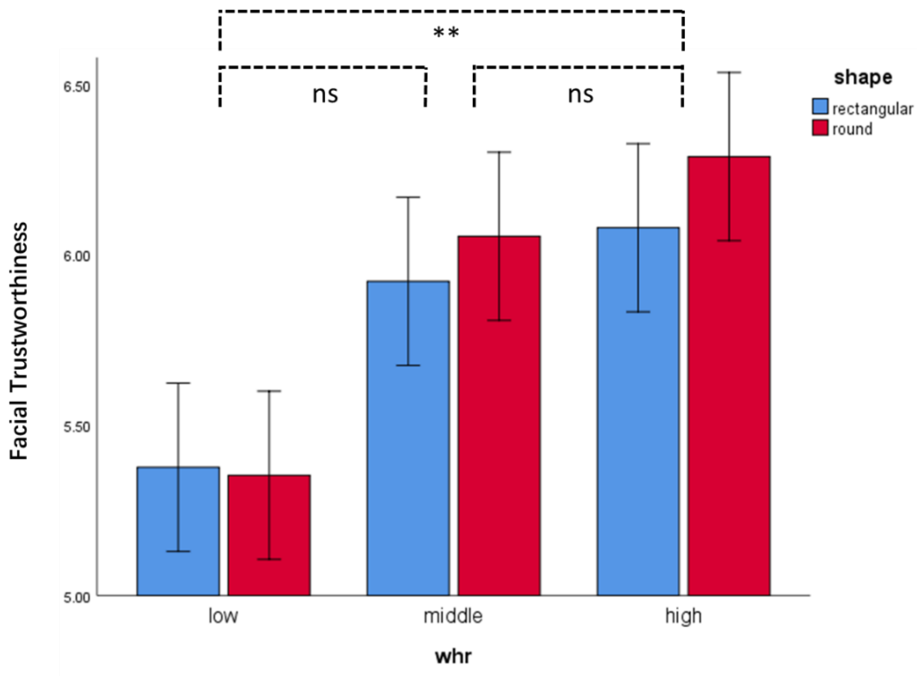


Figure 6.3 The effect of fWHR and face shape on facial anthropomorphic trustworthiness in study 3

Note: ** means $p < 0.01$

Last, additional investigation was performed to analyze the correlation relationship between demographics and facial anthropomorphic trustworthiness, results of the Pearson test showed no significant correlation between facial anthropomorphic trustworthiness and age ($p = 0.06$) or gender ($p = 0.51$).

6.4 Summary and Discussion

Considering limited research has tried to examine how robotic appearance, such as the fWHR and shape of robot face, in influencing facial anthropomorphic trustworthiness. To address this research question, this study adopted an experiment to analyze the influence of fWHR and face shape in signaling facial anthropomorphic trustworthiness. Results showed that: 1) fWHR works as a salient element in impacting on facial anthropomorphic trustworthiness; 2) individuals are inclined to have an increased facial anthropomorphic trustworthiness toward social robots with a high level of fWHR (vs. a low level of fWHR); 3) there seems no significant difference evaluating facial anthropomorphic trustworthiness between a high level of fWHR and a medium level of fWHR and between a low level of fWHR and a medium level of fWHR; 4) neither face shape nor its interaction effect has a significant impact on facial anthropomorphic trustworthiness.

This experiment could have the following academic contributions. To begin with, though prior studies about facial trustworthiness have long attracted theoretical interest, they might be concentrated in the context of human facial processing. Limited research has tried to bolster this boundary to a larger stage. For example, through a social robot with an anthropomorphic head was examined to have a positive effect on people's robot attitude (McGinn, 2019), it is still fixed on the overall morphology of the social robot, neglecting consideration of particular facial features. Correspondingly, it could be intriguing to scrutinize if the social robot would enjoy similar trustworthiness benefits when introducing and applying human facial trustworthiness features in robot design. Through an experimental method, this work indicated that external facial features, such as fWHR, might be applied in robot design to promote facial anthropomorphic trustworthiness.

Moreover, this research advances our understanding of human-robot interaction by illustrating how external facial features, such as fWHR, could significantly influence facial anthropomorphic trustworthiness. Prior literature on human-robot interaction has scrutinized the overall association between robot design principles, such as “beauty premium” or “plainness penalty,” and individuals’ associated attitudes; nevertheless, there seem few attempts into the potential influence in facial anthropomorphic trustworthiness. Hinged on the theoretical background of human trustworthiness, the current study suggests that, regarding facial cues of social robots, there might exist a counter-intuitive facial evaluation association between robot and human. Although individuals with a high level of fWHR could be considered as less trustworthy, social robots with a high level of fWHR could be deemed as more trustworthy.

CHAPTER SEVEN. STATIC FEATURE COMBINATIONS IN COMMUNICATING FACIAL ANTHROPOMORPHIC TRUSTWORTHINESS

This chapter explores the effect of static feature combinations on facial anthropomorphic trustworthiness. Through examining feature combinations (size and position of facial features), this chapter empirically investigates their influence on perceived trustworthiness, contributing to our knowledge of feature combinations in influencing facial anthropomorphic trustworthiness.

7.1 Introduction

Considering the significance of the baby schema in reproductive success, such instinct affective reactions might also arise especially regarding the artificial entities, such as social robots, with similar facial features of the baby schema (Miesler et al., 2011). On the one hand, the latest research has provided preliminary evidence that we might also be sensitive and have specific responses to the social robot with babyish features (Borgi et al., 2014). For instance, social robots that looked childlike might be acknowledged as more approachable, warm, and trustworthy (Reeves et al., 2020). The presence of lifelike eyes, rather than abstract or absence of eyes, is believed to be more personable and suitable for the home (Luria et al., 2018) while the absence of mouth, rather than presence, might aid emotional expressions (Pollmann et al., 2019). Moreover, Kalegina and her colleagues (2018) systematically summarized the most common features in the current market and created a synthetic face as the benchmark. By altering only one-dimension feature, they found that the presence of one face element (blue eyes, extreme close eyes, ears, eyelids, hair, no mouth, no pupil, small eyes, white face) might decrease people's perceived trustworthiness. On the other hand, the difference between a screen-based face and a human face was still distinct that could not be neglected. Different from a general human face, a rendered screen, though convenient and flexible, might be able to create faces with feature displacements, such as a "scattered" feature face (eye and mouth are scattered on a face; interocular eye distance is high with the low vertical position of the mouth) or a "huddled" feature face (eye and mouth are positioned closely together; interocular eye distance is low with a high vertical position of the mouth). Therefore, a social robot could have different combinations of feature size and displacement. Indeed, people are highly sensitive to feature displacement in both human and nonhuman faces (Yovel and Duchaine, 2006). Inappropriate or unnatural feature displacement might significantly reduce social judgments (Jones et al., 2001; Penton-Voak et al., 2001), such as perceived facial attractiveness (Friedenberg, 2001). Although the inward tendency of facial features might stimulate baby schema, its effect could be neutralized by extremely close positioning of eye and mouth. Accordingly, a medium vertical and horizontal positioning might enjoy a high level of trustworthiness.

Based on the prior research on evolutionary psychology and HRI, this study tries to extend the theory on facial anthropomorphic trustworthiness via a systematic size and position displacement of eye and mouth. Thus, we have the following hypotheses (see Figure 7.1):

H 7.1: Social robot with big (vs. small) eye size would be perceived as more trustworthy.

H 7.2: Social robot with a small (vs. big) mouth size would be perceived as more trustworthy.

H 7.3: Social robot with a medium (vs. high or low) vertical eye position would be perceived as more trustworthy.

H 7.4: Social robot with a medium (vs. high or low) horizontal eye position would be perceived as more trustworthy.

H 7.5: Social robot with a medium (vs. high or low) vertical mouth position would be perceived as more trustworthy.

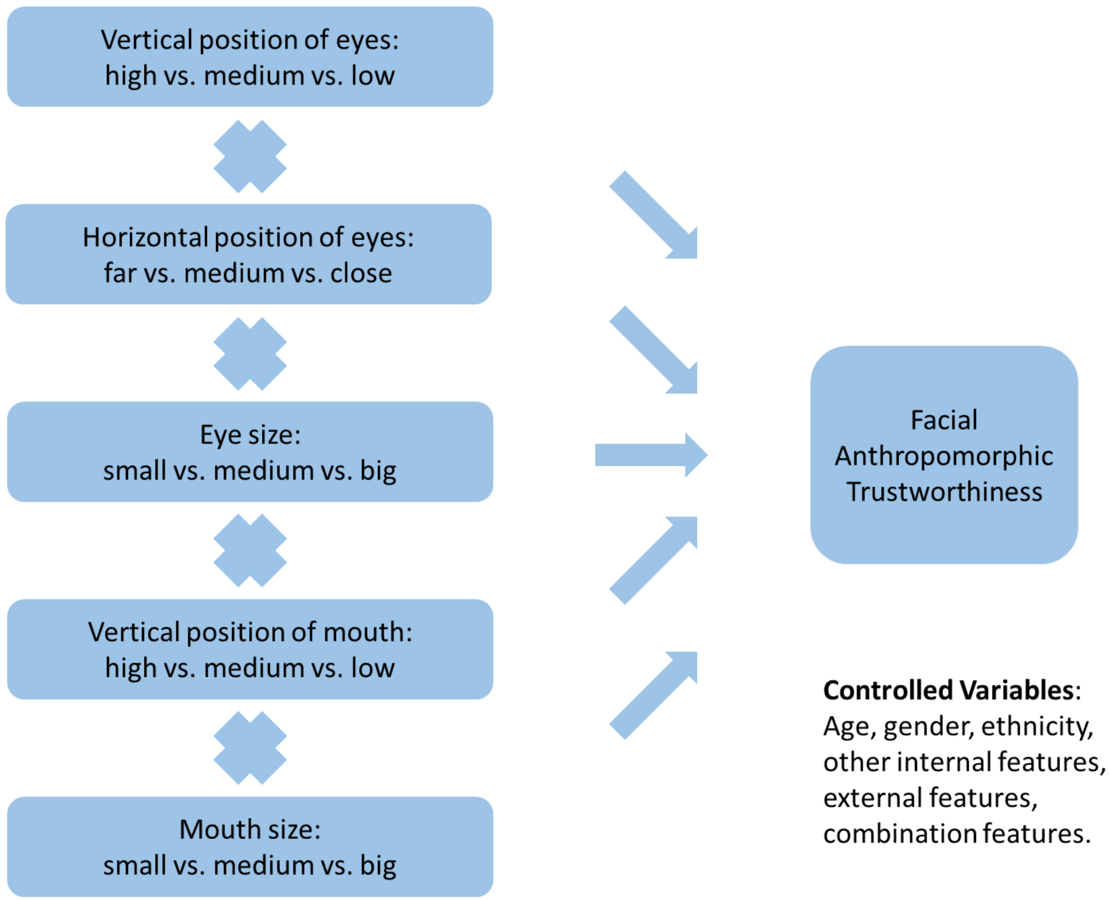


Figure 7.1 Overview of research model of study 4

7.2 Method

7.2.1 Stimuli and Experiment Design

All the robot stimuli were designed as planned. To be more specific, detailed configurations were coded as follows: three levels of eye size (small/medium/big), three levels of mouth size (small/ medium/ big), three levels of horizontal positions of eyes (far/ medium/ close), three levels of vertical positions of eyes (high/ medium/ low), three levels of vertical positions of mouth (high/ medium/ low). Regarding the shape of a baby's head, facial features, and the head traditions of the current social robots, detailed facial variations were manipulated on the prior research (Ferstl et al., 2017, 2016). For instance, high fWHR (vs. low fWHR) was settled as the default facial ratio (Song and Luximon, 2021). Other factors were also controlled to reduce the effect of potential confounding factors. Figure 7.2 shows the configured metrics for the social robot. In addition, other confounding factors, i.e. names, logos, or division, were maintained and removed to confirm the current distinctness in the market. Similarly, additional characteristics, i.e. emotional expressions, robot shape, robot posture, or presentation background, were all controlled the same. Thus, a configuration set of 243 (3 by 3 by 3 by 3 by3) various robot faces were made (see Appendix. D).

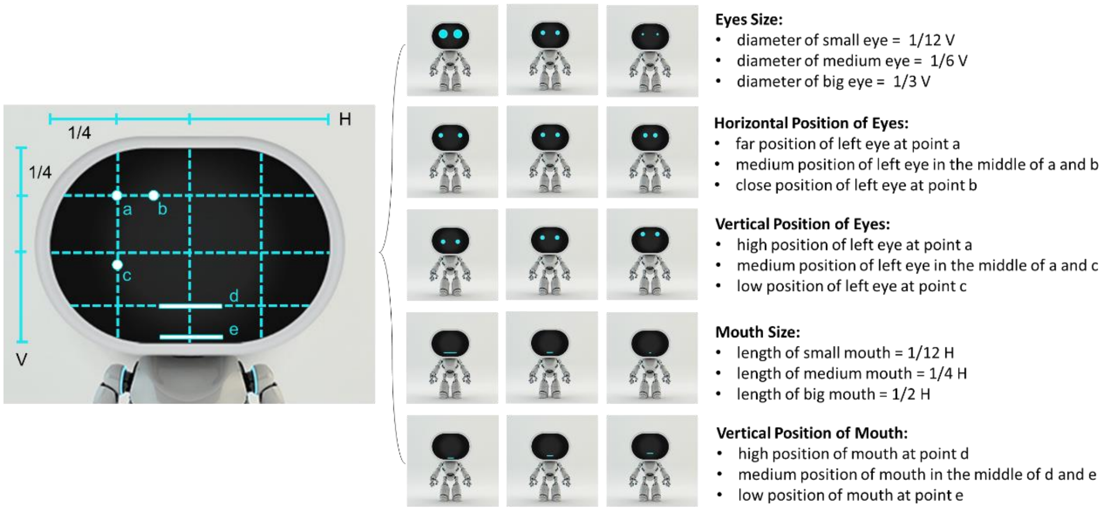


Figure 7.2 The detailed facial feature metrics of the social robot in study 4

In addition, not only the facial features of the social robot could influence trustworthiness evaluation, but also people's prior robot experience might have an impact on this process. For instance, as an emerging creature in our daily lives, most people might not be familiar with it (De Rie, 2016). To deal with uncertainties in knowledge, beliefs, and reasoning of users, Bayes infers user states from prior theory and data, updated by observations (Kardes et al., 2004). Therefore, prior experience on the social robot would act as an influential factor, which should be controlled to have a more precise examination.

7.2.2 Participants and Experiment Procedure

A mixed experiment was designed with the size of the eye and mouth as between-subjects variables and vertical and horizontal positions of eyes and mouth as within-subject variables. A total number of 270 participants were enrolled to take part in this research via the Amazon Mechanical Turk platform (AMT). To specify, the average age of this sample was 36.49 years (SD = 10.498). Table 7.1 showed detailed demographic information.

Table 7.1 The demographic information of the sample for the study of feature combination in study 4

Index	Frequency	Percentage	Index	Frequency	Percentage
Age			**Education Level**		
17-25	24	8.9%	High school graduates or lower	22	8.1%
26-30	79	29.2%	Some college	71	26.3%
31-35	59	21.9%	College graduate or above	177	65.6%
36-40	29	10.7%			
41-	79	29.3%	**Robot Experience**		
			Never used before	165	61.1%
Education			0-1 year use (exclusive for 1 year)	58	21.5%
Male	174	64.4%	1-2 years use (exclusive for 2 years)	36	13.3%
Female	96	35.6%	More than 2 years use	11	4.1%

For the manipulation check, people were asked the extent of agreement on five statements via a nine-point Likert scale when exposed to the certain robot design (I think this robot's eyes are big; this robot's mouth is big; this robot's eyes are positioned high on the face; this robot's eyes are spaced far apart; this robot's mouth is positioned low on the face). Regarding the measurement items on facial anthropomorphic trustworthiness, people were asked to agree on the extent of five items via a nine-point Likert scale (Gorn et al., 2008).

After consenting to participate, 270 individuals were recruited and briefly introduced to the current study. Then, they were asked to provide the demographic information and randomly assigned to one of nine experiment scenarios with different sizes of eye and mouth (each scenario contained thirty individuals equally). In each scenario, they would expose to twenty-seven robot faces with various eye positions (three different vertical and three different horizontal positions) and mouth positions (three different vertical positions). Specifically, the sequence of twenty-seven robot faces was randomized to control the learning effect in within-subjects design (Bosmans and Baumgartner, 2005). For each stimulus, they were asked to pay attention to the robot face, complete the questionnaire, and the related modification checks. After finished the questionnaire, they were told that they have finished the experiment.

7.3 Data Analysis and Result

In order to examine the hypotheses regarding the effect of baby schema on facial anthropomorphic trustworthiness, SPSS was utilized to perform descriptive analysis, manipulation check, and five-way mixed ANOVA of facial feature size and position.

To examine the normality of univariate, we performed the kurtosis and skewness test on five items. Results suggested all the kurtosis and skewness of each item was within the threshold, suggesting a general normal distribution (Groeneveld and Meeden, 1984). Then, we conducted a descriptive analysis of different factors in this study (see Table 7.2)

In addition, manipulation check were performed by a mixed ANOVA (between-subjects variables: eye size and mouth size; within-subjects variable: eye and mouth positions), revealing all the five modifications were successful: robots with bigger eyes were considered as having bigger eyes (Mean = 6.86 vs. 5.59 vs. 3.96; $F(2, 267) = 42.95$, $p < 0.01$); robots with bigger mouth were considered as having bigger mouth (Mean = 5.32 vs. 4.62 vs. 3.17; $F(2, 267) = 16.98$, $p < 0.01$); robot's eyes positioned high were considered as positioned high (Mean = 4.99 vs. 4.00 vs. 3.81; $F(2, 538) = 101.23$, $p < 0.01$); robot's eyes spaced far apart were considered as positioned far apart (Mean = 5.71 vs. 4.13 vs. 3.68; $F(2, 538) = 162.74$, $p < 0.01$); robot's mouth positioned low were considered as positioned low (Mean = 6.68 vs. 4.97 vs. 4.12; $F(2, 538) = 180.66$, $p < 0.01$).

Table 7.2 A summary of descriptive analysis of different factors for the study of feature combinations in study 4

				95% Confidence Interval	
Factors	**Levels**	**Mean**	**SE**	**Lower Bound**	**Upper Bound**
Eye size	small	5.77	0.13	5.52	6.01
	medium	5.89	0.13	5.64	6.13
	big	6.20	0.13	5.95	6.45
Mouth size	small	5.84	0.13	5.60	6.09
	medium	6.07	0.13	5.82	6.32
	big	5.94	0.13	5.69	6.19
Eye height	low	5.83	0.08	5.66	5.99
	medium	6.07	0.07	5.93	6.21
	high	5.95	0.08	5.80	6.11
Eye width	close	5.79	0.08	5.63	5.95
	medium	6.17	0.07	6.03	6.30
	far	5.89	0.08	5.75	6.05
Mouth height	low	5.60	0.09	5.42	5.77
	medium	6.13	0.07	5.98	6.28
	high	6.12	0.07	5.98	6.27

Five items on trustworthiness were averaged and treated as a whole with a satisfactory Cronbach's alpha coefficient (0.98), suggesting a high consistency of the current measurement. Then, a five-way mixed ANCOVA (between-subjects variables: eye size and mouth size; within-subjects variable: eye and mouth positions; covariate variable: prior robot experience) was then conducted. Table 7.3 shows the summarized results of ANCOVA. We mainly focus on the main effect of feature size and position since they are theoretically relevant to the research questions at hand.

Table 7.3 A summary of main and significant interactions in the ANCOVA in study 4

Sources	**df**	**F-statistic**	**p-Value**	**Effect size**
Experience (EXP)	1.00	26.60	$p < 0.01$	0.09
Eye size (ES)	2.00	3.15	$p < 0.05$	0.02
Mouth size (MS)	2.00	0.79	0.45	0.01
Eye height (EH)	2.00	13.89	$p < 0.01$	0.05
Eye width (EW)	2.00	41.75	$p < 0.01$	0.14
Mouth height (MH)	2.00	53.19	$p < 0.01$	0.17
EW * ES	4.00	13.61	$p < 0.01$	0.09
MH * ES	4.00	5.01	$p < 0.01$	0.04
EH * EW	4.00	3.37	$p < 0.01$	0.01
EH * MH	4.00	28.21	$p < 0.01$	0.10
EW * MH	4.00	6.95	$p < 0.01$	0.03

EH * EW * MH	8.00	3.29	$p < 0.01$	0.01
EH * MH * ES	8.00	2.23	$p < 0.05$	0.02
EW * MH * ES	8.00	3.31	$p < 0.01$	0.02
EW * ES * MS	8.00	2.93	$p < 0.01$	0.04

As for the main effect of feature size, we have found a significant impact of eye size ($F(2, 260) = 3.15$, $p < 0.05$, $\eta2 = 0.02$) and robot experience ($F(1, 260) = 26.60$, $p < 0.01$, $\eta2 = 0.09$) while the effect of mouth size was nonsignificant ($F(2, 260) = 0.79$, $p = 0.45$, $\eta2 = 0.01$). Specifically, (1) Robots with bigger eye size enjoyed a higher level of trustworthiness: A post-hoc Bonferroni corrected comparisons showed the significant difference between big and small eye size ($p < 0.05$) while the difference of eye size between big and medium and between medium and small was not significant (see Table 7.4); (2) There seems a nonsignificant difference between different mouth size: people showed no clear preference for a specific mouth size; and (3) we can also find a significant effect of the covariate variable, robot experience: people with more robot experience have an increased tendency to trust a robot. Thus, H 7.1 was supported while H 7.2 was not supported.

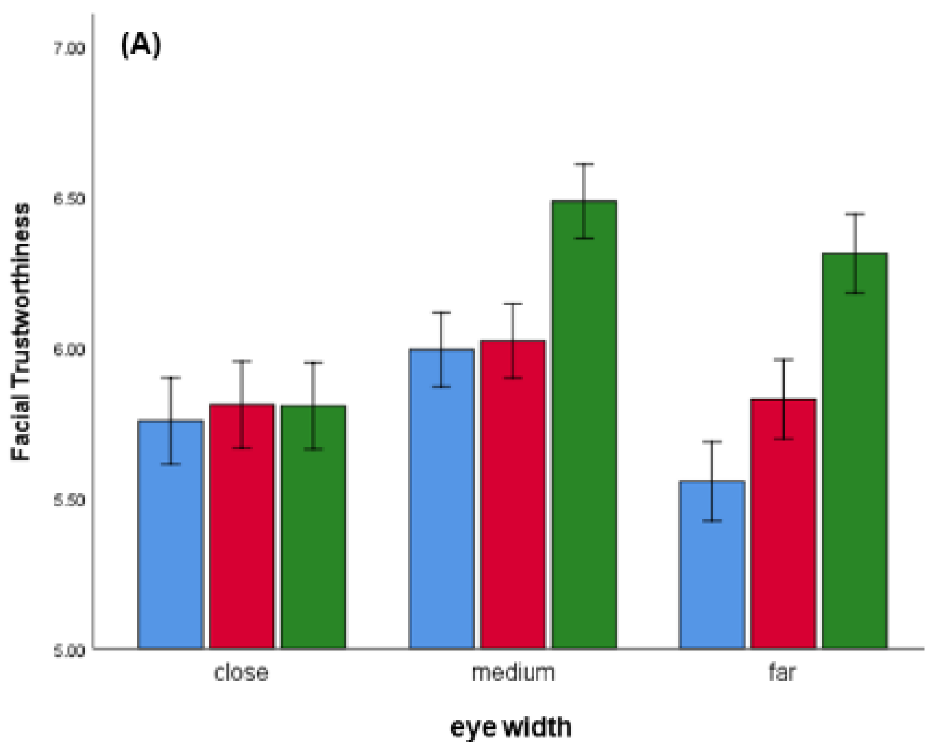
Table 7.4 A summary of the comparison of different levels and hypotheses testing in study 4

			95% Confidence Interval		
	Difference	**SE**	**Lower Bound**	**Upper Bound**	**Hypothesis**
Eye size					
small vs. medium	-0.12	0.18	-0.55	0.31	H1 was supported
small vs. big	-0.43*	0.18	-0.86	0.00	
medium vs. big	-0.31	0.18	-0.74	0.11	
Mouth size					
small vs. medium	-0.22	0.18	-0.66	0.21	H2 was not supported
small vs. big	-0.09	0.18	-0.52	0.34	
medium vs. big	0.13	0.18	-0.30	0.56	
Eye height					
low vs. medium	-0.25**	0.05	-0.36	-0.13	H3 was supported
low vs. high	-0.13	0.05	-0.26	0.00	
medium vs. high	0.12*	0.05	0.01	0.23	
Eye width					
close vs. middle	-0.38**	0.04	-0.47	-0.28	H4 was supported
close vs. far	-0.11	0.05	-0.22	0.01	
middle vs. far	0.27**	0.04	0.18	0.36	
Mouth height					
low vs. medium	-0.53**	0.06	-0.68	-0.38	H5 was supported
low vs. high	-0.52**	0.06	-0.67	-0.37	
medium vs. high	0.01	0.04	-0.08	0.10	

Note: * denotes difference significant at 0.05; ** significant at 0.01

The significant interaction of eye size with eye width ($F(4, 520) = 13.61$, $p < 0.01$, $\eta2 = 0.10$) and of eye size with mouth height ($F(4, 520) = 18.83$, $p < 0.01$, $\eta2 = 0.04$)

demonstrated that although increased eye size with an inward tendency of facial features could generally enjoy a high level of trustworthiness, this desire might be counteracted by extreme close displacement of facial features (big eyes with the close horizontal position of eye or big eyes with a high vertical position of the mouth), resulting in a declining trustworthiness perception (see Figure 7.3).



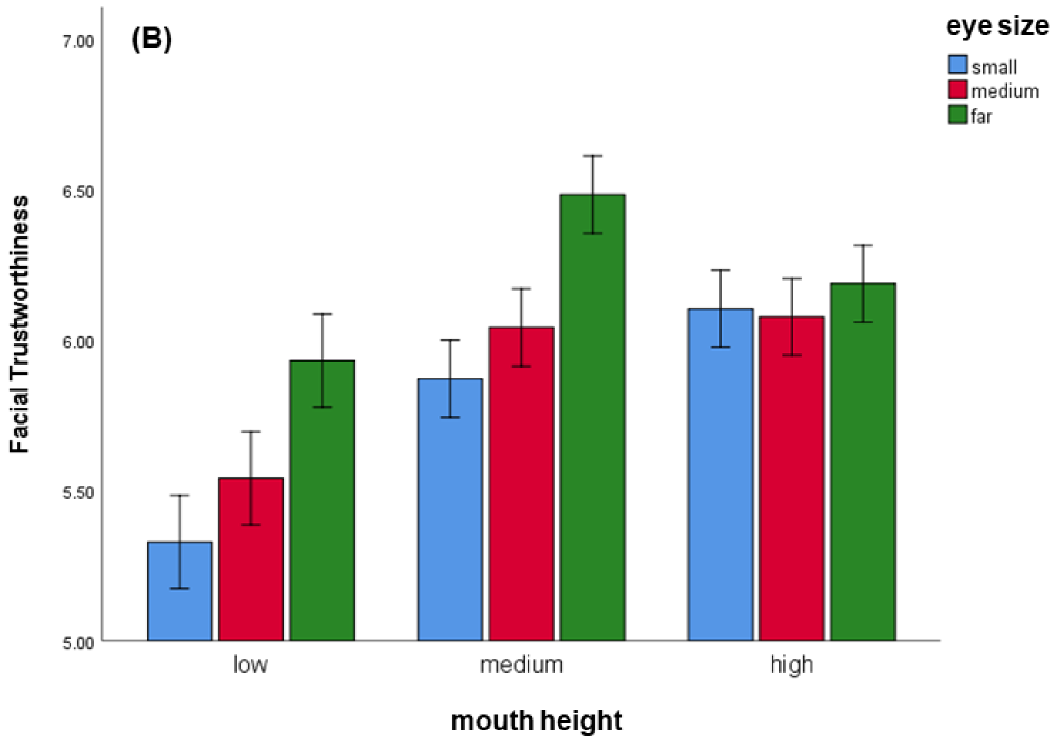
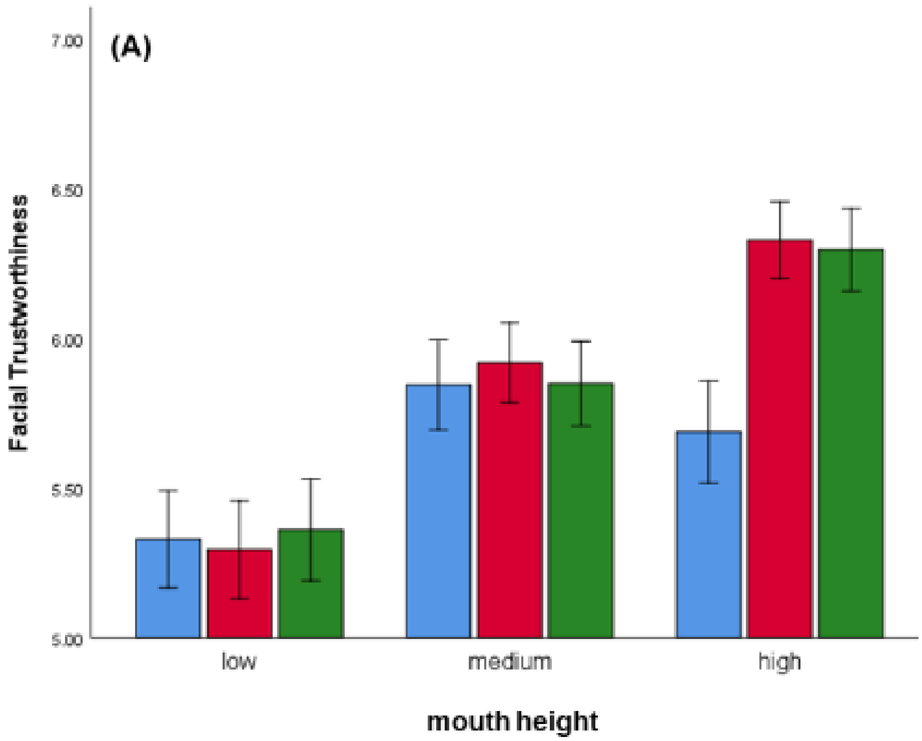


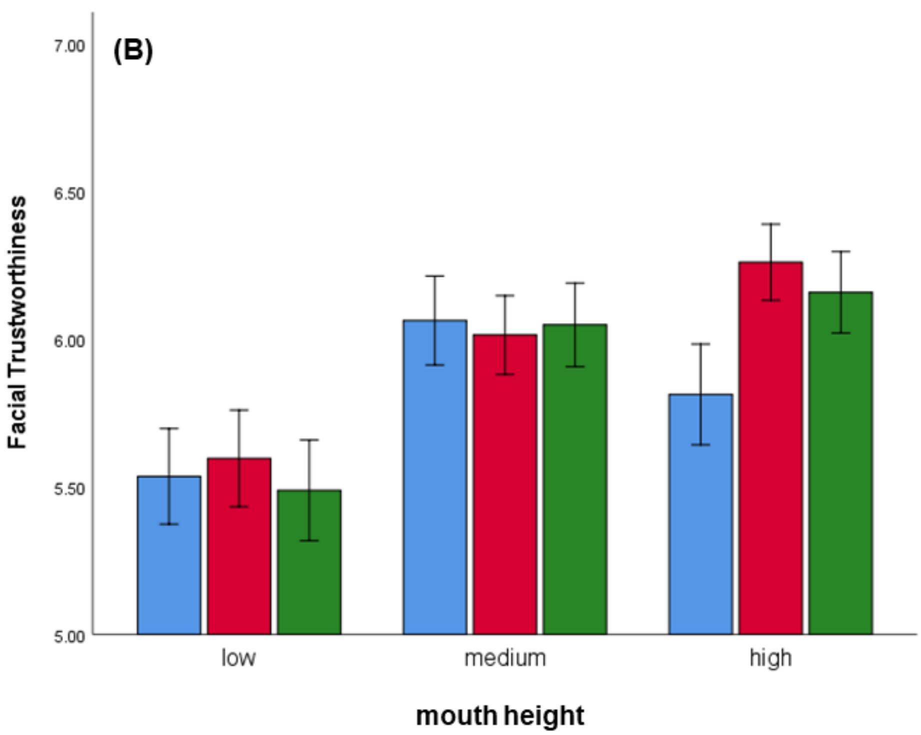
Figure 7.3 Bar chart representing the results of the 2-way interactions in facial trustworthiness ratings in study 4: (A) eye width x eye size; (B) mouth height x eye size.

Note: Error bars represent ±1 SE.

Similar observations could also be found at three-way interactions of eye size with eye width and mouth height ($F(8, 1040) = 3.31$, $p < 0.01$, $\eta 2 = 0.02$). When the mouth was moved upward (from bottom to medium), people would experience an increased level of trustworthiness when eyes were displaced inward (horizontally positioned from far to medium), regardless of eye size. When mouth was continued shifting upward (from medium to high), people would show their reluctance to the concentrated facial features (horizontally positioned from medium to close), especially regarding big eye size. Correspondingly, eye size similarly interacted with eye height and mouth height ($F(8, 1040) = 2.23$, $p < 0.05$, $\eta^2 = 0.02$), revealing as the eye size increases, people's trust toward high mouth keeps decreasing (See Figure 7.4 & 7.5).

Last, additional investigation was performed to analyze the correlation relationship between demographics and facial anthropomorphic trustworthiness, results of the Pearson test showed no significant correlation between facial anthropomorphic trustworthiness and age ($p = 0.58$) or gender ($p = 0.52$).


(A)
7.00
6.50
6.00
5.50
5.00
Facial Trustworthiness
low
medium
high
mouth height


(B)
7.00
6.50
6.00
5.50
5.00
Facial Trustworthiness
low
medium
high
mouth height

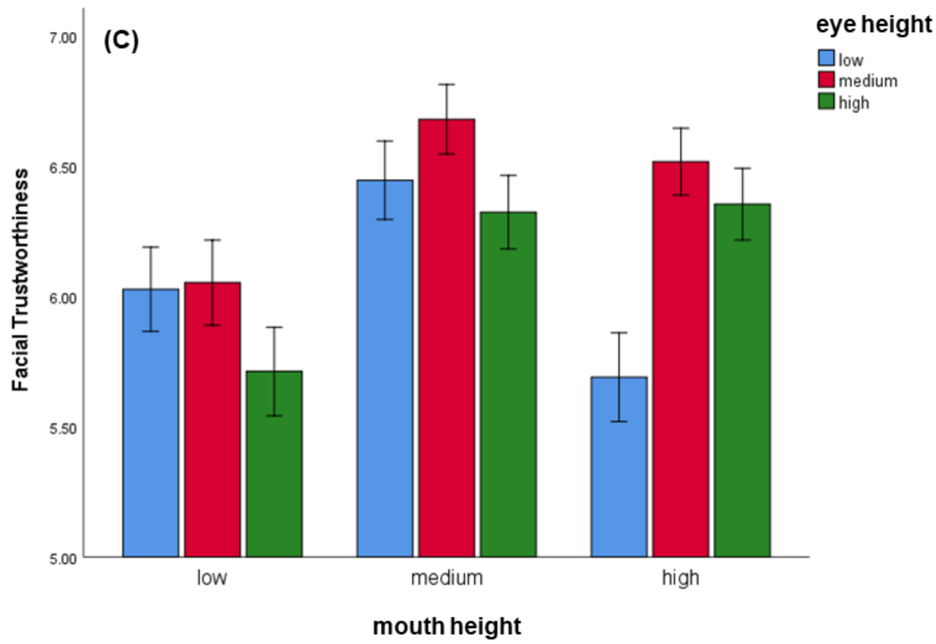
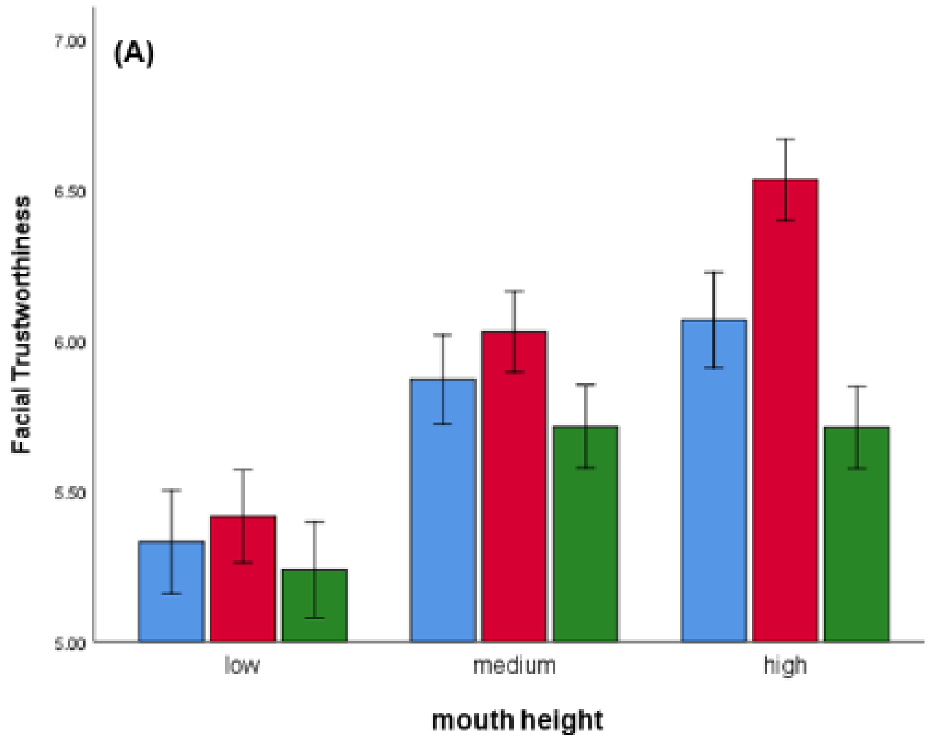
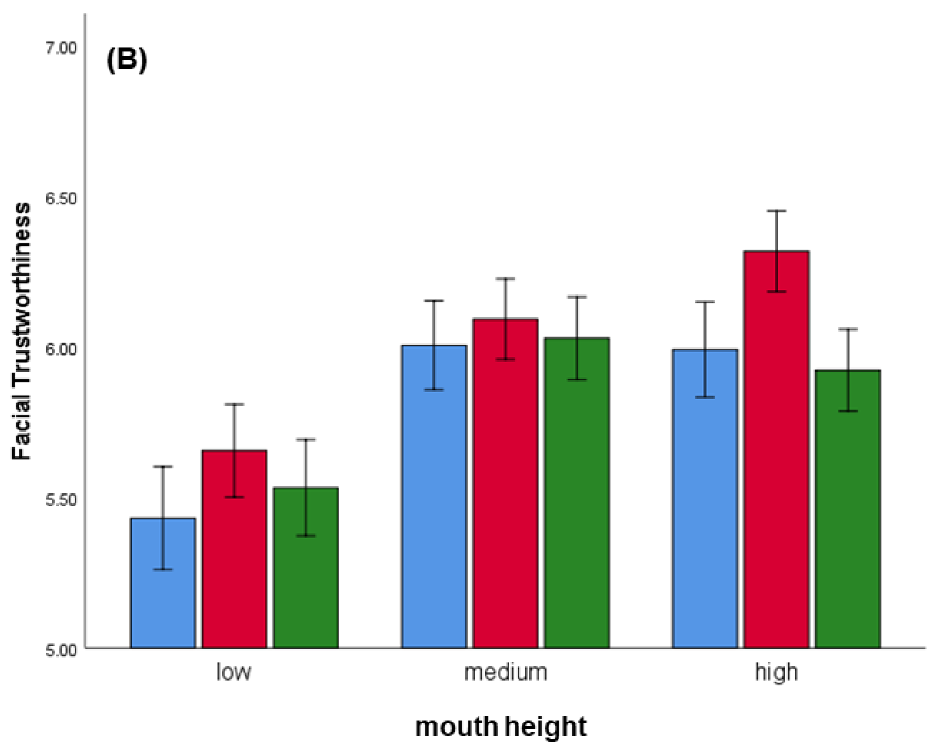


Figure 7.4 Bar chart representing the results of the 3-way interactions in facial trustworthiness ratings in study 4: (A) mouth height x eye height at small eye size; (B) mouth height x eye height at medium eye size; (C) mouth height x eye height at big eye size.

Note: Error bars represent ±1 SE.


(A)
7.00
6.50
6.00
5.50
5.00
Facial Trustworthiness
low
medium
high
mouth height


(B)
7.00
6.50
6.00
5.50
5.00
Facial Trustworthiness
low
medium
high
mouth height

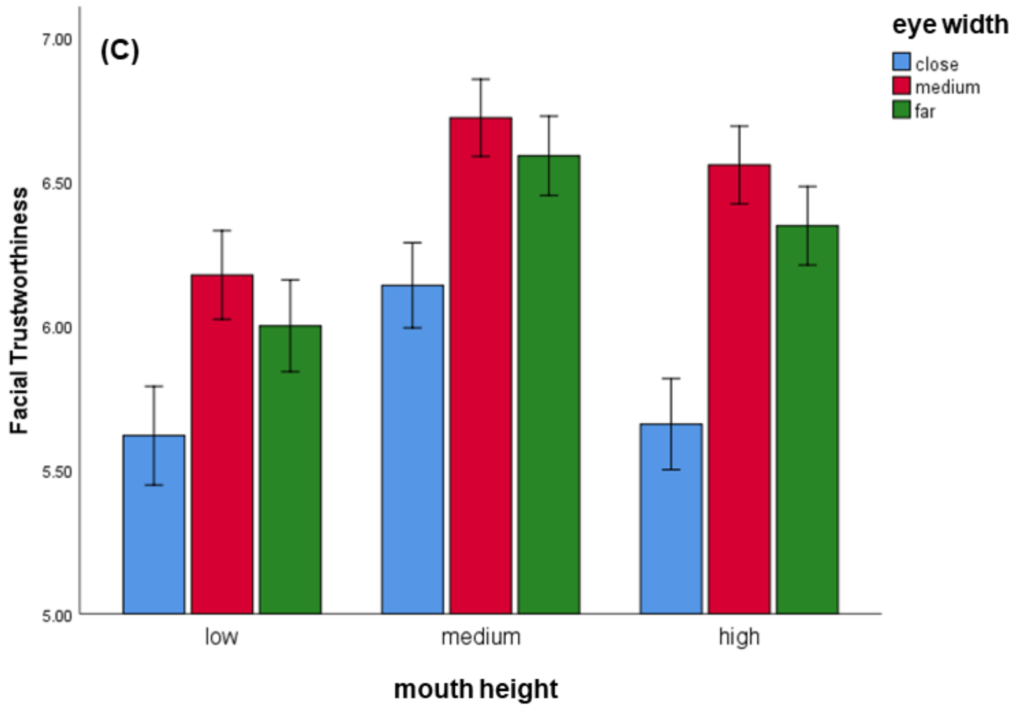
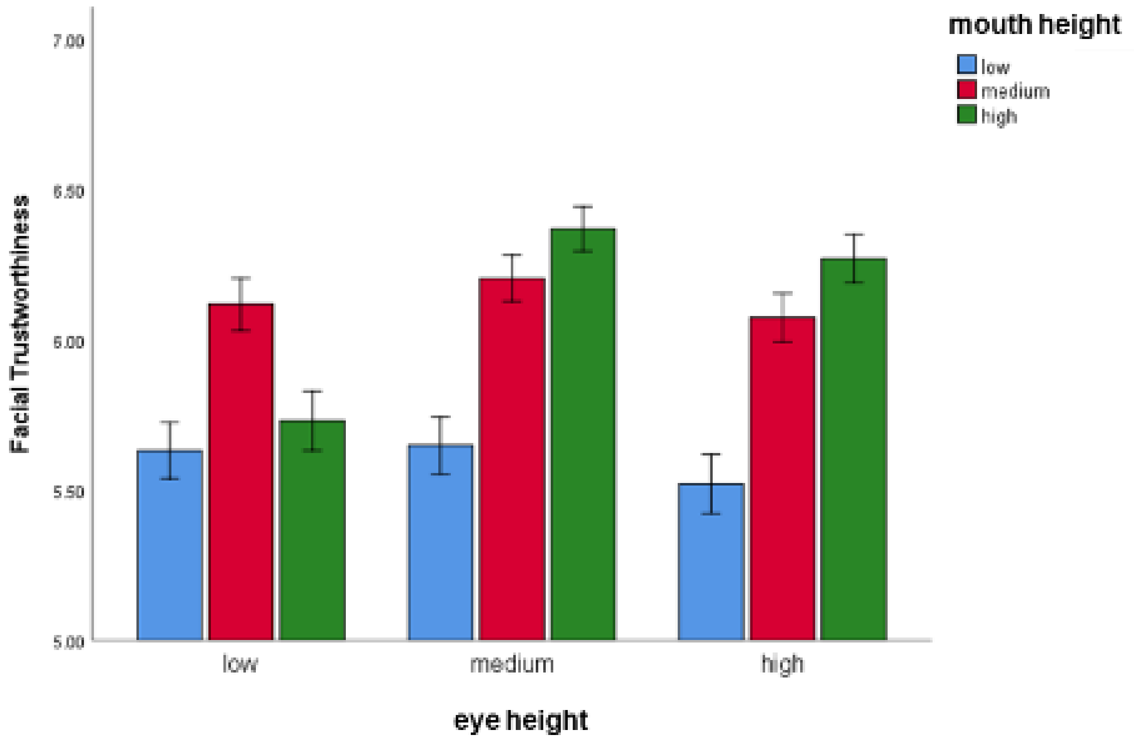


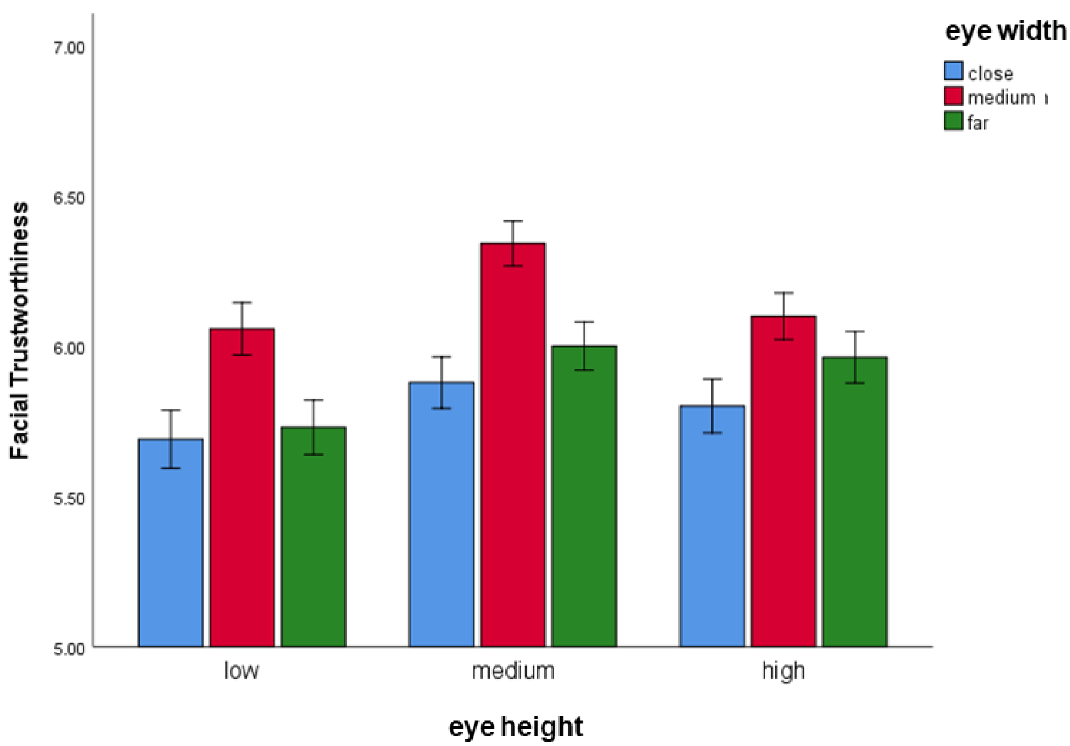
Figure 7.5 Bar chart representing the results of the 3-way interactions in facial trustworthiness ratings in study 4: (A) mouth height x eye width at small eye size; (B) mouth height x eye width at medium eye size; (C) mouth height x eye width at big eye size.

Note: Error bars represent ±1 SE.

As depicted in Table 7.3, the result of ANCOVA revealed a strong main effect for eye height ($F(2, 520) = 13.89$, $p < 0.01$, $\eta 2 = 0.05$), eye width ($F(2, 520) = 41.75$, $p < 0.01$, $\eta 2 = 0.14$), and mouth height ($F(2, 520) = 53.19$, $p < 0.05$, $\eta 2 = 0.17$): a general inward tendency for feature displacement was associated with an increased level of facial trustworthiness, however, extreme concentrated feature displacement might have a counteracted effect on facial trustworthiness. In other words, although the inward tendency of feature displacement might lead to more facial trustworthiness, people did not desire a "huddled" feature displacement (i.e. centered eye position with a high vertical mouth). Thus, H 7.3, H 7.4, and H 7.5 were all supported.

We could also find similar observations at the interaction effects among feature positions. As expected, eye height interacted with mouth height, $F(4, 1040) = 28.21$, $p < 0.01$, $\eta 2 = 0.10$; eye height interacted with eye width, $F(4, 1040) = 3.37$, $p < 0.01$, $\eta 2 = 0.01$; eye width interacted with mouth height, $F(4, 1040) = 6.95$, $p < 0.01$, $\eta 2 = 0.03$, showing people generally have an increased level of trustworthiness toward the centering direction of eye and mouth positioning (see Figure 7.6). However, as the distance between eye and mouth was getting closer, people's facial trustworthiness perception would decrease. Moreover, a reliable three-way interaction of eye height, mouth width, and mouth height also exemplified this conclusion, $F(8, 2080) = 3.29$, $p < 0.01$, $\eta 2 = 0.01$. No other theoretically significant effects were observed.


mouth height
low
medium
high
7.00
6.50
6.00
5.50
5.00
Facial Trustworthiness
low
medium
high
eye height


eye width
close
medium
far
7.00
6.50
6.00
5.50
5.00
Facial Trustworthiness
low
medium
high
eye height

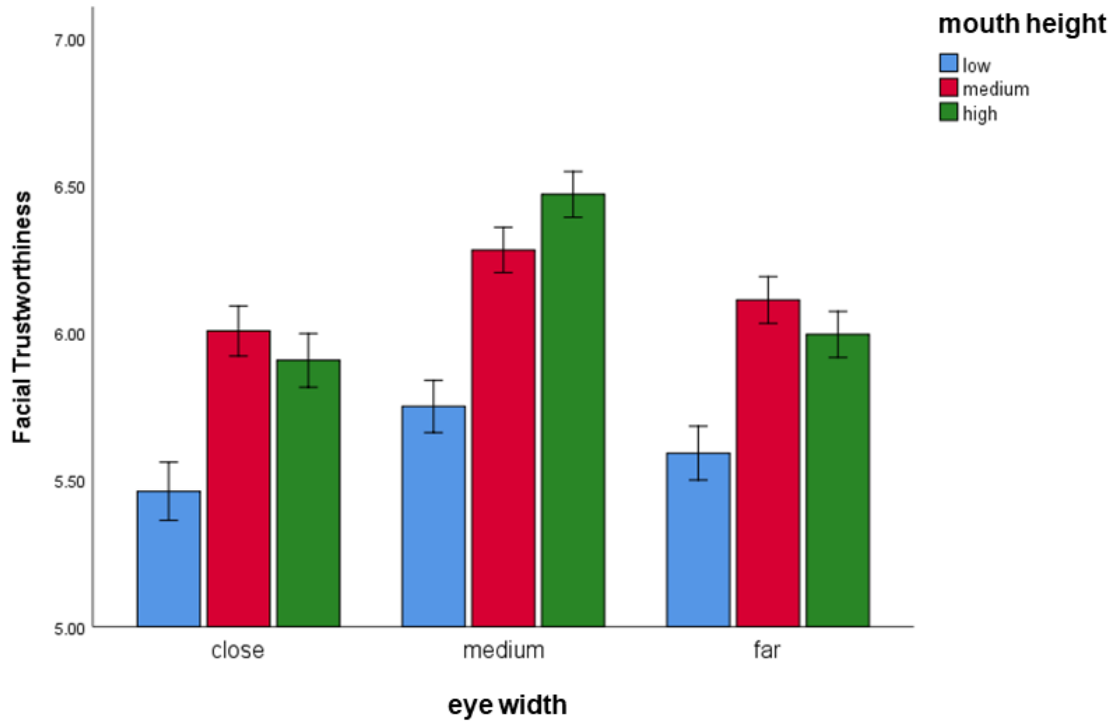


Figure 7.6 Bar chart representing the results of the 2-way interactions in facial trustworthiness ratings in study 4: (A) eye height x mouth height; (B) eye height x eye width; (C) eye width x mouth height.

Note: Error bars represent ±1 SE.

According to the results of ANCOVA, we could reveal the top three examples of social robots with the most and least trustworthy facial features (see Figure 7.7). In order to further validate whether the examples of the social robot could indeed be perceived as more babyish. We additionally recruited 180 participants (Mean age =

37.43, SD = 12.02; 118 males and 62 females; As for education level, 7 reported high school graduates or lower, 39 reported some college, and 134 reported college graduate or above. Regarding the robot use experience 74 participants never used it, 46 participants had 0-1 year use (exclusive for 1 year), 42 participants had 1-2 years use (exclusive for 2 years), and 18 participants had more than 2 years use) from the same resource and conducted a between-subject one-way ANCOVA (the six examples were treated as between-subject independent variables, robot experience was treated as a covariate, and perceived babyishness was treated as dependent variables). The perceived facial babyishness was measured on a 9-point Likert scale (a single item measure: I think this robot looks baby-faced; Berry, 1991).

Results of one-way ANCOVA revealed a significant effect of different examples, $F(5, 173) = 11.28$, $p < 0.01$, $\eta 2 = 0.25$, indicating people faced with trustworthy robots might experience a significant higher level of facial babyishness (Mean = 6.89; SD = 2.07) than those faced with untrustworthy robots (Mean = 4.62; SD = 2.10).

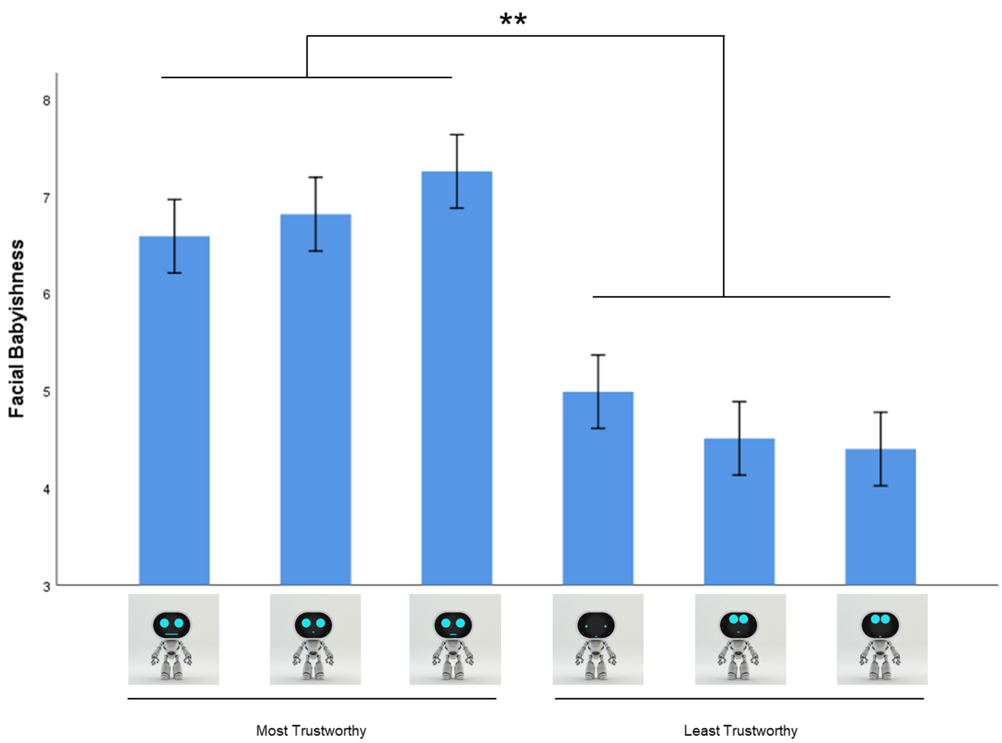


Figure 7.7 Bar chart representing the results of the 1-way ANCOVA in facial babyishness ratings in study 4

Note: Error bars represent ±1 SE. ** significant at 0.01.

7.4 Summary and Discussion

The current study tried to examine the effect of major facial anthropomorphic features on trustworthiness in HRI. Through a full-factorial mixed experiment, the result of this study indicated: (1) eye size has a significant impact on facial anthropomorphic trustworthiness: big eye could enjoy a relatively high level of facial anthropomorphic trustworthiness; (2) mouth size does not have a significant impact on facial anthropomorphic trustworthiness: either big mouth or medium mouth could enjoy a relatively high level of facial anthropomorphic trustworthiness; (3) eye positions have a significant impact on facial anthropomorphic trustworthiness: medium vertical and horizontal position of eye could enjoy a relatively high level of facial anthropomorphic trustworthiness; (4) mouth positions have a significant impact on facial anthropomorphic trustworthiness: both high and medium vertical position of mouth could enjoy a relatively high level of facial anthropomorphic trustworthiness; (5) people generally trust a social robot with inward facial features, however, those features should not be "huddled together".

Study 4 tries to contribute to the literature of trustworthiness in HRI from three perspectives. To begin with, this study provided preliminary evidence that certain facial features of baby schema could be applied and extended into the context of social robots to improve facial anthropomorphic trustworthiness toward the social robot. Moreover, this study further examined the effect of close displacement of facial features, which rarely appears on the human face. Considering the flexibility of rendering faces, it might be relatively convenient to have different combinations of feature size and positioning. Few works have attempted to explore extreme feature displacement in an anthropomorphic medium, i.e. social robot, thus the results of this study might facilitate our understanding of peculiar features in signaling facial trustworthiness. Last, although prior research has discussed the relationship between facial feature displacement and perceived trustworthiness, it might focus more on the one-dimension effect of separate facial features and neglect to analyze the effect of co-dependency on anthropomorphic trustworthiness. Through a mixed experiment design, the current work tries to analyze both separate and interaction effects of feature size and positioning, providing a relatively comprehensive understanding of eye and mouth in communicating facial trustworthiness.

CHAPTER EIGHT. STATIC FEATURE MODELING IN COMMUNICATING FACIAL ANTHROPOMORPHIC TRUSTWORTHINESS

This chapter conducts a modeling study to examine the effect of static features on facial anthropomorphic trustworthiness. Through modeling various static features, this chapter empirically investigates their influence on perceived trustworthiness, the difference for ethnic groups, and for anthropomorphic trustworthiness prediction. Based on the result of a series of experiments, the model enjoys satisfactory reliability, contributing to our knowledge of static features in influencing facial anthropomorphic trustworthiness.

8.1 Introduction

Through the results of studies 2-4, the effect of internal features, external features, and their combinations on facial anthropomorphic trustworthiness has been empirically examined. The objective of this chapter is three-fold: Firstly, the dataset from previous studies was utilized to model the relationship between different features and facial anthropomorphic trustworthiness in a relatively holistic way.

Second, based on the model, the results from prior social robot design and predict the social robot design, which does not be explored before, were predicted and validated. Thus, additional participants were recruited to validate the predicted values. Last, this chapter may also explore whether cultural differences have an impact on facial anthropomorphic trustworthiness. In other words, this study would further validate the conclusion from human facial trustworthiness that facia trustworthiness enjoyed a universal evaluation pattern: both Chinese and Caucasians shared similar cues in facial trustworthiness judgments (Xu et al., 2012).

8.2 Method

8.2.1 Modeling and Validation Rationale

We followed a structural rationale for modeling and validation study (see Figure 8.1).

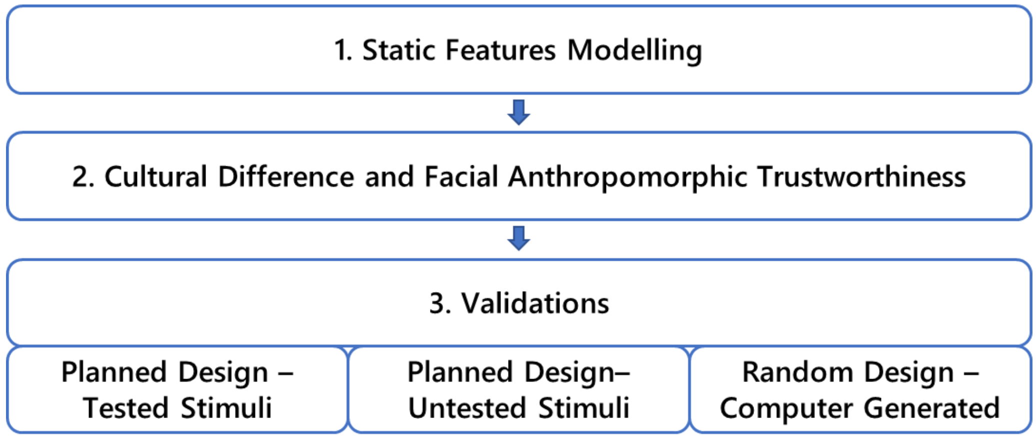


Figure 8.1 The overview of research model of study 5

Firstly, the dataset was extracted from studies 2-4 and was merged as a whole for further analysis (991 observations in total; Age = 36.36, SD = 10.69; 61% Male). Moreover, based on the configurations from previous studies 2-4, the current modeling study followed the same configuration patterns. To be more specific, fWHR ranged from 1/2 to 2; distribution was coded 1 for centered and 2 for scattered distribution; eye shape was coded 1 for narrow and 2 for round shape; eye size ranged from 1 to 4; mouth size ranged from 1 to 6.

A stepwise modeling regression was conducted to model the static features and provided the 95% prediction interval for different configurations. With regard to the validation process, it was divided into five different steps (see Figure 8.2). To specify, social robot design was categorized into the planned design and random design.

For planned design (step 1-4 in Figure 8.2), it focused on the stimuli within the previous settings (discrete configurations from studies 2-4): it contained the tested stimuli (for validating the previous results) and untested stimuli (for confirming the prediction validity). Accordingly, validation with planned design stimuli has two-fold objectives: 1) to validate whether the rating values from the validation experiment is consistent with the 95% prediction interval of the model; 2) to examine the difference between the US and Chinese participants in evaluating facial anthropomorphic trustworthiness.

For random design (step 5 in Figure 8.2), it focused on the stimuli beyond the previous setting (continuous configurations from studies 2-4). Planned design (both tested and untested stimuli) was still within the theoretical configuration, however, some parameters for the facial feature could have continuous, rather than discrete, configurations. With the help of an Excel random number generator, random configurations were generated for validation. Thus, additional participants were recruited (in step 5) to investigate the reliability of the prediction for the model.

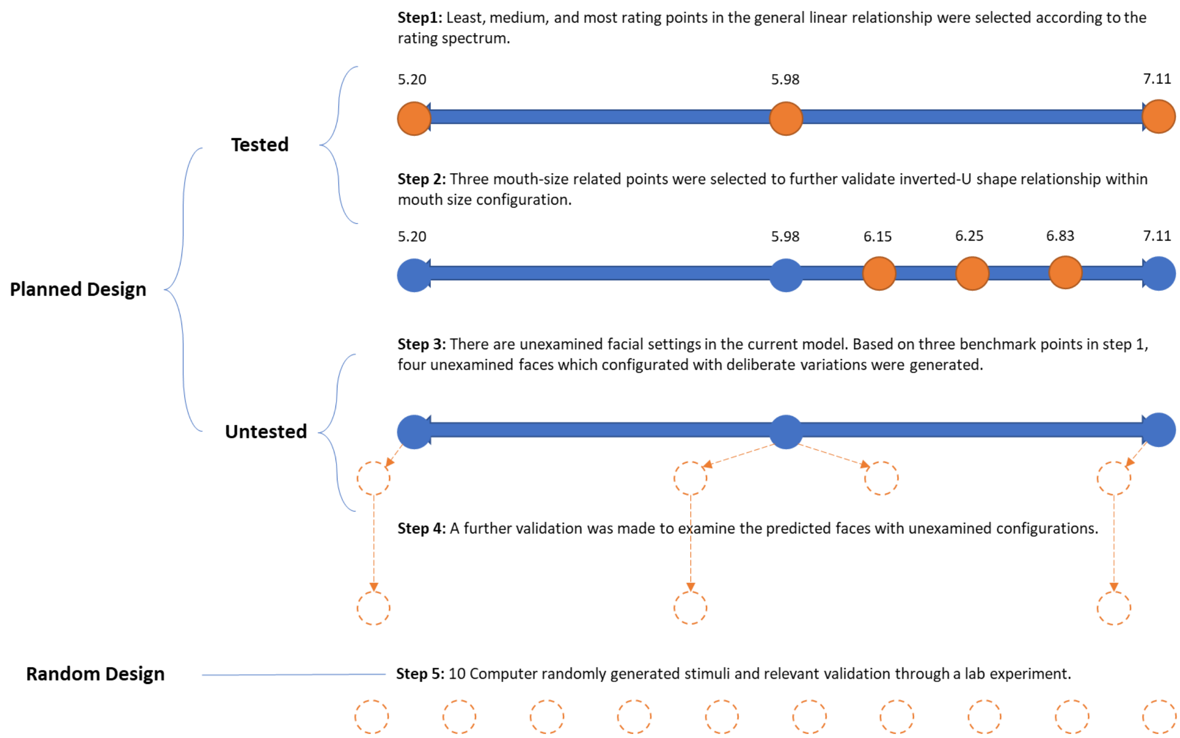


Figure 8.2 The rationale for validation process in five steps in study 5

8.2.2 Validation Stimuli and Experiment Design

Followed the validation rationale in Figure 8.2, different stimuli were generated with specific configurations as shown in Figure 8.3.

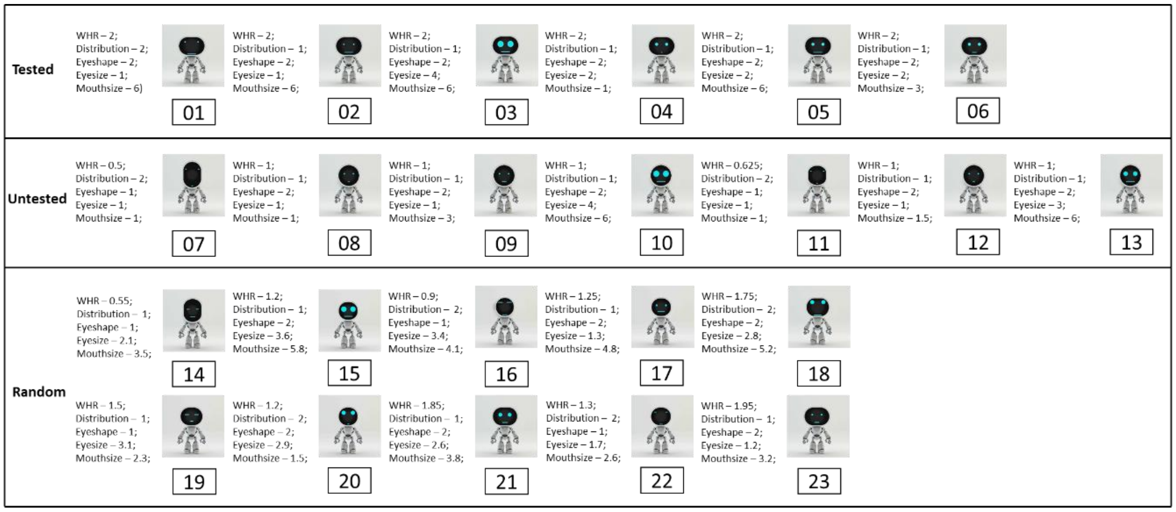


Figure 8.3 The stimuli with specific configurations in study 5

To specify, a one-way experiment was designed with 13 planned designs and 10 random designs as within-subjects variables while facial anthropomorphic trustworthiness was the dependent variable (Gorn et al., 2008).

8.2.3 Participants and Experiment Procedure in Validation Study

The planned design scenario recruited 60 participants from China and the US (30 Chinese participants were recruited via Wenjuanxing while 30 American participants were recruited from AMT). Random design scenario recruited additional 35 Chinese participants from the same source.

For the experiment procedure, after consenting to participate, individuals were recruited and briefly introduced to the current study. Then, they were asked to provide demographic information. In each scenario, they would expose to thirteen different stimuli. Specifically, the sequence of robot design (thirteen for planned design and ten for random design) was randomized to control the learning effect in within-subjects design (Bosmans and Baumgartner, 2005). For each stimulus, they were asked to pay attention to the robot face and complete the questionnaire. After finished the questionnaire, they were told that they have finished the experiment. Table 8.1 showed the detailed demographic information of participants in study 5.

Table 8.1 The demographic information of the sample in study 5

Group	Index	Values	Percentage
Planned Design (Chinese Participants)	**Age**		
	18-20	8	26.67%
	21-22	12	40.00%
	23-25	6	20.00%
	26-30	2	6.67%
	31-	2	6.67%
	Mean	22.67	
	Standard Deviation	3.58	
	Gender		
	Male	13	43.33%
	Female	17	56.67%
	Education Level		
	High school graduates or lower	0	0.00%
	Some college	3	10.00%

	College graduate or above	27	90.00%
	Robot Experience		
	Never used before	17	56.67%
	0-1 year use (exclusive for 1 year)	8	26.67%
	1-2 years use (exclusive for 2 years)	3	10.00%
	More than 2 years of use	2	6.67%
Planned Design (American Participants)	**Age**		
	18-20	1	3.33%
	21-22	2	6.67%
	23-25	4	13.33%
	26-30	4	13.33%
	31-	19	63.33%
	Mean	34.17	
	Standard Deviation	8.48	
	Gender		
	Male	20	66.67%
	Female	10	33.33%
	Education Level		
	High school graduates or lower	5	16.67%
	Some college	10	33.33%
	College graduate or above	15	50.00%
	Robot Experience		
	Never used before	13	43.33%
	0-1 year use (exclusive for 1 year)	10	33.33%
	1-2 years use (exclusive for 2 years)	4	13.33%
	More than 2 years of use	3	10.00%
Random Design (Chinese Participants)	**Age**		
	18-20	9	25.71%
	21-22	19	54.29%
	23-25	4	11.43%
	26-30	3	8.57%
	31-	0	0.00%
	Mean	21.60	
	Standard Deviation	1.87	
	Gender		
	Male	17	48.57%

Female	18	51.43%
Education Level		
High school graduates or lower	0	0.00%
Some college	3	8.57%
College graduate or above	32	91.43%
Robot Experience		0
Never used before	18	51.43%
0-1 year use (exclusive for 1 year)	11	31.43%
1-2 years use (exclusive for 2 years)	5	14.29%
More than 2 years of use	1	2.86%

8.3 Data Analysis and Result

8.3.1 Data Analysis and Result for Static Feature Modeling

In order to conduct modeling analysis and consequent validations on facial anthropomorphic trustworthiness, SPSS was utilized to perform the stepwise regression for modeling analysis and one-way ANOVA for validation research.

As for modeling analysis, significant factors were revealed and added in the stepwise regression in a sequence (factors: fWHR, distribution, eye shape, mouth size, a quadratic term of mouth size): fWHR showed a positive relationship with facial anthropomorphic trustworthiness; people would have a higher level of facial anthropomorphic trustworthiness toward an inward feature distribution (medium vertical and horizontal positions of eyes and mouth) than scattered feature distribution (high vertical and horizontal positions of eyes and low vertical position of the mouth); round eye or big eye might enjoy a higher rating in facial anthropomorphic trustworthiness; mouth size and facial anthropomorphic trustworthiness followed an inverted-U shape relationship: both big and small mouth could have a relatively low level of trustworthiness. Followed the suggestions of stepwise regression (Thompson, 1995), a quadratic term of mouth size was also introduced in the model, thus, six factors were all added in the modeling. Therefore, the merged dataset was utilized and stepwise modeled to illustrate the effect of different features on facial anthropomorphic trustworthiness. Table 8.2 shows the summarized results of the stepwise regression using standardized and unstandardized coefficients (β).

Table 8.2 The summarized result of stepwise regression in study 5

Model	**1**	**2**	**3**	**4**	**5**	**6**
fWHR	0.09***	0.15***	0.18***	0.14***	0.14***	0.16***
Distribution		-0.20***	-0.25***	-0.26***	-0.26***	-0.24***
Eye shape			0.17***	0.17***	0.17***	0.20***
Eye size				0.12***	0.12***	0.13***
Mouth size					-0.01	0.36**
Mouth size^2						-0.38***

Note: * denotes significant at 0.05; ** significant at 0.01; *** significant at 0.001

See equations for unstandardized models 1-6:

1) Facial Trustworthiness = 5.37 + 0.18 fWHR

2) Facial Trustworthiness = 6.03 + 0.27 fWHR – 0.79 Distribution

3) Facial Trustworthiness = 4.12 + 0.35 fWHR – 0.98 Distribution + 0.99 Eye shape

4) Facial Trustworthiness = 4.10 + 0.27 fWHR – 1.04 Distribution + 0.99 Eye shape + 0.20 Eye size

5) Facial Trustworthiness = 4.10 + 0.27 fWHR – 1.04 Distribution + 0.99 Eye shape + 0.20 Eye size - 0.002 Mouth size

6) Facial Trustworthiness = 2.94 + 0.31 fWHR – 0.96 Distribution + 1.13 Eye shape + 0.22 Eye size + 0.42 Mouth size – 0.06 Mouth size^2

Consistent with previous studies, the results, as a whole, indicated that the robots with higher fWHR ($\beta = 0.16$, $p < 0.001$) and round eyes ($\beta = 0.20$, $p < 0.001$) tended to have a higher level of facial anthropomorphic trustworthiness. Also, the eye size ($\beta = 0.13$, $p < 0.001$) and feature distribution ($\beta = -0.24$, $p < 0.001$) were found to be associated with facial anthropomorphic trustworthiness positively and negatively respectively. Moreover, mouth size and perceived trustworthiness followed an inverted-U shape: a positive relationship with mouth size ($\beta = 0.36$, $p < 0.01$) but a negative relationship with its quadratic term ($\beta = -0.38$, $p < 0.001$). Then, this model was utilized to make a 95% prediction interval for planned and random design with specific configurations, as suggested in Figure 8.3.

8.3.2 Data Analysis and Result for Validation Study

A validation study (steps 1-5) recruited 95 participants were conducted accordingly (see Figure 8.2). To specific, as for planned design for a social robot (step 1-4), a one-way ANOVA was conducted with 13 planned stimuli. Based on the modeling analysis, 95% prediction intervals were calculated for each stimulus. Table 8.3 shows: 1) the observed values from previous studies 2-4; 2) the results (60 individuals as a whole) from the current validation experiment; 3) the trustworthiness evaluation difference between American (30 individuals) and Chinese participant

Results showed high reliability of prediction for the model: all the results from the current experiment were consistent with the 95% prediction intervals for each configuration. In addition, one-way ANOVA further confirmed no significant difference between American and Chinese in evaluating 13 robot stimuli from steps 1-4: Chinese and Americans might share similar patterns or perceptions in evaluating facial anthropomorphic trustworthiness for various robots.

Table 8.3 The summary of the prediction of steps 1-4 and the comparison of different ethnical groups in study 5 (N = 60)

Steps	Stimuli	Observed Values	95% Prediction Interval	Results	US	CN	US-CN Sig
Step 1	1	5.20	4.83-5.47	5.13	5.38	4.89	ns

	2	5.98	5.80-6.42	6.05	6.27	5.83	ns
	3	7.11	6.79-7.38	6.92	7.04	6.81	ns
	4	6.15	5.99-6.55	6.09	6.25	5.94	ns
Step 2	5	6.25	6.05-6.61	6.16	6.28	6.11	ns
	6	6.83	6.46-6.83	6.65	6.79	6.51	ns
Step 3	7	-	2.35-3.72	3.28	3.65	2.92	ns
	8	-	5.43-6.05	5.59	5.62	5.57	ns
	9	-	5.59-6.01	5.93	6.15	5.72	ns
	10	-	5.71-6.60	6.22	6.31	6.13	ns
Step 4	11	-	2.45-3.78	3.01	3.19	2.83	ns
	12	-	5.29-6.85	5.56	5.71	5.41	ns
	13	-	5.54-6.32	6.19	6.20	6.17	ns

In order to have further reliability testing, an additional experiment was conducted with 10 random stimuli (step 5). A similar process was performed, and Table 8.4 shows the 95% prediction interval and the results of step 5. Results indicated that 9 entries were within the 95% prediction interval (90% successful rate), suggesting satisfactory reliability of the current modeling.

Table 8.4 The summary of the prediction of step 5 in study 5 (N = 35)

Steps	Stimuli	95% Prediction Interval	Results
Step 5	14	4.16-5.16	4.17
	15	5.79-6.56	6.31

16	3.58-4.74	3.49
17	5.75-6.17	6.01
18	5.31-5.84	5.77
19	4.83-5.66	4.92
20	4.82-5.49	5.36
21	6.50-6.90	6.79
22	4.49-5.45	4.54
23	6.25-6.64	6.55

8.4 Summary and Discussion

Study 5 tried to model the effect of static features on anthropomorphic trustworthiness and further validate the model and cultural difference between Chinese and Americans. Through a stepwise modeling analysis, this study has provided the preliminary modeling metrics of different levels for static facial features in determining anthropomorphic trustworthiness.

Results of modeling analysis were consistent with the previous studies 2-4:

1) fWHR showed a positive relationship with facial anthropomorphic trustworthiness (effect size = 0.16, $p < 0.001$) in which high fWHR leaded to a high level of facial anthropomorphic trustworthiness;

2) Feature distribution revealed a negative relationship with facial anthropomorphic trustworthiness (effect size = -0.24, $p < 0.001$) where people would have a higher level of facial anthropomorphic trustworthiness toward an inward feature distribution (medium vertical and horizontal positions of eyes and mouth);

3) Eye shape unveiled a positive relationship with facial anthropomorphic trustworthiness (effect size = 0.20, $p < 0.001$) where round eyes might enjoy a higher rating in facial anthropomorphic trustworthiness.

4) Eye size also indicated a positive relationship with facial anthropomorphic trustworthiness (effect size = 0.13, $p < 0.001$) where big eyes might enjoy a higher rating in facial anthropomorphic trustworthiness.

5) Mouth size suggested an inverted-U shape relationship with facial anthropomorphic trustworthiness (mouth size, effect size = 0.36, $p < 0.01$; quadratic term, effect size = -0.38, $p < 0.001$).

With regard to validations, a series of validation experiments (steps 1-5) was conducted to examine the reliability of the current model and further explore the cultural difference in influencing facial anthropomorphic trustworthiness. According to the results of steps 1-4, there is no statistically significant difference for Chinese and American to facial anthropomorphic trustworthiness evaluation, suggesting American and Chinese tended to share a similar pattern in perceiving the trustworthiness of robot faces. More importantly, almost all the results were within the 95% prediction interval estimated by the model (only one, configuration 16, was slightly outside of the interval), reaching a 95.6% accuracy rate (22/23*100%). Thus, it showed that the current model achieved relatively high reliability for predicting facial anthropomorphic trustworthiness with different configurations.

One thing worth noting is the role of culture in signaling human facial trustworthiness. As trustworthiness evaluation plays a paramount role in almost every ethnic group, Xu et al., (2012) indicated that people would use facial cues as an intuitive tool for social judgments, especially meeting at the first time when extensive personality information is missing. As suggested in chapter 3, previous literature has shown a nuanced influence of facial anthropomorphic trustworthiness. On the one hand, different ethnic groups, e.g. Chinese vs. Caucasian (Xu et al., 2012) or Caucasian vs. African vs. East Asian vs. South Asian (Birkás et al., 2014), tended to share and adopt similar perceptions of trustworthiness and attractiveness toward the same facial stimuli. On the other hand, different ethnic groups, i.e. Japanese/Israeli, might have preferences for faces of their own ethnicity (Sofer et al., 2017). The reason accounting for this phenomenon is that the former literature is exploring the effect of different ethnic groups on the same stimuli while the latter literature is focusing on the effect of various ethnic groups on the different stimuli (contained stimuli of their own ethnicity and other ethnicities). In the current experiment setting, different ethnical groups, Chinese and American in particular, are responding to the same robot facial stimuli. Thus, in accordance with the previous research that different ethnic groups have similar trustworthiness perception toward the same stimuli (Birkás et al., 2014; Xu et al., 2012), the current study expands human face to robot face, contributing to the literature of AI agent.

CHAPTER NINE. DYNAMIC FEATURES IN COMMUNICATING FACIAL ANTHROPOMORPHIC TRUSTWORTHINESS

This chapter explores the effect of dynamic features on facial anthropomorphic trustworthiness in different daily contexts. Through exploring the fine-grained nature of emotional expressions, valence, and arousal, this chapter empirically investigates their specific influence for perceived trustworthiness and their interaction with context with a different regulatory focus in shaping facial anthropomorphic trustworthiness. Through two experiments (Study 6 and 7), the chapter discusses the potential optimum combination of dynamic expressions and contexts with a different regulatory focus, contributing to our knowledge of dynamic features in influencing facial anthropomorphic trustworthiness. Besides, study 7 also compared the scale from Gorn (2008) and study 1, suggesting both scales could enjoy sufficient reliability and validity to predict facial anthropomorphic trustworthiness

9.1 Introduction

Facial expressions and their associated emotions play a crucial role in interpersonal interaction (Weiβ et al., 2019). Previous research on the dynamic features mainly concentrates on two facial areas: the eye/brow and mouth region. To specify, positive emotions, excited or relaxed expressions, in particular, is usually featured with an upturned U-shaped mouth while lip corners and eyebrows are raised while negative emotions, depressed, or afraid expressions, in particular, are characterized by an inverted U-shaped mouth while lip corners and eyebrows are lowered (Calvo et al., 2017; Gill et al., 2014; Johnston et al., 2010; Ma et al., 2015). When expressing positive emotions, the mouth region might attract more visual attention which works as a salient predictor for facial happiness and trustworthiness (Calvo et al., 2017). However, when expressing negative emotions, that observer might rely less on the mouth region but more on the eye/brow region for additional information (Calvo et al., 2019).

Generally speaking, facial valence might work as a significant predictor for anthropomorphic trustworthiness where the positive emotions are believed to have a high level of trustworthiness and visual attention while negative emotions are believed to have a low level of trustworthiness and visual attention (Gutiérrez-García and Calvo, 2016a; Oosterhof and Todorov, 2009; Woodall et al., 1980). However, as for visual arousal, few studies have explicitly explored the effect of facial arousal and found a nuanced effect on facial trustworthiness and visual attention (Chiller-Glaus et al., 2011; Karbauskaite et al., 2020; Sanchez and Vazquez, 2014; Weiβ et al., 2019). On the one hand, facial sadness and anger showed no difference in communicating trustworthiness attracting visual attention (Okubo et al., 2018; Sanchez and Vazquez, 2014). On the other hand, the intensity of the smile was significantly correlated with visual fixations (2019).

Besides, emotional expressions should also consider the associated contexts (Song and Luximon, 2020a). For example, when facing a prevention-focused scenario, such as warning information, Reed and DeScioli (1974) suggested the best combination (regulatory fit) would be warning information repressed by an emotional fearful expression, instead of a neutral face or even happy face. This indicated that negative expression could also promote trustworthiness in certain contexts when it fitted its scenario. Thus, regulatory focus theory, therefore, can serve as a theoretical framework providing helpful insights into message framing in daily contexts. Tailored messages demonstrating a different regulatory focus (promotion vs. prevention) can influence the target audience's trustworthiness evaluation in daily contexts, such as reminding. However, there is still a lack of empirical evidence on whether the emotional expressions of social robots and the contextual messages with different regulatory focus could also create the regulatory fit and promote facial anthropomorphic trustworthiness toward message senders in the context of human-robot interaction. The current study tries to investigate the role of a regulatory fit, using different dynamic facial expressions paired with a message describing a daily context, in determining the facial anthropomorphic trustworthiness toward the social robot. Specifically, we first analyze the dynamic expressions with different valence and arousal in communicating anthropomorphic trustworthiness in a context-free scenario (study 6); then, we examine the regulatory fit between different dynamic expressions and contexts in communicating facial anthropomorphic trustworthiness (study 7; see Figure 9.1). Based on the results of study 7, a dynamic features model was developed. Thus, the hypotheses are summarized as follows:

H 9.1: For facial valence, social robots with positive (vs. negative) facial expressions would be perceived as more trustworthy.

H 9.2: For visual arousal, social robot with active (vs. inactive) facial expressions would be perceived as more trustworthy.

H 9.3: For facial valence, social robots with positive (vs. negative) facial expressions would attract more visual attention (fixation duration/ counts).

H 9.4: For facial arousal, social robots with positive (vs. negative) facial expressions would attract more visual attention (fixation duration/ counts).

H 9.5: When individuals in promotion-focused contexts are exposed to positive expressions of social robots, the occurrence of regulatory fit will evoke an increased level of trustworthiness (a low level of EDA) toward social robots than when no regulatory fit occurs.

H 9.6: When individuals in prevention-focused contexts are exposed to negative expressions of social robots, the occurrence of regulatory fit will evoke an increased level of trustworthiness (a low level of EDA) toward social robots than when no regulatory fit occurs.

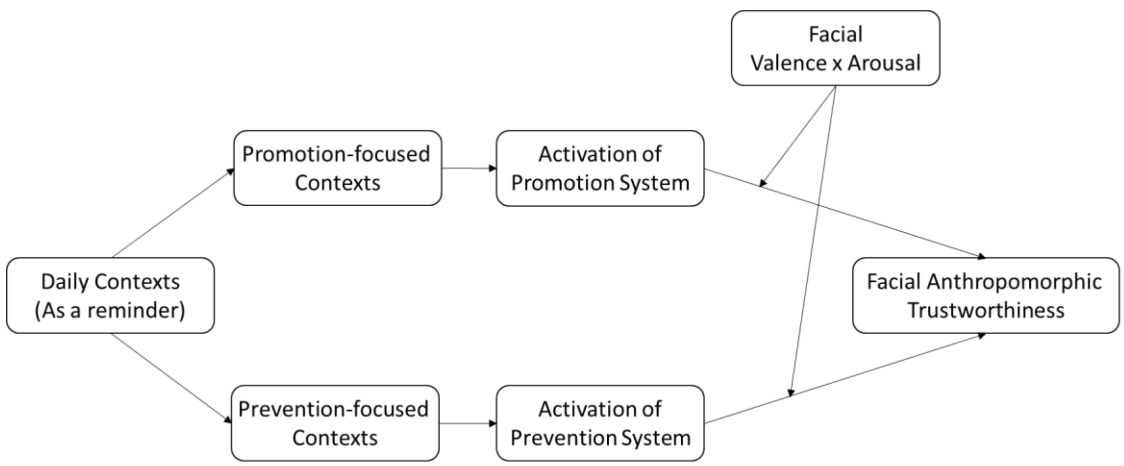


Figure 9.1 The overview of research model of study 6 and 7

9.2 Method

9.2.1 Physiological and Subjective Measures

This study utilized subjective ratings of facial anthropomorphic trustworthiness and physiological measures (eye-tracking and electrodermal activity) to explore facial anthropomorphic trustworthiness, visual patterns, and related stress during the human-robot interaction.

Eye-tracking is an apparatus of estimation of an individual's attention either from the direction of gaze (the position where an individual is looking at) or from the movement of eyes during a period of time (Riegelsberger et al., 2004). In order to measure it, an eye tracker work as a device for recording eye directions, positions, and motion in the given duration (Calvo et al., 2019). Indeed, it has been applied in various kinds of disciplines, such as marketing (Riegelsberger et al., 2004), psychology (Calvo et al., 2019), and HCI (Guo et al., 2019).

Within the field of facial analysis, eye-tracking is also widely used to analyze the relationship between emotional faces and people's attention (Ellingsen et al., 2019). To specify, fixation duration and its ratio are the most common indicators to measure visual attention when interacting with selective engagement (Pavlov et al., 2015). For instance, attractive images could enjoy a longer fixation duration than unattractive ones (Guo et al., 2019). A similar observation of the positive association between facial trustworthiness and visual attention could also be found in the field of facial evaluation: eye-tracking assessment works as an efficient tool to evaluate facial trustworthiness where a trustworthy face enjoys longer fixation duration, compared with a non-trustworthy face (Calvo et al., 2019; Ellingsen et al., 2019; Leder et al., 2011; Ma et al., 2015; Mollahosseini et al., 2018; Oh and Ju, 2020; Sanchez and Vazquez, 2014; Stanton and Stevens, 2017; Toet et al., 2017). To further examine the visual patterns for specific facial regions of a social robot, the ratios of fixation duration of a specific area of interest (AOI) for each expression was used to illustrate the featural attention allocation within the particular expression. Specifically, as depicted in Figure 9.4, four typical emotional expressions (afraid, depressed, excited, and relaxed expressions) were analyzed in study 6. Within each expression, the ratios of fixation duration of facial expressions were examined, as depicted in Figure 9.7.

Considering there was no substantial difference in fixation patterns between human and robot face detection (Palinko et al., 2015), the eye-tracking assessment could work as an effective tool to evaluate anthropomorphic trustworthiness for a social robot.

Besides, the electrodermal activity skin conductance (EDA) of the human body adjusts with the developments in the function of the skin sweat glands (Braithwaite et al., 2013). These measurable changes (the unit of measurement is micro-Siemens, μS) in skin electrical activity are also called skin conductance (SC). Figure 9.2 shows the EDA electrodes attaching to fingers (left).

Compared with heart rate and respiration (Stellar et al., 2015), EDA is especially a highly sensitive signal of stress or arousal in physical interaction since it is an indicator of sympathetic nerve activity (Haag et al., 2004). It is associated with the variation of the electrical conductance of the skin in response to sweat secretion (Ding et al., 2020). When the human body is stimulated by the outside or affected by psychological emotions, its sympathetic nerves are excited, and the sweat glands secrete sweat, which contains water and electrolytes, through its postganglionic nerve fibers, thereby increasing skin conductance, which could be recorded via an EDA device (Sharma et al., 2016). The mean of skin conductance (SC mean; unit:

micro-Siemens, μS) within a specific time domain was a common signal, which works as the indicator for measuring EDA (Ding et al., 2020). As depicted in Figure 9.2 (right), as the stimulus disappears, the skin conductance returns to the original level (Sharma et al., 2016).

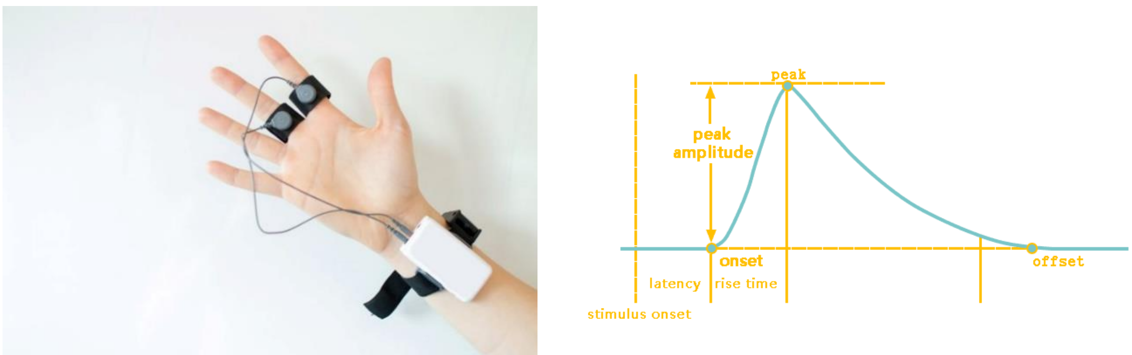


Figure 9.2 The setup (left) and procedure (right) of skin conductance response in study 7

EDA is especially appropriate long been applied in the area of trustworthiness research, such as deception (Pennebaker and Chew, 1985) and lie detection (Tomash and Reed, 2015). Prior research suggested the reasons (Aguado et al., 2011): when people meet and interact with those individuals they trust, they tend to be relaxed and comfortable, thus having a lower level of stress (lower EDA). On the contrary, when people meet and communicate with those individuals they distrust, they tend to stay alert and vigilant, thus having a higher level of stress (higher EDA).

In order to explore the effect of dynamic features on facial anthropomorphic trustworthiness, eye-tracking, and EDA, this study examines the effect of dynamic features via a robot prototype in a lab experiment. Hence, it is expected that both attention-based and electrode-based measures correlate substantially and specifically with the corresponding subjective reports.

9.2.2 Stimuli and Experiment Design for Study 6

With regard to stimuli preparation of dynamic expressions, we followed three stages to ensure that the dynamic expressions were accurate and appropriate (see Figure 9.3) was the first step was especially for study 6 and the second and third steps were for study 7.

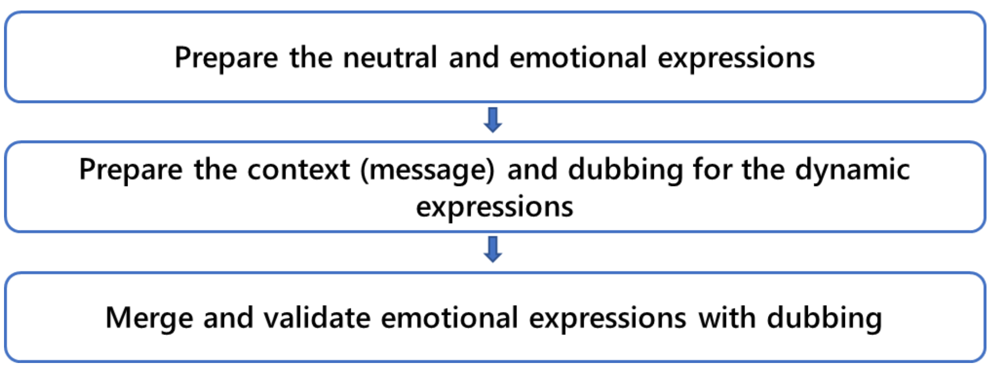


Figure 9.3 The stimuli preparation process for study 6 and 7

- Neutral and Emotional Expressions Preparations

An animation designer made all stimuli of dynamic expressions as requested (excited, afraid, depressed, and relaxed). Dynamic expressions were designed following human facial dynamic characteristics (Campellone and Kring, 2013; Gutiérrez-García and Calvo, 2016a; Hildebrandt and Fitzgerald, 1979; Kramer, 2015; Vesker et al., 2018). For example, positive expressions have an enlarged and upturned mouth, while negative expressions have a shrink downturned mouth with a frown. Similarly, we also discussed the presentational features of facial arousals, such as enlarged pupils (Q. Liu et al., 2019). Several versions of drafts were made and iterated to ensure the overall style as natural as possible. As a result, we agreed on and designed the dynamic facial transitions from a neutral expression to emotional expressions, which included four different emotional scenarios, two levels of valence (positive vs. negative) by two levels of arousal (arousal vs. non-arousal). They were used as the stimuli both for context-free scenarios and context scenarios (see Figure 9.4).

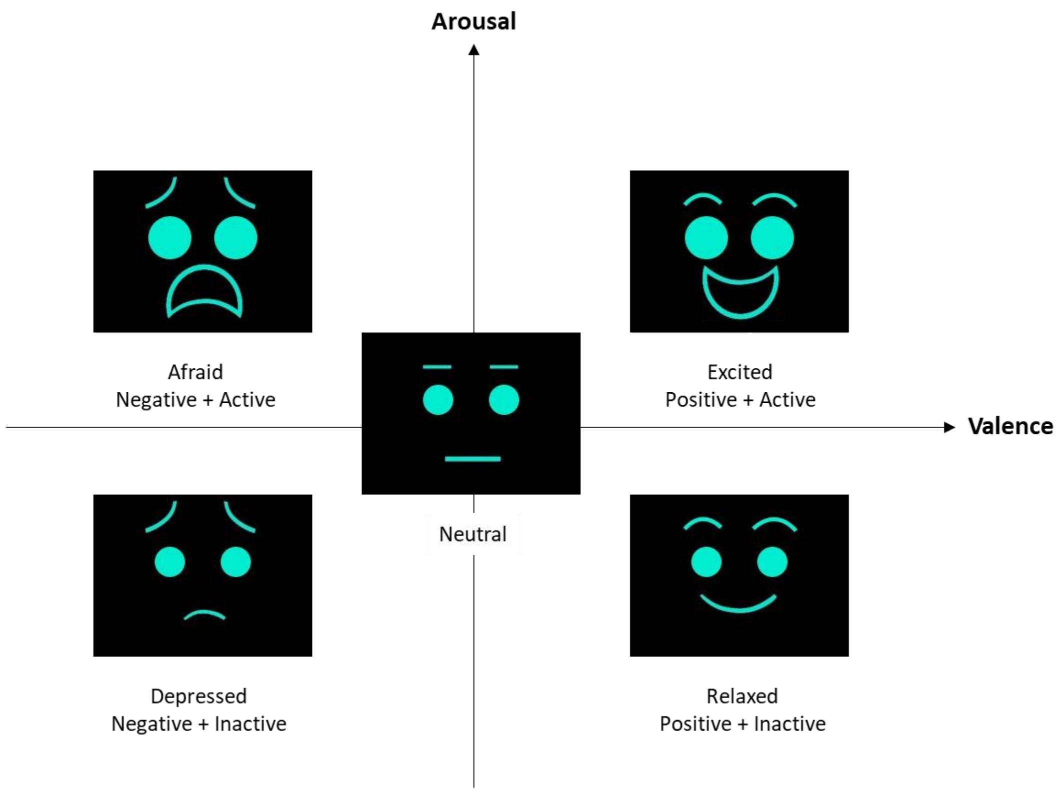


Figure 9.4 The neutral and dynamic expressions in study 6

9.2.3 Stimuli and Experiment Design for Study 7

- Context, Message, and Dubbing

After finishing the facial animations, four messages from specific contexts merged with different regulatory focus. We set the same social roles of the robot, as a reminder, in scenarios to control the potential confounding factor of a specific social role. To be more specific, the message of the contexts followed the same structure of reminding things (Hello – Current Situation – Reasons – Suggestions) with different degrees of regulatory focus (prevention-focused contexts for prevention-1 and prevention-2 and promotion-focused contexts for promotion-3 and promotion-4). Four scenarios focused on promotion-focused context and prevention-focused contexts were developed accordingly (see Table 9.1)

Table 9.1 The summary of contexts with a different focus in study 7

Focus	**Texts**
Prevention-1	“Hello! The kitchen gas is still on! It's dangerous if you don’t close it quickly! Turn it off!” “你好！厨房煤气还没关呢！不赶紧关危险很大呐！赶紧关吧！”
Prevention-2	“Hello! You had too much Cola today! Too much Cola is bad for your health! Stop drinking!” “你好！你今天喝太多可乐啦！喝太多对身体不好！别再喝啦！”
Promotion-3	“Hello! I have prepared water for you! Water is good for your health! Have some water!” “你好！我特意为你准备了水！多喝水对身体好哦！喝点水吧！”
Promotion-4	“Hello! Your pizza is here! It smells so good! Open the door and get it!” “你好！你的披萨外卖到了！闻起来味道好香啊！快开门拿吧！”

One professional voice actor with five-years relevant experience was recruited for dubbing. We carefully discussed and agreed on the tone and voice of a specific scenario. After a rehearsal several times, she dubbed her voice into a specific scenario appropriately. The duration for each scenario (around six seconds) stayed almost the same to exclude its potential cofounding effect of time.

- Finalization and Validation of Emotional Expressions

The dynamic expressions and dubbing were carefully and appropriately synchronized together, generating 16 different scenarios (two levels of valence by two levels of arousal by four levels of regulatory focus). Each scenario lasted for around seven seconds which included one second for facial dynamic transition (from neutral face to the given emotions) and six seconds for the action of a given scenario (emotional expressions). As a result, four emotional expressions were generated for study 6 while eight scenarios were generated for study 7 (see Table 9.2).

Table 9.2 The experiment scenarios summarization in study 6 and 7

Studies	**Valence**	**Arousal**	**Regulatory focus**
Study 6	Negative	Inactive	-
	Negative	Active	-
	Positive	Inactive	-
	Positive	Active	-
Study 7	Negative	Inactive	Prevention-1/2
	Negative	Active	Prevention-1/2
	Positive	Inactive	Prevention-1/2
	Positive	Active	Prevention-1/2
	Negative	Inactive	Promotion-3/4
	Negative	Active	Promotion-3/4
	Positive	Inactive	Promotion-3/4
	Positive	Active	Promotion-3/4

A pilot study was conducted for manipulation check of the stimuli. A convenience sample of 15 people (Mean age = 30.93, SD = 11.25; 6 male and 9 female) was invited to rate their agreement on three statements of valence, arousal and regulatory focus via a 9-point Likert scale. A within-subjective ANOVA was performed, revealing all the three manipulation of stimuli were successful: robots with positive expression were considered as having positive expression (Mean = 2.62 vs. 7.17; $F(1, 14) = 179.10$, $p < 0.01$); robots with arousal expressions were considered as having arousal expression (Mean = 2.53 vs. 6.95; $F(1, 14) = 168.35$, $p < 0.01$); promotion-focused context was considered as promotion-focused ($Mean_{prevention\text{-}1} = 2.13$; $Mean_{prevention\text{-}2} = 2.77$; vs. $Mean_{promotion\text{-}3} = 7.25$; $Mean_{promotion\text{-}4} = 7.43$; $F(3, 42) = 332.14$, $p < 0.01$).

9.2.4 Participants and Experiment Procedure

As for the main experiment, thirty-five students enrolled in a design discipline class at a major university in southern China participated in study 6 as partial fulfillment of a research participation requirement for the discipline. To specify, the average age of this sample was 21.37 years (SD = 1.31). Detailed demographic information in studies 6 and 7 could be found in Table 9.3.

Table 9.3 The experiment scenarios summarization in study 6 and 7

	Frequency	Percent		Frequency	Percent
Gender			**Robot interaction experience**		
Male	16	45.7%	Never	4	11.4%
Female	19	54.3%	0-1 year (1 year not included)	16	45.7%
			1-2 years (2 years not included)	8	22.9%
Age			2- year	7	20.0%
18-20	6	17.1%			
21-22	26	74.3%	**Education**		
23-25	2	5.7%	High school graduate or lower	0	0.0%
26-30	1	2.9%	Some college	0	0.0%
31-	0	0.0%	College graduate or above	35	100.0%

The main experiment was conducted in a lab where the physical environment was controlled stable (around 23.5 degrees Celsius) to minimize the environmental influence in eye-tracking and EDA recordings (see Figure 9.5). Participants were recruited to the lab to finish the test while others are not allowed to disturb during the experiment. As for the experiment procedure, it contained two steps. The first step was intended to examine the effect of four emotional expressions in a context-free scenario (study 6) while the later step was further to explore the effect of dynamic expressions in different contexts (study 7).

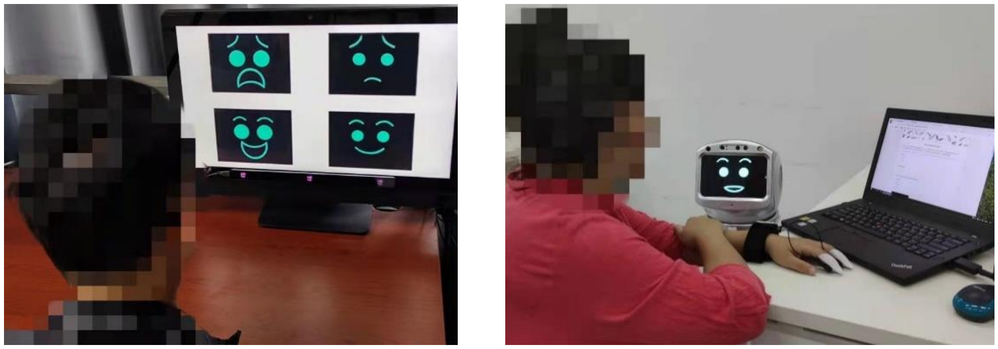


Figure 9.5 The process of an experiment in study 6 and 7: a participant was in study 6 (left), and a participant was in study 7 (right)

To specify, study 6 was a within-subjective designed experiment. After consenting to participate, individuals were instructed to sit in an adjustable chair and have briefly been introduced to the aim of the current study. Then, participants were asked to provide the demographic information, asked to view a set of robot faces on the screen, and was instructed to "look at the face as you normally trust". During this process, data of visual gaze was collected via a screen-based eye-tracking device (see Figure 9.6; X3-120, Tobii). A 23'' monitor was arranged around seventy centimeters before the participants. Before each trial, the eye-tracker was calibrated for every individual through a five-point calibration approach embedded in the eye-tracking program. To be more specific, the eye-tracker was adjusted and recalibrated till the eye-tracking error rates reach within 0.5 degrees of the visual angle (for both x-axis and y-axis). During each trial, the eye-tracker recorded the coordinates of gaze information on the monitor (frequency equals to 250 Hz).

Regarding recorded fixation, eye movements that existed above the minimal fixation duration, fifty milliseconds (ms), within one degree of the visual angle were considered as eye fixations. Any loss of signal, which was caused by blinking or off-screen gazes, was spontaneously screened. Four rectangle-shaped areas of interest (AOI) corresponded with four emotional expressions were also recognized and labeled for each trial (see Figure 9.6).

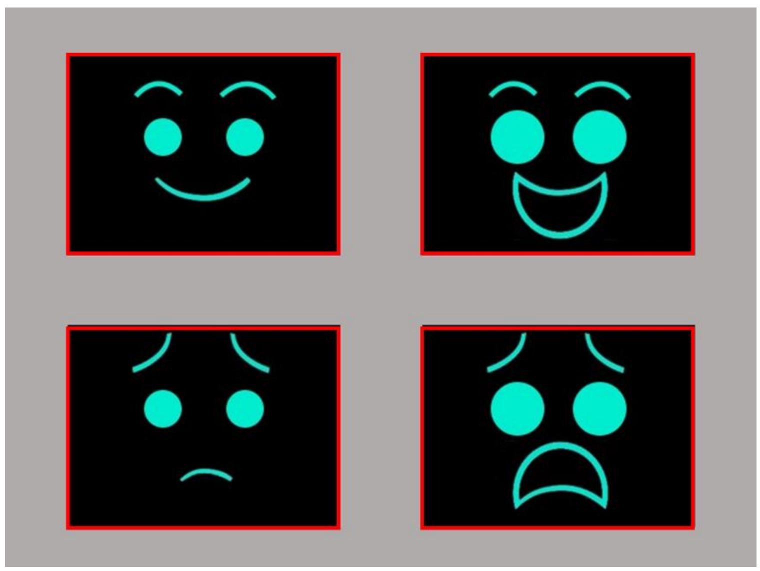


Figure 9.6 The AOI for the eye-tracking experiment

For study 6, each trial started with a blank screen for two to three seconds of duration, followed by a fixation cross located at the center of the monitor. To proceed to the main task, individuals needed to focus on the cross for a minimum of two seconds. The target image was presented for ten seconds (see Figure 9.7). Last, they were instructed to finish a questionnaire (contains both trustworthiness scales from Gorn (2008) and FATSR-17) on facial anthropomorphic trustworthiness for four emotional expressions.

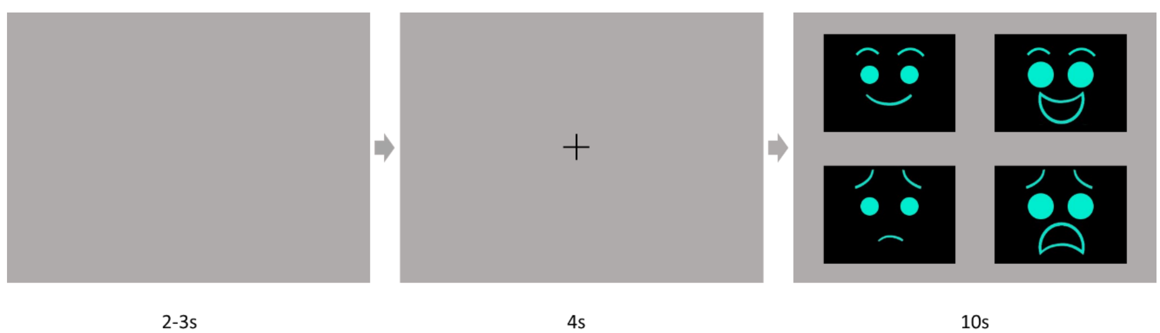


Figure 9.7 The experiment process in study 6

For study 7, participants were instructed to move to another area and sat at a distance of around sixty centimeters from the social robot in the same room. Electrodermal activity (EDA) was collected for the physiological signal, which was recorded and analyzed via the ErgoLAB human-machine-environment testing cloud platform (Kingfar International Inc., Beijing, China). A cotton swab with scrubbing cream was utilized to decrease the impedance of skin. Thus, EDA signals were collected from the index and middle fingers of the left hand with a sample rate of 32 Hz.

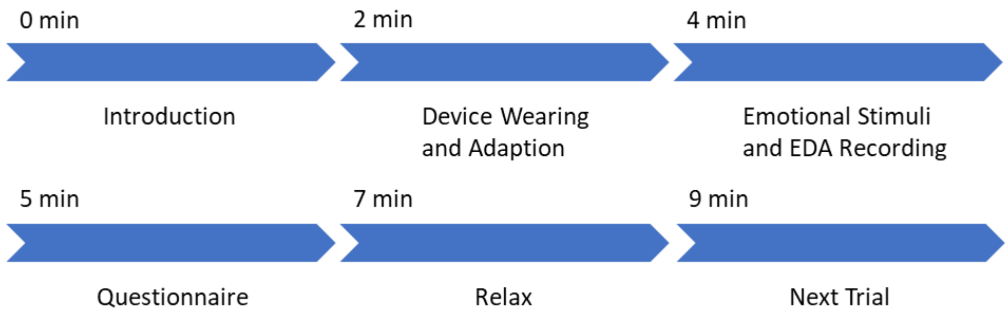


Figure 9.8 The experiment process in study 7

Figure 9.8 indicates the trial procedure. First of all, the individuals were instructed on the aim of this study. Subsequently, the investigator assisted them to equip the EDA recording device on their non-dominant hand, as suggested in Figure 9.2. Then, the individuals were given two minutes to get used to the device. During their period of familiarization, the transmission quality of EDA signals was checked to ensure the signal was well received and recorded. Later, the robot began to interact with participants in a given scenario. To specify, the participants were asked to sit on a comfortable chair while the robot is in front of them (around 50-60 cm apart). During each scenario, the social robot performed specific emotional expressions while introducing a specific context with a given regulatory focus. Each participant saw the robot with different expressions and focus in a randomized sequence. Participants pay attention to the facial expressions of the robot and the contextual information sent by the robot, during which their skin conductance was recorded. Then, participants were asked to complete the questionnaire to evaluate the facial anthropomorphic trustworthiness they had just experienced (FATSR-17). After two minutes relax, another trial was iterated and conducted in a similar process. The whole experiment lasted about 90 minutes. The environment and temperature for both studies were the same, and Mandarin was used throughout all the trials.

9.3 Data Analysis and Result

9.3.1 Scale Validation for Trustworthiness Scale and FATSR-17

As discussed in section 5.2.2, studies 2-5 adopted the trustworthiness scale from Gorn (2008) due to the parallel research during the period of exploration on the meaning of facial anthropomorphic trustworthiness and examination about the effect of specific features on trustworthiness evaluation. Accordingly, it would be necessary to compare the difference between FATSR-17 from study 1 and the trustworthiness scale from Gorn (2008) in study 6.

Following a theory-driven approach for validating the trust-related scale (Bhattacherjee, 2002), the AMOS software and confirmatory factor analysis (CFA) technique was adopted for scale validation between these two scales where five items from Gorn (2008) and seventeen items from study 1.

The skewness and kurtosis were firstly severed as the indicator to check the normal distribution of the responses from study 6. Results showed the responses followed a normal distribution: the value of skewness ranged from −0.40 to -0.17 while the value kurtosis ranged from −1.15 to 0.75, which were all within their threshold (Groeneveld and Meeden, 1984).

Further, the reliability and validity of all the factors were examined to check whether they have achieved adequate reliability and validity (J. Zhang et al., 2019). To specify, the composite reliability (CR) of all factors should exceed the threshold, 0.7; the average variance extracted (AVE) should exceed 0.5. Table 9.4 summarized all the Cronbach's Alpha, CR, AVE of all the factors, suggesting they have achieved adequate reliability and validity.

Table 9.4 The reliability and validity of two scales

Factors	**Cronbach's Alpha**	**Variable**	**Standardized Factor Loading**	**C.R. (t-value)**	**SMC**	**AVE**	**Composite Reliability**

Trustworthiness (Gorn, 2008)	0.979	TR1	0.940	-	0.883	0.903	0.925
		TR2	0.941	21.338	0.885		
		TR3	0.905	19.017	0.819		
		TR4	0.896	19.667	0.803		
		TR5	0.926	22.793	0.857		
Ethics Concern (Study 1)	0.960	EC1	0.910	-	0.827	0.96	0.936
		EC2	0.919	18.503	0.845		
		EC3	0.911	17.986	0.829		
		EC4	0.903	17.542	0.815		
		EC5	0.909	17.871	0.825		
Capability (Study 1)	0.947	CAP1	0.891	-	0.794	0.808	0.947
		CAP2	0.911	16.892	0.830		
		CAP3	0.903	16.513	0.815		
		CAP4	0.914	17.038	0.835		
Anthropomorphism (Study 1)	0.949	AN1	0.877	-	0.769	0.848	0.94
		AN2	0.913	16.342	0.834		
		AN3	0.889	15.368	0.790		
		AN4	0.892	15.491	0.796		
Positive Affects (Study 1)	0.947	AFF1	0.912	-	0.832	0.874	0.946
		AFF2	0.889	16.907	0.791		
		AFF3	0.913	18.21	0.834		
		AFF4	0.902	17.587	0.814		

Next, a second-order factor model was established where four constructs of FATSR-17 served as the latent first-order factors (ethical concern, capability, anthropomorphism, and positive affect) while the five items of trustworthiness served as the latent second-order factor. The results of the CFA analysis are shown in Figure 9.9 and Table 9.5.

The association between the theory and the empirical results was examined via the goodness-of-fit indicators. Following the previous literature (Bhattacherjee, 2002), chi-square (χ^2) and adjusted chi-square (χ^2/df) were 563.33 and 2.75, respectively, which were all within the threshold, suggesting adequate model fit. Additional fit indices included IFI (Incremental Fit Index), NNFI (Non-Normed Fit Index), and CFI (Comparative Fit Index) and their threshold should all exceed 0.9. Estimation of IFI, NNFI, CFI for the second-order model were 0.92, 0.91, and 0.92, respectively, suggesting the current model adequately fits the responses.

Further, path coefficients between trust and four dimensions and factor loading for each construct were examined. According to the previous literature that path coefficients should be significant ($p < 0.05$) and the loading factors should exceed 0.7, each construct is explained by its items rather than error accordingly (Bhattacherjee, 2002). As depicted in Figure 9.9 and Table 9.5, all the factor loadings and path coefficients were within the threshold, suggesting a high correlation between the two scales.

Thus, both scales enjoyed high validity and reliability to measure facial anthropomorphic trustworthiness. Considering the fine-grained nature of trustworthiness, it might be appropriate to evaluate facial anthropomorphic trustworthiness as a second-order construct since it could provide enriched information.

Table 9.5 The path coefficient for the second-order factor model

Path	Standardized coefficient	Standard error	C.R. (t-value)
Trust - > Ethics Concern	0.974***	0.059	18.445
Trust - > Capability	0.968***	0.064	17.076
Trust - > Anthropomorphism	0.979***	0.065	18.570
Trust - > Positive Affect	0.974***	0.063	16.781

Note: * p<0.1; ** p<0.05; *** p<0.01

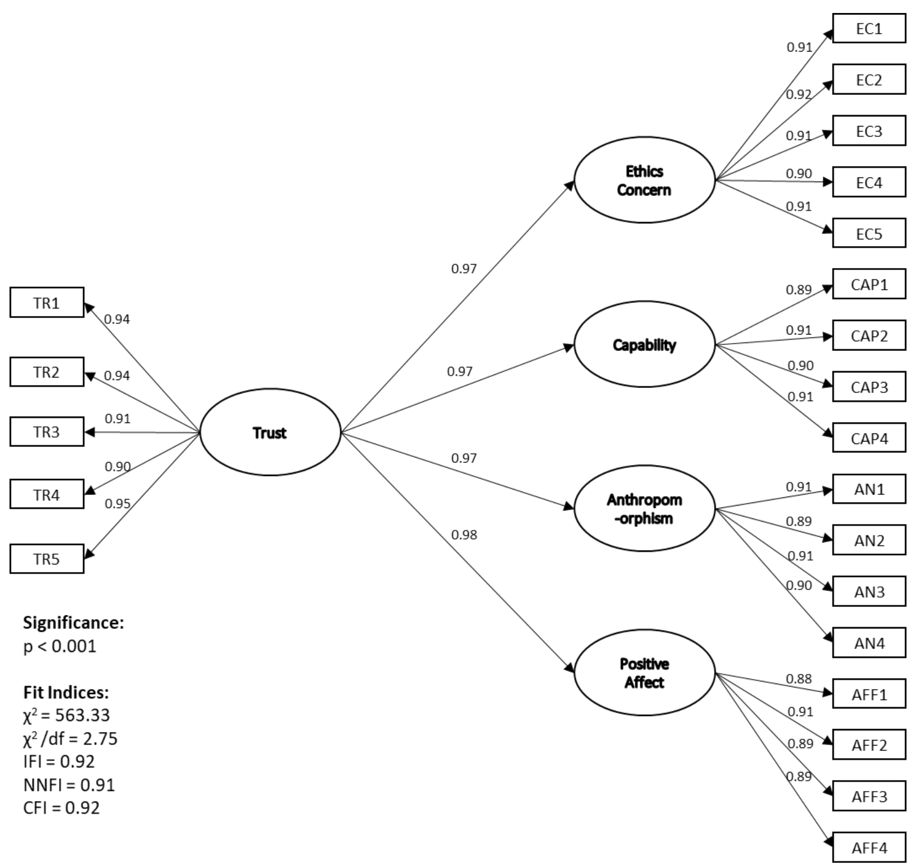


Figure 9.9 The model fit and path analysis for the second-order factor model

9.3.2 Study 6: Emotional Expressions in a Context-free scenario

With regard to subjective rating on facial anthropomorphic trustworthiness, the score was averaged of the ratings of 17 items in FATSR-17. A two-way ANOVA was performed, with valence (positive vs. negative) and arousal (active vs. inactive) as the independent variables and with FATSR-17 as the dependent variable. The results of the two-way ANOVA showed that the main effect of valence on trustworthiness evaluation was significant ($F(1, 34) = 33.64$, $p < 0.01$), while the effect of arousal ($F(1, 34) = 0.68$, $p = 0.42$) and the interaction effect were not significant ($F(1, 34) = 2.91$, $p = 0.10$). Specifically, post-hoc tests revealed that robots with positive expressions showed significantly higher trustworthiness perceptions than those with negative expressions. Thus, H 9.1 was supported while H 9.2 was not supported (see Tables 9.6).

Table 9.6 Descriptive statistics for trustworthiness, fixation duration, and fixation count for different facial valence and arousal in study 6

Valence	**Arousal**	**Mean**	**SD**
Facial Anthropomorphic Trustworthiness (Likert scale)			
Negative	Inactive	4.28	1.91

	Active	3.74	1.86
Positive	Inactive	6.36	1.46
	Active	6.55	1.48
Fixation Duration (Millisecond)			
Negative	Inactive	949.83	462.51
	Active	1090.63	577.55
Positive	Inactive	1939.46	546.38
	Active	1860.31	1075.95
Fixation Count (Frequency)			
Negative	Inactive	6.11	3.11
	Active	5.31	2.23
Positive	Inactive	9.83	2.94
	Active	8.86	2.38

As for eye-tracking measures on facial anthropomorphic trustworthiness, a similar two-way ANOVA was performed, with valence (positive vs. negative) and arousal (active vs. inactive) as the independent variables and with fixation duration and fixation count as the dependent variables. The results of the two-way ANOVA showed that the main effect of valence on fixation duration ($F(1, 34) = 38.67$, $p < 0.01$) and fixation count ($F(1, 34) = 44.68$, $p < 0.01$), while the effect of arousal on fixation duration ($F(1, 34) = 0.06$, $p = 0.80$) and fixation count ($F(1, 34) = 3.56$, $p = 0.06$) were not significant. Besides, the interaction effect of valence and arousal on fixation duration ($F(1, 34) = 0.79$, $p = 0.37$) and fixation count ($F(1, 34) = 0.03$, $p = 0.85$) were also non-significant. Thus, H 9.3 was supported while H 9.4 was not supported.

Last, additional investigation was performed to analyze the correlation relationship between demographics and facial anthropomorphic trustworthiness, results of the Pearson test showed no significant correlation between facial anthropomorphic trustworthiness and age ($p = 0.35$) or gender ($p = 0.23$).

9.3.3 Study 7: Dynamic Expressions in Contexts with Regulatory Focus

With regard to subjective rating on facial anthropomorphic trustworthiness (FATSR-17), a three-way ANOVA was performed, with valence (positive vs. negative), arousal (active vs. inactive), and focus (prevention-focused vs. promotion-focused) as the independent variables and with FATSR-17 as the dependent variable. Table 9.7 summarizes the descriptive statistics for trustworthiness and EDA for different facial valence, arousal, and focus in study 7. The results of the three-way ANOVA showed that the main effect of valence on trustworthiness evaluation was significant ($F(1, 34) = 6.55$, $p < 0.05$), while the main effect of arousal ($F(1, 34) = 2.69$, $p = 0.11$) and focus ($F(3, 102) = 0.53$, $p = 0.66$) were not significant. For interaction effect, only the interaction between valence and focus ($F(3, 102) = 61.88$, $p < 0.01$) while the other interactions were all non-significant (see Figure 9.10).

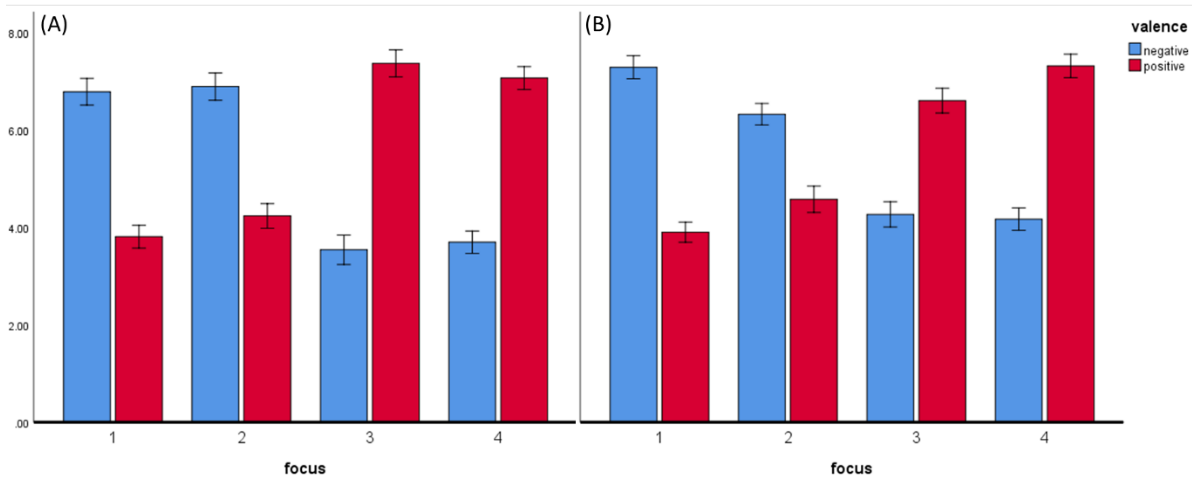


Figure 9.10 Bar chart representing the results of the 3-way interactions in facial trustworthiness ratings of study 6: (A) valence x focus at inactive; (B) valence x focus at active.

Note: focus 1 refers to the scenario, prevention-1; focus 2 refers to the scenario, prevention-2; focus 3 refers to the scenario, promotion-3; focus 4 refers to the scenario, promotion-4; Error bars represent ±1 SE. Unit: Likert-scale

Table 9.7 Descriptive statistics for trustworthiness and EDA for different facial valence, arousal, and focus in study 7

			FATSR-17		EDA (uS)	
Valence	**Arousal**	**Focus**	**Mean**	**SD**	**Mean**	**SD**
Negative	Inactive	Prevention-1	6.78	1.63	12.77	3.46
		Prevention-2	6.89	1.66	12.34	3.37
		Promotion-3	3.54	1.78	13.55	3.41
		Promotion-4	3.69	1.36	14.78	3.24

	Active	Prevention-1	7.28	1.40	12.89	3.36
		Prevention-2	6.32	1.30	12.48	3.46
		Promotion-3	4.26	1.53	13.44	3.39
		Promotion-4	4.17	1.36	14.97	3.27
Positive	Inactive	Prevention-1	3.81	1.39	14.69	3.42
		Prevention-2	4.23	1.50	12.84	3.31
		Promotion-3	7.36	1.64	11.80	3.33
		Promotion-4	7.06	1.40	12.77	3.37
	Active	Prevention-1	3.90	1.22	14.84	3.52
		Prevention-2	4.57	1.59	12.89	3.50
		Promotion-3	6.60	1.50	11.86	3.34
		Promotion-4	7.31	1.43	12.81	3.43

As for the physiological measure of EDA in different scenarios, a three-way ANOVA was performed, with valence (positive vs. negative), arousal (active vs. inactive), and focus (prevention-focused vs. promotion-focused) as the independent variables and with EDA as the dependent variable. The results of the three-way ANOVA showed that the main effect of valence was significant ($F(1, 34) = 57.34$, $p < 0.01$), while the main effect of arousal ($F(1, 34) = 2.89$, $p = 0.10$) and focus ($F(1, 34) = 0.24$, $p = 0.63$) were not significant. For interaction effects, only the interaction between valence and context was significant ($F(1, 34) = 1723.84$, $p < 0.01$). To specify, the regulatory fit scenarios that positive expressions in promotion focus and negative expressions in prevention focus tend to have a low EDA and a high level of trustworthiness than the regulatory unfit scenarios that positive expressions in prevention focus and negative expressions in promotion focus. The other interactions were all non-significant (see Table 9.7 and Figure 9.11). Thus, H 9.5 and H 9.6 were supported.

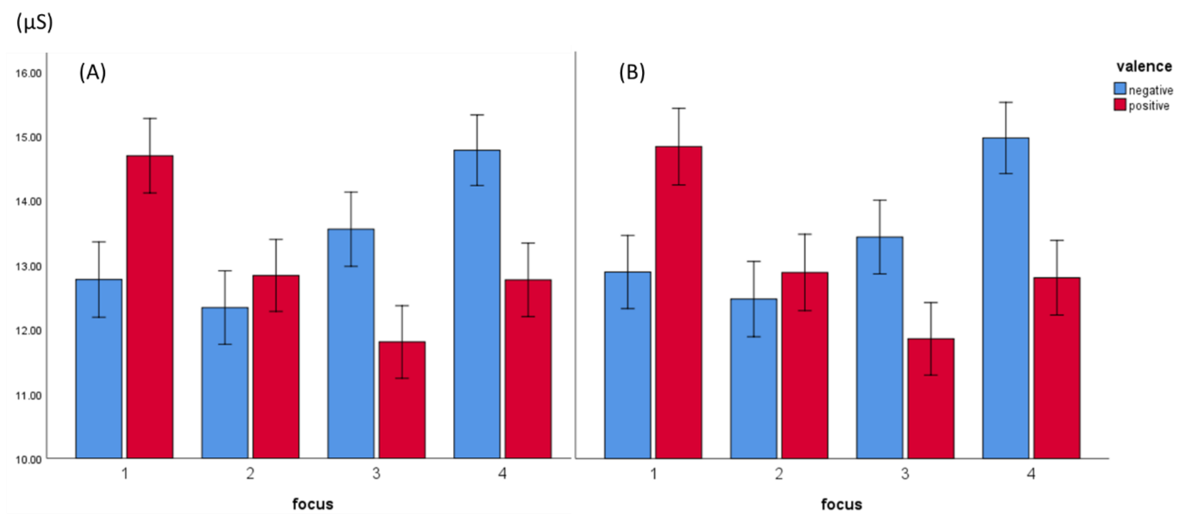


Figure 9.11 Bar chart representing the results of the 3-way interactions in EDA of study 6. (A) valence x focus at inactive; (B) valence x focus at active.

Note: Focus 1 refers to the scenario, prevention-1; focus 2 refers to the scenario, prevention-2; focus 3 refers to the scenario, promotion-3; focus 4 refers to the scenario, promotion-4; Error bars represent ±1 SE. Unit: uS

9.3.4 Modeling for Dynamic Features

In order to conduct modeling analysis on facial anthropomorphic trustworthiness from dynamic features, SPSS was utilized to perform the stepwise regression for this modeling analysis.

As for dynamic feature modeling, significant factors were revealed and added in the stepwise regression in a sequence. Factors included valence (coded one for negative expression while two for positive expression), arousal (coded one for inactive expression while two for active expression), regulatory-focus (the degree ranged one to four from prevention to promotion), and the interaction term of valence and focus. Table 9.8 shows the summarized results of the stepwise regression using standardized and unstandardized coefficients (β).

Table 9.8 The summarized standardized result of stepwise regression in study 7

Model	**1**	**2**	**3**	**4**
Valence	0.06*	0.06*	0.06*	-1.41***
Arousal		0.03	0.03	0.03
Focus			0.15	-1.95***
Valence x Focus				2.54***

Note: * denotes significant at 0.05; ** significant at 0.01; *** significant at 0.001

See equations for unstandardized models 1-4:

1) Facial Trustworthiness = 5.13 + 0.24 Valence

2) Facial Trustworthiness = 4.92 + 0.24 Valence + 0.13 Arousal

3) Facial Trustworthiness = 4.86 + 0.24 Valence + 0.13 Arousal + 0.03 Focus

4) Facial Trustworthiness = 14.07 - 5.91 Valence + 0.13 Arousal - 3.65 Focus + 2.46 Valence * Focus

Last, additional investigation was performed to analyze the correlation relationship between demographics and facial anthropomorphic trustworthiness, results of the Pearson test showed no significant correlation between facial anthropomorphic trustworthiness and age ($p = 0.40$) or gender ($p = 0.66$).

9.4 Summary and Discussion

Considering few studies have expanded and validated results of human facial expressions in the context of HRI, studies 6 and 7 conducted two experiments to explore the effect of dynamic expressions on facial anthropomorphic trustworthiness via a combination of subjective rating and physiological measures. To specify, study 6 examined the effect of dynamic expressions on facial anthropomorphic trustworthiness in a context-free scenario, attentional allocation for each expression and further validated two scales for facial anthropomorphic trustworthiness; study 7 then examined the regulatory fit theory in the context of social robot via analyzing the role of dynamic expressions under different regulatory-focused contexts (prevention-focused or promotion-focused) in communicating facial anthropomorphic trustworthiness. Further, To specify, before the main studies, the difference between FATSR-17 and the trustworthiness scale adopted in studies 2-5 were analyzed. According to the results of the second-order factor model, the evaluations of the trustworthiness scale by Gorn (2008) and FATSR-17 were correlated and both of them enjoyed a high validity and reliability. Considering the fine-grained nature of trustworthiness, FATSR-17 could provide enriched information which makes it more appropriate to determine the facial anthropomorphic trustworthiness evaluation for a social robot.

In study 6, we investigated the effect of four emotional expressions (two levels of valence by two levels of arousal) on facial anthropomorphic trustworthiness and physiological indicator (fixation duration/ counts). Consistent with the H 9.1, results showed that positive (vs. negative) emotional expressions enjoyed a higher level of facial anthropomorphic trustworthiness. Thus, H 9.1 was supported. However, arousal in emotional expressions and its interaction with valence did not have a significant influence on facial anthropomorphic trustworthiness. H 9.2 was not supported. In order to have a further examination of attentional allocation in processing faces, four facial expressions were identified and suggested different visual patterns when evaluating emotional expressions: faced with happy expressions, individuals automatically tend to have visual attention (fixation duration and counts), resulting in a higher level of trustworthiness evaluation, which is consistent with the literature regarding human facial recognition and evaluation (Beaudry et al., 2014; Bombari et al., 2013; Calvo and Nummenmaa, 2008). Accordingly, H 9.3 was supported and H 9.4 was not supported.

Based on the result from study 6, we further examined the effect of four dynamic expressions (two levels of valence by two levels of arousal) in contexts with different regulatory focus (prevention-focused or promotion-focused) via a lab experiment (study 7). Results showed that: 1) Similar to study 6, positive dynamic expressions have a high level of facial anthropomorphic trustworthiness and a low level of EDA; 2) facial arousal of social robot still did not have a significant impact on facial anthropomorphic trustworthiness and EDA; 3) inspection of the means of facial anthropomorphic trustworthiness and EDA demonstrated that participants had a higher level of perceived trustworthiness (a lower level of EDA) toward the social robot with positive expressions when they were presented with the message that was promotion-focused than when they were presented with the prevention-focused context. Conversely, participants had a higher level of perceived trustworthiness (a lower level of EDA) toward the social robots with negative expressions when they were confronted with the prevention-focused context than when they were confronted with the promotion-focused. These results confirm H 9.5 and H 9.6.

Studies 6 and 7 might have the following theoretical contributions. First of all, although plenty of prior research has discussed the relationship between emotional expressions and facial trustworthiness, they were all within the field of human faces.

Few studies have attempted to explore the effect of emotional expressions on facial trustworthiness in the context of anthropomorphism. In order to fill this research gap, study 6 and 7 conducted an experiment to expand and generalize the conclusions of human facial trustworthiness to facial anthropomorphic trustworthiness, where people tend to share a similar strategy to evaluate both human and robot face (Palinko et al., 2015). Although due to technical constrain or ethical considerations, anthropomorphic faces might be distinct from human faces, instinct might prompt humans to identify and respond to anthropomorphic faces, during which similar perceptions might be automatically evoked as interacting with another human being (Sproull et al., 1996). Accordingly, this study contributes to the understanding of anthropomorphism and anthropomorphic objects (Epley et al., 2007).

In addition, this study contributed to the research of anthropomorphic trustworthiness from both subjective rating and physiological measures. Although prior research on facial trustworthiness has adopted various methods, ranging from subjective rating to physiological measures, it mainly concentrates on the context of human facial evaluation (Dong et al., 2018; Todorov et al., 2008b; van 't Wout and Sanfey, 2008). Few studies have tried to adopt both subjective and physiological approaches in analyzing facial anthropomorphic trustworthiness for a social robot. Through combining subjective evaluation and physiological measures, this study provides a relatively reliable insight into the relationship between dynamic features, regulatory focus, and facial anthropomorphic trustworthiness.

Still, a few limitations are worth noting. The participants in this study were college students. Although the previous study did not explicitly suggest age might work as an efficient factor influencing people's attitude and behavioral intentions toward the social robot, they indeed had more opportunity to be exposed to advanced technology and show a more favorable attitude toward emerging creatures (Turner, 2015). A future study could recruit different groups to validate the current findings. Furthermore, even within positive expressions, such as smile, prior works have suggested enjoyment smiles and non-enjoyment smiles varied in their degree of "genuineness" or "convincingness" (Johnston et al., 2010; Slessor et al., 2010). For example, Centorrino et al. (2015) indicated that smiles rated as more genuine strongly predict judgments about the trustworthiness of trustees, and willingness to cooperate. Regarding social robot could also have non-enjoyment smiles, it would be theoretically interesting to explore whether a human could recognize the fake "smile" on an anthropomorphic face and respond in a similar manner as it did in interpersonal relationship. Last, though the current sample size (N = 35) is adequate for a behavioral experiment, it might be relatively small to build a precise model for dynamic features and examine the difference between the two scales. A future study could try to draw a large sample to validate the current model and further discuss the difference between the two scales.

CHAPTER TEN. DISCUSSION AND FINALIZATION OF FACIAL ANTHROPOMORPHIC TRUSTOWRHTINESS MODEL

This chapter discusses the results of each study to answer the research questions proposed in chapter one. By integrating the findings from various methods, the research proposes models of facial anthropomorphic trustworthiness for static and dynamic features. Moreover, a list of design guidelines for trustworthy social robot design and managerial implications were also discussed in this chapter. Last, It discusses the limitations of the research and proposes possible areas for further study.

10.1 Discussion of Research Questions

As the human mindset and its adaptive behavioral patterns are increasingly affected by their surrounding eco-system, people are still easily influenced by the appearance of an entity. Acting as various social roles in our daily lives, the social robot's physical characteristics, such as facial features, could still shape its personality attribution, such as trustworthiness (Fortunati, 2015). As facial anthropomorphic trustworthiness plays an important role in HRI, it could prompt communication and accelerate further interaction, even at first sight (Luo et al., 2006). To address the RQs 1-3, this study focused on the effects of two groups of facial features, static and dynamic features, on facial trustworthiness in the first impression.

10.1.1 Understanding of Facial Anthropomorphic Trustworthiness (RQ 1)

A social robot is an artificial intelligence autonomous system, which is specially designed to communicate and interact with humans or other intelligent agents by following social behaviors or acting as one of the social roles. With increased technology and equipment applied in social robots, it is becoming a medium or a communication partner between human and digital data in our daily lives, supporting us in physical and emotional ways (Hoorn, 2015). Considering the significant role of ethics evaluation at the initial step of HRI, facial anthropomorphic trustworthiness indeed plays a crucial role in building the initial credibility and the approaching intention at a later stage (Hoorn, 2015). Within the domain of social robots, however, it still lacks a better understanding and a valid scale to measure the constructs of interest: facial anthropomorphic trustworthiness. Accordingly, study 1 (phase 1) fulfilled the RQ 1 by conducting a hybrid deep learning approach to explore the meaning and develop a scale to measure facial anthropomorphic trustworthiness toward social robots (FATSR-17).

The current result is partially consistent with the concept of human trustworthiness where it has three dimensions: ability, benevolence, and integrity. Particularly, ability in interpersonal trustworthiness, which refers to the individual's evaluation whether others' related capability and knowledge in the given task (Mayer et al., 1995), is consistent with the capability dimension of a social robot, that concerns the evaluation of the power or ability of a social robot to get tasks done. Benevolence in human trustworthiness, which refers to the degree that people tend to do well to themselves, even beyond their profit motivation (Mayer et al., 1995), might be related to the positive affect dimension of a social robot, where an individual subjectively experiences positive moods when interacting with a social robot, such as happiness and interest. Integrity, which refers to the individual's evaluation of whether others would obey a set of social rules in the interpersonal interaction (Mayer et al., 1995), might be related to the ethics concern dimension for a social robot, which mainly concerns the ethical consideration toward the social robot, such as the anxiety toward the integrity of the programmer. However, trustworthiness for social robots indeed has its own distinction from human trustworthiness: its anthropomorphic nature (Rosenthal-von der Pütten et al., 2019; Y. Zhang et al., 2019). While individuals might expect it to behave like a human and obey specific social norms, which makes them as friends or companions for humans (Hoorn, 2018), others might consider an as a non-human object or a product without extensive emotional bonding (Ames et al., 2010; Bart et al., 2005; Kocsor and Bereczkei, 2017; Song and Luximon, 2020a; van 't Wout and Sanfey, 2008). Therefore, robot trustworthiness stands in the middle between human trustworthiness and product trustworthiness, which would be influenced by not only the affective reactions, i.e. human trustworthiness, but also the logical reasoning, i.e. product trustworthiness. Accordingly, anthropomorphism, which refers to the degree that attributing human characteristics, emotions, or intentions to non-human entities, would work as a distinct dimension from previous research (Al-Qaderi et al., 2018; Förster et al., 2019) and facial anthropomorphic trustworthiness for social robots could be defined as "an impression-based trustworthiness where people relied on facial cues of social robots to evaluate their capability, ethics, anthropomorphism, and positive affect" (see Figure 10.1).

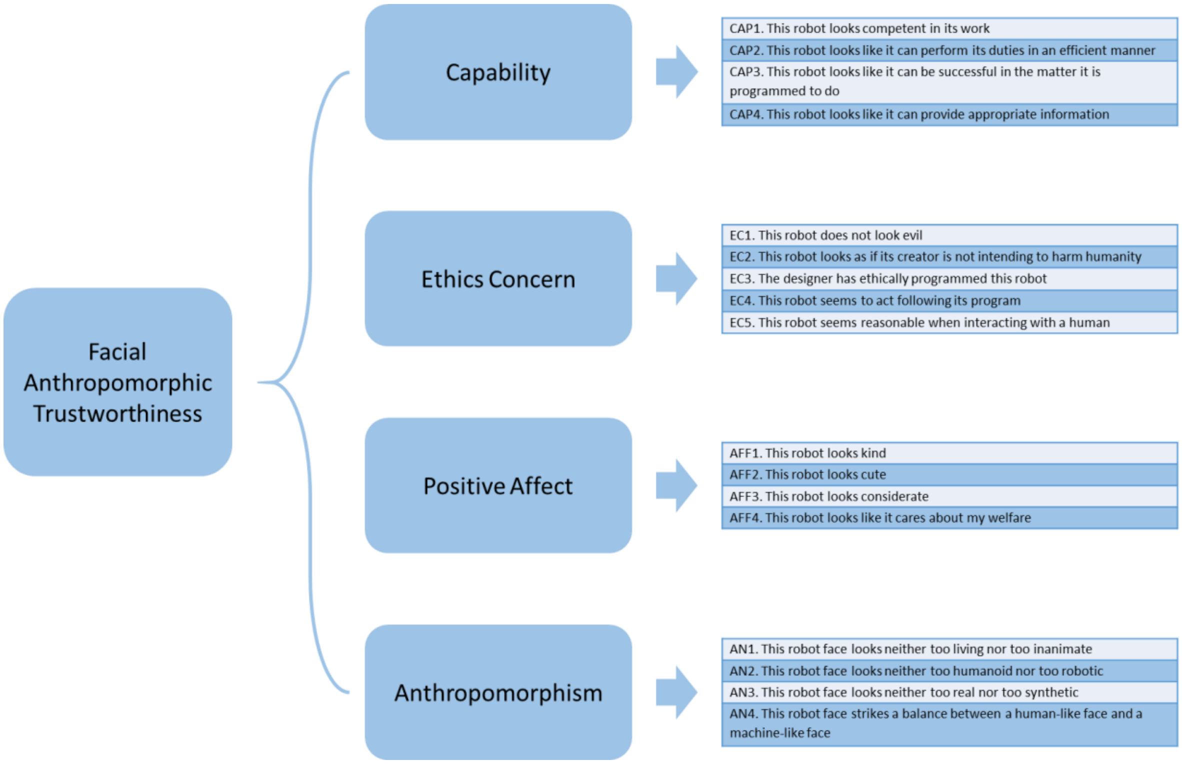


Figure 10.1 The meaning and constructs of facial anthropomorphic trustworthiness.

Study 1 aims to contribute to the literature on HRI and scale development from three perspectives. First of all, this scale, for the first time, set out to develop a facial anthropomorphic trustworthiness scale, which tried to merge the theory among interpersonal trustworthiness, the uncanny valley, and general robot trustworthiness. To some extent, it could provide a relatively holistic picture when we evaluated the trustworthiness of a social robot in the first impression. In addition, this study used a relatively large sample of qualitative data as the input for the item generation process. Considering the relatively high cost (time and money) of the traditional qualitative method (i.e. interview and focus group), the current study recruited a relatively larger sample of qualitative data via a crowdsourcing platform. Lastly, to avoid the unconscious bias in interpreting corpus of a qualitative method (Frey and Fontana, 1991), the current study used a SOTA natural language processing technique to cluster the corpus in a more objective way (Timoshenko and Hauser, 2019). With the rapid development in the field of deep learning, quantifying the qualitative data might be an emerging field that helps design research to examine the collected data from various perspectives, such as sentiment analysis, autonomous translation, etc.

10.1.2 Static Features of Facial Anthropomorphic Trustworthiness (RQ 2)

Based on the result of the literature review, the facial trustworthiness for social robots at first sight mainly depended on static features and dynamic features. Within the static features, studies 2-5 (phase 2) addressed RQ 2, which deliberately examined the separate features (eye shape, mouth shape, face shape, and fWHR) and their combinations (interactions effect of size and position of eye and mouth) in communicating facial anthropomorphic trustworthiness for a social robot.

First of all, eye shape had a significantly positive impact on facial anthropomorphic trustworthiness where round eyes (vs. narrow eyes) were considered as more trustworthy; mouth shape had a significantly positive impact on facial anthropomorphic trustworthiness where upturned and neutral shaped mouth (vs. downturned shaped mouth) were considered as more trustworthy.

This finding on internal features was consistent with the prior research on facial trustworthiness where individuals might believe round eyes (vs. narrow eyes) as significant indicators for facial babyishness (Haselhuhn et al., 2013; Maoz, 2012), thus improving trustworthiness (Ferstl et al., 2017; Masip et al., 2004). Similarly, compared with a downturned mouth (sad mouth) (Landwehr et al., 2011), human faces with an upturned mouth (smiling mouth) or a neutral-shaped mouth might enjoy a high level of trustworthiness and friendliness (Arminjon et al., 2015; Kleisner et al., 2013; Landwehr et al., 2011; Maeng and Aggarwal, 2018). Accordingly, the current study, for the first time, tried to give preliminary evidence to suggest that the static facial features could also work for social robots: round eyes and an upturned mouth (or neutral mouth) could improve people's trustworthiness towards the robot.

In addition, fWHR had a significantly positive impact on facial anthropomorphic trustworthiness where high fWHR (vs. low fWHR) was considered as more trustworthy; face shape had a nonsignificant impact on facial anthropomorphic trustworthiness where rounded shape and rectangular shape shared similar perceived trustworthiness.

This finding on external features was, counter-intuitively, against the previous literature on fWHR. While people with high fWHR might be less evaluated (Haselhuhn et al., 2013; Kramer, 2015; Welker et al., 2016), a robot with high fWHR might be considered as more trustworthy. The reason might lie in the theory of extended self (Ladik et al., 2015) as individuals might consider the social robot as an extension of bodies. Since the face of the robot is processed in a similar manner as in processing human face, it is reasonable to predict that a robot with more capable and competitive looking could, instead, enjoy a high level of trustworthiness evaluation (Song and Luximon, 2021). Moreover, results on face shape were consistent with the previous finding that individuals' preference for a specific shape depends on the specific situation. On the one hand, individuals have generally shown a preference for a rounded shape (Westerman et al., 2012) since it might be considered safe, natural, and approachable (Sevilla and Kahn, 2014). On the other hand, rectangular shape, as a typical component in the history of the robot (Hwang et al., 2013), might also be valued since it could help us easily categorizing a certain object (Meeden and Blank, 2006). Thus, the desire for a rounded shape might be, in turn, counteracted or neutralized by people's typicality preference, resulting in the insignificant effect of face shape on trustworthiness evaluation.

For feature combinations, eye size had a significantly positive impact on facial anthropomorphic trustworthiness where big eyes (vs. small eyes) were considered as more trustworthy; mouth size had a nonsignificant impact on facial anthropomorphic trustworthiness where small mouth or big mouth enjoyed similar perceived trustworthiness; inward positioning had a significantly positive impact on facial anthropomorphic trustworthiness, however, extremely centralized feature positioning would dampen facial anthropomorphic trustworthiness.

With regard to feature size, on the one hand, this finding is consistent with prior research on trustworthiness in humans and HRI: both people and robots with big eyes were considered to be more honest and innocent (Chen et al., 2010; Ferstl et al., 2016; Kalegina et al., 2018). Even, people wearing glasses could be perceived as having bigger eyes and more trustworthy and intelligent (Leder et al., 2011). On the other hand, the nuanced relationship between mouth size and perceived trustworthiness might account for this nonsignificant effect. While a small mouth is often associated with an infant face (Glocker et al., 2009b), people have shown their preference for a face with a big mouth since it is a signal of confidence, capability, and trustworthiness (Re and Rule, 2016). According to the theory of robot communication (TORC), we sometimes treated social robots as "hyper-persons" whose negative cues are filtered out while positive features are filtered in (Konijn and Hoorn, 2017).

As for feature positioning, these results were consistent with Berry and McArthur's work (1985) which discussed the adaptive covariations between baby-faced cues and appearance-based stereotypes. Although the centralized tendency of eye and mouth positioning might enjoy a higher level of trustworthiness, unnatural displacement, such as extremely concentrated positioning, might lead to undesirable reactions or responses (Chen et al., 2010; Miesler et al., 2011) and dampen facial anthropomorphic trustworthiness: compared with the most centralized position of eye and mouth, medium level of centralized (medium vertical and horizontal position of eyes, and medium/high vertical position of the mouth) was deemed highest level of trustworthiness. In other words, it is suggested that striking a balance between ordinal positions and extreme centralized positions in which the face of a social robot could retain the evolutionary benefits of the baby schema. This finding is also consistent with previous work that extremely close positioning of eyes would reduce the perceived trustworthiness (Kalegina et al., 2018).

Based on the observations from studies 2-4, study 5 modeled the static facial features to predict anthropomorphic trustworthiness in a holistic way. To be more specific, 911 observations were merged and examined via stepwise regression. Overall, the finding was consistent with the conclusions in studies 2-5 and configurated the different features in determining the subjective rating of facial anthropomorphic trustworthiness (see Figure 10.2).

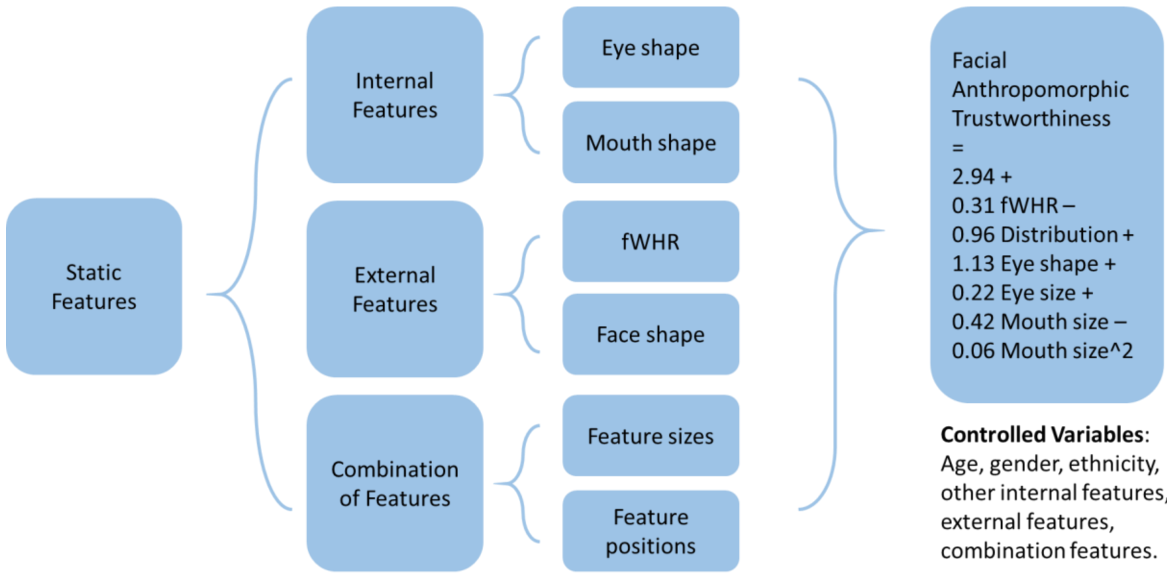


Figure 10.2 Static features modeling for anthropomorphic trustworthiness

In order to validate the trustworthiness model, a series of experiments, which contained planned design and random design for the social robot, was conducted. As for the planned design, additionally tested stimuli and untested stimuli were introduced and rated to be compared with the 95% predicted interval estimated by the model. The finding showed it reached a 95.6% accuracy rate (22/23*100%). Thus, it showed that the current model achieved relatively high reliability for predicting facial anthropomorphic trustworthiness with different configurations.

Also, some variables are controlled in the current model. For example, study 5 examined the difference between ethnic groups (Chinese and American) in perceiving facial anthropomorphic trustworthiness. Results indicated that there seems no significant difference for different ethnic groups in evaluating facial anthropomorphic trustworthiness for a social robot, which is consistent with the previous literature (Birkás et al., 2014), suggesting different ethnical groups indeed developed similar evaluation strategies to evaluate facial trustworthiness with no significant difference (Etcoff et al., 2011). Moreover, age and gender were also examined from study 2-7. Results indicated that they have no significant influence in perceiving facial anthropomorphic trustworthiness, which was consistent with the finding in human facial evaluation (Ma et al., 2015). Other facial features, such as facial color and eye color, all severed as control variables, which were not included in the model due to ethical issues guidelines and universal design of the robot and AI (Torresen, 2018; Winfield, 2019).

10.1.3 Dynamic Features of Facial Anthropomorphic Trustworthiness (RQ 3)

Followed the results from static features, studies 6 and 7 (phase 3) addressed RQ 3, which investigated the effect of dynamic features on facial anthropomorphic trustworthiness and further examined their effect under different regulatory-focused contexts on facial anthropomorphic trustworthiness via a lab experiment.

By analyzing people's subjective and physiological responses, studies 6 and 7 provided sights into how dynamic expressions within different regulatory contexts affect individuals' trustworthiness evaluations. The findings of study 6 confirmed that facial valence worked as a significant factor in influencing facial anthropomorphic trustworthiness where positive expressions (vs. negative expressions) enjoyed a higher level of trustworthiness. However, facial arousal did not have a significant effect on facial anthropomorphic trustworthiness where active or inactive emotions tended to have similar trustworthy evaluations and visual attention.

This result of facial anthropomorphic trustworthiness and fixations were consistent with the previous study that positive expressions (e.g. smile or happy face) are positively associated with facial trustworthiness while negative expressions (e.g. sad or unhappy face) are negatively associated with facial trustworthiness (Engell et al., 2010; Gutiérrez-García and Calvo, 2016b). Since positive expressions, such as a smile, are universally considered as indications of positive experience (Calvo et al., 2017; Elfenbein and Ambady, 2002), it usually could work as supplementary nonverbal signals for social judgments. Similar to human emotional expressions, avatars, or social robots with human facial features could also have related emotional expressions (Ku et al., 2005). Indeed, this study suggested that this phenomenon was not exclusive to the human facial recognitions only, instead, it could potentially be generalized to the field of a social robot: Robots with positive expressions could similarly enjoy a higher level of anthropomorphic trustworthiness. Considering the strong association between emotional expressions and facial trustworthiness (i.e. happy faces are generally perceived to be more trustworthy), this finding tries to expand the visual processing pattern from interpersonal interaction to human-robot interaction.

Regarding facial arousal, its effect on facial anthropomorphic trustworthiness, visual attention, and, EDA were not significantly. There are several explanations that might account for this phenomenon. To begin with, human-robot interaction could be a long duration where individuals and robots communicate, connect, and bond with each other (Sandry, 2015). However, in the experiment setting, it could be relatively hard to elicit a deeper bonding experience for the participants, thus less involvement in the interaction with robots. Even within the human context, there seems no significant difference between two comparable negative facial expressions, such as anger and sadness, in communicating trustworthiness attracting visual attention (Okubo et al., 2018; Sanchez and Vazquez, 2014). Compared with facial valence, facial arousal might play a minor role in attracting visual attention and eliciting facial anthropomorphic trustworthiness, especially at the first sight of a social robot.

Moreover, the study analyzed the role of contexts with different regulatory focus interacted with different dynamic expressions. Regarding daily contexts could have different regulatory focus: promotion-focused events and prevention-focused events. The associations between different expressions and contexts are necessary to be explored to infer the appropriate combination. Overall, as predicted by regulatory fit theory, the result of scale rating and EDA found that promotion-focused context would be compatible with positive expressions in eliciting a higher level of facial anthropomorphic trustworthiness and a lower level of EDA; prevention-focused context would be compatible with positive expressions in eliciting a higher level of facial anthropomorphic trustworthiness and a higher level of EDA. However, when negative expressions are met with the promotion-focused context or positive expressions met with prevention-focused context, individuals would experience great difficulty in trusting those social agents. Besides, arousal and its interaction with valence seem to have a non-significant influence in influencing facial anthropomorphic trustworthiness.

This result was consistent with previous literature and tried to contribute to regulatory focus theory in terms of the situational activation of an individual's self-regulatory system via a nonhuman interaction that influences trustworthiness evaluation. As the extant literature provides evidence on the influence of message framing associated with a regulatory focus in the interpersonal context (Cesario et al., 2013; Ewe et al., 2018). The present study contributes to the existing knowledge by providing empirical evidence that a combination of anthropomorphic dynamic expressions and context associated with the regulatory focus can act as a prime to activate facial anthropomorphic trustworthiness.

Further, based on the observations from study 7, it modeled the dynamic facial features to predict anthropomorphic trustworthiness. Overall, the finding configurated the different dynamic features in determining the subjective rating of facial anthropomorphic trustworthiness (see Figure 10.3) while similar variables in static feature modeling, such as age and gender, were also controlled in this model.

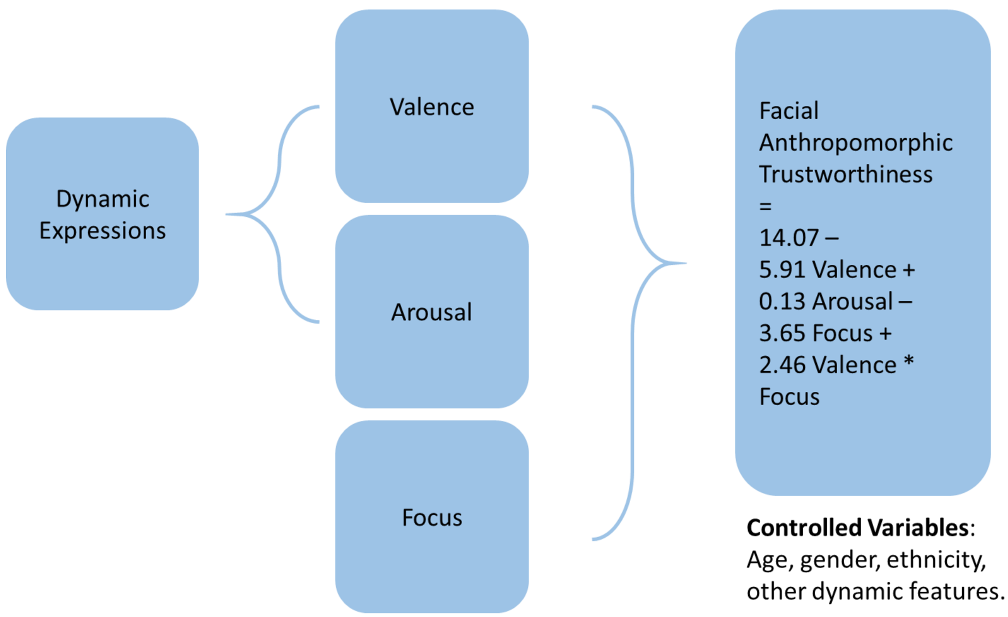


Figure 10.3 Dynamic features modeling for anthropomorphic trustworthiness

10.1.4 Design Guidelines and Managerial Implications

Based on the results from studies 1-7, this research explored how to configure facial features of the social robot to elicit people's anthropomorphic trustworthiness at first sight. Through a novel approach of processing qualitative data, this study first examined the specific meaning of facial anthropomorphic trustworthiness for social robots and its relevant sub-constructs to have an in-depth understanding of such emerging mediums. Then, the researchers conducted a series of experiments to investigate the effect of separate and combinations of static features and further model and validate the features for facial anthropomorphic trustworthiness. Last, this research extended subjective rating and physiological reactions to analyze the role of dynamic expressions and regulatory contexts in communicating facial anthropomorphic trustworthiness.

By examining people's trustworthiness reactions toward different facial configurations for the social robot, this study provides insights into how design patterns of social robots promote or dampen an individual's facial anthropomorphic trustworthiness and further introduces a set of appropriate design considerations. A list of design implications and guidelines is developed with the objective to promote the robot designer's understanding of anthropomorphic trustworthiness by facilitating more intuitive social robot design implications and guidelines, shown in Table 10.1. The design implications and guidelines contained five sections: the first section contained the general rules for designing a trustworthy robot, which works as the basis for comprehension of robot anthropomorphic trustworthiness; the other sections summarized the practical implications from the results of previous experiments.

One thing that worth noticing is that, in the practical application of the guidelines for designing a social robot, robot designers might face vague design situations where creators should appropriately arrange various design elements. For example, if a robot company is famous for its sharp-styled design, it might be inappropriate to adopt guideline No.6 to introduce a round eye for the robot. Indeed, robot design, or product design in general, should follow design rules which emphasize balance and harmony (Song and Luximon, 2020a). Accordingly, the facial design implications for a social robot in Table 10.1 could serve as general directions for making a trustworthy-looking robot. Additional design elements, such as brand style and local specific preference, should also be considered during the design process.

Table 10.1 Facial design implications and guidelines for a social robot

Principles	No.	Design Implications and Guidelines
General rules	1	To elicit affective reactions from a user, such as kindness, cuteness, and carefulness.
	2	To evoke confidence in the helpfulness of the social robot, such as the perceived capability.
	3	To strike a balance between a human-like appearance or a machine-like appearance.
	4	To have an integrity and reasonability image for the robot and its programmer.
Internal features	5	To design a relatively large eye size for a social robot, i.e. the diameter of the eye could be 1/3 of facial vertical height.
	6	To design a round eye shape for a social robot.
	7	To design an upturned or neutral mouth shape for a social robot.

	8	Either big or small mouth size would be fine, i.e. mouth size could be either 1/12 or 1/2 of facial horizontal width. It should be depended on the overall facial style.
External features	9	To design a high fWHR face for a social robot, i.e. the width height ratio could be around 1/2 to 2/3.
	10	Either rectangular or rounded face shape would be fine. It should depend on the overall style.
Feature combinations	11	To have an inward positioning of facial features where features are centralized positioned while avoiding extremely close positioning, i.e. eye displacement could be positioned at around 1/3 of facial vertical and horizontal position while mouth could be placed at around 1/4 of facial vertical position.
Dynamic expressions and contexts	12	To have a positive expression for the social robot in a context-free scenario.
	13	Either active or inactive expressions would be fine. It should depend on the usage and overall facial style.
	14	Positive expressions are compatible with promotion-focused contexts to increase perceived trustworthiness.
	15	Negative expressions are compatible with prevention-focused context to increase perceived trustworthiness.

As for managerial implications, the design guidelines for making a trustworthy-looking appearance for a social robot not only provide intuitions for robot designers but also benefits the related business and market. For example, Song and Luximon (2021) adopted a social robot as the research context to show the significant role of facial anthropomorphic trustworthiness in influencing purchase intentions. To be more specific, individuals might have a more willingness to buy a social robot with a high level of facial anthropomorphic trustworthiness while having a less willingness to buy a social robot with a low level of facial anthropomorphic trustworthiness. During this process, the effect of specific facial features on purchase intentions was mediated by facial anthropomorphic trustworthiness.

Accordingly, the current design guidelines could, to some extent, help relevant business sectors to increase the related-use and business sales, which could, in turn, promote the acceptance of such emerging AI agents and benefit the ecosystem as a whole.

10.2 Limitations and Future Work

The current research explored the effect of different facial features of the social robot on facial anthropomorphic trustworthiness via a structurally mixed method. However, there are still some limitations in this study due to the social issues, limited time, and research scope.

First, the current study mainly explored the different facial features in communicating trustworthiness for social robots via the current experiment settings. Despite a relatively limited exploration of facial features (such as size, position, and shape of eye and mouth) in this study, it could provide some preliminary evidence to illustrate the influence of facial features on the perceived trustworthiness. The current exploration of social robot either considers it in a context-free scenario or set it as a reminder. However, specific affiliations or occupations might also influence people's initial trust. For instance, a salesman was strongly correlated with untrustworthiness and unintelligence while a highly educated job, such as a professor, was usually believed to be trustworthy, intelligent, and helpful (Bonnefon et al., 2013; Hellström and Tekle, 1994). Considering different social robots could take various social roles, such as a financial consultant or an educational tutor, it is both theoretically and practically interesting to examine the appropriate association between jobs and their typical looking.

Second, the current study only focuses on facial anthropomorphic trustworthiness when interacting with social robots. Particularly, this research explored the specific meaning of facial anthropormophic trustworthiness via a structural approach of scale development. Although the process follows a rigous and systematic procedure of scale development, it might tend to provide a more descriptive conclusion. Futher study might use grouded theory to constructe hypotheses and theories to revalidate the current finding in this thesis. Besides, there are various kinds of potential perceptions or feelings that could be evoked during human-robot interaction. For example, when interacting with social robots, people would also have negative reactions, such as embarrassment (Bartneck et al., 2010), psychological distance (Kim et al., 2013), or even horribleness (Gray and Wegner, 2012). How to dampen or ease up the undesirable consequences when interacting with such an emerging medium is also related to facial anthropomorphic design since previous research on human facial attractiveness has also long enjoyed academic attention (Kim et al., 2013). The appropriate facial configuration could not only increase anthropomorphic trustworthiness but also lessen the negative reactions (Brenton et al., 2005). Reluctance to accept such an emerging medium is common in human-robot interaction, especially at first sight. Accordingly, future research could try to configurate facial features to avoid unfavorable perceptions or feelings aroused by social robot.

Third, when conducting this Ph.D. study, I started to learn machine learning and deep learning knowledge from the very beginning since I do think it could help me to improve the quality of this research. Though I have a solid mathematical and statistical background in business research, the study of machine learning and deep learning still takes me a lot of time since 2018. That is the main reason why studies 2-6 were conducted in parallel with study 1. Since study 1 is the study that I try to implement the NLP technique to explore the meaning of facial anthropomorphic trustworthiness and develop a scale. As discussed in study 6, although studies 2-5 have used the human trustworthiness scale as a protocol (Gorn et al., 2008) and both of them seems to have adequate reliability and validity to measure facial anthropomorphic trustworthiness, it is still necessary to validate the result of studies 2-5 via FATSR-17 and further examined the difference between these two scales through a large sample.

Fourth, although this Ph.D. project tries to control all other factors, such as gender and age, there are still some factors, such as prior experience interacting with the robot, which could potentially influence facial anthropomorphic trustworthiness. Indeed, a social robot was an emerging creature in our daily lives that most people might not be familiar with it (De Rie, 2016). Accordingly, a relatively large sample size was needed since individuals might rely on their previous experience to help them make a decision (Kardes et al., 2004). However, this factor was only explored in study 4 since study 4 recruited the largest sample (N = 270) among all the studies, which could help to have a more precise estimation of its effect. A future study might recruit another large sample to validate the effect of robot experience on facial anthropomorphic trustworthiness evaluation. Moreover, the current study also makes efforts to address the potentially biased sampling issue in this study: this research combines both online experiments and lab experiments to explore the effect of anthropomorphic features on trustworthiness evaluation for social robot. To be more specific, online experiments could reach more diverse samples, recruit large subject pools, and could be conducted in a more efficient way, which could enjoy a greater external validity and generalizability (Salganik et al., 2006) while lab experiments could have strong experimental controls, which could confirm the reliability and internal validity (Bond et al., 2012). Considering the feasibility of experiment implementation, this study recruited an American online sample in studies 2-4 and a Chinese lab sample in studies 6-7 to ensure the validity and reliability of this research. Under this experiment sampling setting, the role of culture in signaling human facial trustworthiness for different ethnic groups could also be further discussed. Since people could use facial cues as an intuitive tool for social judgments, especially meeting at the first time when extensive personality information is missing (Xu et al., 2012), little is known whether the conclusion from interpersonal cross-cultural trustworthy evaluation could be directly applied into the context of human-robot interaction. In accordance with the previous research that different ethnic groups have similar trustworthiness perception toward the same stimuli (Birkás et al., 2014; Xu et al., 2012), the current study provides preliminary evidence that different ethnical groups might share similar strategies or patterns to evaluate the trustworthiness for a social robot. Future studies would try to recruit an American sample for the lab experiments and a Chinese sample for the online experiments to reconfirm the current finding in this research.

Last, although the current research has tried to conducting modeling studies to analyze the effect of different features on facial anthropomorphic trustworthiness, they are static features modeling and dynamic feature modeling separately. Considering merging two models as a whole, it might be difficult to integrate the dynamic features and static features since dynamic features might not be coded or double coded in different features for stepwise regression modeling. For example, happy and active expressions first started with a neutral face then turned into a happy face with an increased size of mouth and eyes. Thus, it would violate the assumptions of stepwise regression, thus be inappropriate to use this modeling technique (Thompson, 1995). Although other statistical or machine learning techniques are needed to introduce and address this problem, they might need a relatively large amount of emotional expressions and human reaction responses as training sets and testing sets in order to get a reliable prediction rate (Ngiam et al., 2011). Accordingly, a future study could try to employ the machine learning technique to generate a more comprehensive model, which tries to examine both static and dynamic features in a holistic way, to give an in-depth understanding of facial anthropomorphic trustworthiness.

CHAPTER ELEVEN. CONTRIBUTION AND CONCLUSION

This chapter summarizes the major findings of the research and outlines its contributions and implications.

11.1 Major Findings

The research provides several important findings:

- Based on the human features of facial trustworthiness, anthropomorphic features of robot facial trustworthiness in human-robot interaction was investigated. Overall, human features of facial trustworthiness could be adapted to and shed light for features of facial anthropomorphic trustworthiness for the social robot, promoting perceived trustworthiness toward the social robot.
- Initial trustworthiness toward social robots could be summarized into four facial categories: internal features, external features, combinations of features, and emotional expressions. Essential features arising from the literature were outlined for further analysis: eye shape and mouth shape for internal features;

fWHR and face shape for external features; feature size and position for combinations of features; facial valence, facial arousal, and regulatory contexts for emotional expressions.

- Different from the constructs from human trustworthiness, facial anthropomorphic trustworthiness could have four distinct dimensions: capability, anthropomorphism, positive affect, and ethics concern. Accordingly, facial anthropomorphic trustworthiness could be initially summarized as impression-based trustworthiness where people relied on facial cues of social robots to evaluate their capability, ethics concern, anthropomorphism, and positive affect. Furthermore, a scale of facial anthropomorphic trustworthiness (FATSR-17) was developed and validated, which could work as the foundation for relevant research in the future.
- Internal features, eye shape, and mouth shape were analyzed. In general, eye shape and mouth shape have a significant impact on facial anthropomorphic trustworthiness where a robotic face with round eyes (vs. narrow eyes) and upturned or neutral mouth (vs. downturned mouth) enjoyed a higher level of facial anthropomorphic trustworthiness.
- External features, fWHR, and face shape were analyzed. In general, fWHR has a significant impact on facial anthropomorphic trustworthiness where a robotic face with high fWHR (vs. low fWHR) enjoyed a higher level of facial anthropomorphic trustworthiness. However, face shape (rounded or

rectangular shape) did not have a significant effect on facial anthropomorphic trustworthiness.

- A combination of features, feature size, and position, were analyzed. In general, eye size and positions of eyes and mouth have a significant impact on facial anthropomorphic trustworthiness for social robot where a robotic face with large eyes (vs. small eyes), medium horizontal and vertical position of eye and mouth (vs. high or low horizontal and vertical positions) enjoyed a higher level of facial anthropomorphic trustworthiness. However, mouth size (large or small) did not have a significant effect on facial anthropomorphic trustworthiness.
- A model for facial anthropomorphic trustworthiness in static features was developed. In this model, facial anthropomorphic trustworthiness was significantly predicted by fWHR, facial distribution, eye shape, eye size, mouth size, and quadratic term of mouth size. A validation study via variations in planned and random stimuli was conducted and confirmed the reliability of the current model.
- Dynamic features, facial valence and arousal, and their interaction with regulatory contexts were analyzed. In general, facial valence and its interaction with regulatory contexts have a significant impact on facial anthropomorphic trustworthiness where a robotic face with positive (vs. negative expressions) enjoyed a higher level of facial anthropomorphic trustworthiness while positive expressions were compatible with the

promotion-focused context and negative expressions were compatible with prevention-focused context, in signaling facial anthropomorphic trustworthiness. However, facial arousal (active or inactive) did not have a significant effect on facial anthropomorphic trustworthiness. A dynamic features model was also developed in this study.

11.2 Contributions

rThe findings presented in this thesis have important implications for theoretical exploration and design practice. As for theoretical contributions, this research advanced the knowledge of human-robot interaction:

- This study, for the first time, set out to explore and develop the meaning of facial anthropomorphic trustworthiness, which is rarely discussed in prior literature. Specifically, this research explored the understanding of human facial trustworthiness toward AI agents via synthesizing the theories from two perspectives, interpersonal trustworthiness, and human-robot trustworthiness. While human facial trustworthiness was originated from evolutionary psychology, which emphasizes human facial features in signaling trust, human-robot facial trustworthiness arose from innate intention for anthropomorphism, which focuses on the specific attributes of robots. Thus, this study emerges these two kinds of knowledge together to provide a relatively holistic picture when we evaluated the trustworthiness of a social robot at a first impression.

- Prior research of facial trustworthiness was mainly within the context of human facial properties, thus only a limited number of researchers have tried to figure out whether the rules or conclusions could be applied in designing facial features for social robots. However, trust or trustworthiness, as one of the crucial social judgments, does not belong to human perceptions exclusively. Instead, it is a symbol by which people convey their expectations to themselves and others and this symbolic meaning is known to influence people's preference for a robot. Accordingly, it is academically interesting and necessary to explore the similarities and differences when making facial trustworthiness evaluations toward humans or robots. Through a series of behavioral experiments and modeling, the current study adopted triangulated data to suggest: 1) facial trustworthiness features could be applied to social robot design and improve people's trustworthiness and attitude toward the social robot; 2) some facial features, such as eye size, mouth size, eye shape, and mouth shape, shared similar effects on human facial trustworthiness and facial anthropomorphic trustworthiness; 3) some facial features, such as fWHR, might have a counter-intuitive or even opposite effect on human facial trustworthiness and facial anthropomorphic trustworthiness.

- Considering the emerging role of the social robot in our daily lives, limited research has tried to holistically examine the effect of different facial features on facial anthropomorphic trustworthiness. In order to address this research gap, it is necessary to investigate the separate and combinations of facial features in signaling trustworthiness in a more comprehensive way. Based on the data of series of experiments, the current study modeled the facial anthropomorphic trustworthiness for static and dynamic features. It might work as a pioneering study to address this research gap from an integrated perspective. By analyzing the degree of factors in advancing facial anthropomorphic trustworthiness, it could show the potential significant facial area and help to evaluate the perceived trustworthiness of different robots, contributing to the field of facial trustworthiness toward a social robot.

As for practical implications, although the current market has various social robots and some of them even won some honors (CES, 2018), detailed and specific guidelines are still missing to some extent that companies design a social robot primarily relying on their understandings and intuitions (Vanderborght et al., 2012). Regarding the risk in intuition-based design that might dampen user experience and potential market performance (Ulrich, 1992), this research provides preliminary instructions for designing a trustworthy social robot. This research advanced the design guidelines of a social robot, which could work as preliminary instructions in helping robot designer to make a trustworthy robot:

- Considering the numerous kinds of social robots in the current market, it might be difficult to evaluate the robot design and trustworthiness perceptions associated with the robot design. Since trustworthiness might work as a significant step for further interaction, it is indeed necessary to develop a protocol to evaluate robot design from a more objective perspective. This study provides an assessment toolkit, facial anthropomorphic trustworthiness for the social robot (FATSR-17), to evaluate the trustworthiness of a current or newly designed robot, which helps robot designers or relevant researchers to have an insight into the robot design. For example, FATSR-17 has four dimensions. When evaluating a robot design, a robot could be evaluated through a total score for facial trustworthiness as well as separated scores for anthropomorphism, capability, positive affect, and ethics concern. Accordingly, robot designers and researchers could enjoy more specific directions for making a trustworthy-looking robot.

- As a creative process for designing a social robot, robot designers might rely on their past experience and intuitions to create an "ideal" social robot. However, this process might be uncertain since they might lack sufficient knowledge of relevant user studies for making a trustworthy-looking robot. By empirically investigating facial anthropomorphic trustworthiness in human-robot interaction, design guidelines were specifically developed for each facial anthropomorphic feature, such as eye size, eye shape, mouth size, and mouth shape. Thus, these guidelines could provide design directions for the specific facial feature, avoiding the intuition-based design strategy in robot design and facilitating trustworthiness understanding for robot designers, and contributing to the design system for a social robot. Further, this study provides a systematic model of facial anthropomorphic trustworthiness for static and dynamic features. With the help of the model of facial anthropomorphic trustworthiness for the social robot, it could assist designers to analytically evaluate a relatively large number of proposed social robot designs. By evaluating the effectiveness of different facial features, the results can help designers to decide how to effectively, focusing on the significant facial area to signal perceived trustworthiness and avoid the potential non-significant facial areas. It can also inform designers by identifying possible causes of untrustworthy robot examples, which could severe as a more objective evaluation. With future iterative improvement and validation of the model, the advantages of this prediction tool will be further enhanced, and less time and fewer resources will be required.

- This design guideline could also contribute to the market and business performance of the social robot, promoting the acceptance toward such emerging AI agents. As a rule-of-thumb design principle, "beauty premium" or "plainness penalty", people are naturally prone to be attracted by a cute or beautiful appearance. Considering the vital role of trustworthiness in interpersonal interaction, individuals could enjoy more advantages of being looking trustworthy. Similarly, a trustworthy-looking social robot could also have the potential to attract the public. Take Buddy the Robot for example. The Paris-based company enjoyed a huge market success and had raised more than $600,000 since 2014. Parts of the reasons might go to its cute appearance and emotional expressions. However, seldom guidelines have been summarized for systematically designing a trustworthy-looking robot. Thus, the conclusion of this study could severe as a preliminary toolkit or protocol for designers to create a trustworthy-looking robot. In this way, the current research could help relevant stakeholders in the process of robot commercialization to have a high probability of market and business success.

Appendix A. Summarized facial features on anthropomorphic trustworthiness

Authors	Sample	Country	Application/ Purpose of study	Measure	Processing Technique	Results
Arminjon et al. (2015)	57		To test the effect of lying cues (LC) in guessing behavior.	Yes or no proportion	Repeated measures ANOVA	Compared with NLC, LC was significant to lying decisions and is related to the automatic processing of lying detection.
Calvo et al. (2017)	64	Spanish	To explore the effect of the combination of different mouth and eye on trustworthiness evaluation.	1-9 Likert scale; iNVT	Repeated measured ANOVA	Faces with an unfolding smile and eye looked more trustworthy. The contribution of the mouth was greater for happiness than for trustworthiness.
Dijk et al. (2011)	196	Dutch	To explore the effect of blushing on trustworthiness.	Trust game choice; 1-7 Likert scale	Two-way ANOVA	The blushing people were perceived to be more trustworthy.
Etcoff et al. (2011)	149		To evaluate the effect of color cosmetics on trustworthiness.	1-7 Likert scale	A linear mixed-effects model	Cosmetics can exaggerate cues to sexual dimorphism, improving trustworthiness.
Ferstl et al. (2017)	48		To explore the effect of facial features on the perceived personality and moral decisions.	1-7 Likert scale	A generalized linear mixed model	Human faces trustworthy features might not be consistent with abstract faces.
Gill et al. (2014)	12		To test the effect of phenotypic morphology on the default social features.	1-5 Likert scale	Correlation Analysis	The facial movement could predictably modulate perception of basic social features in face morphology.

Hellström and Tekle (1994)	75	Swedish	To evaluate the effects of different facial attributes (glasses, beard, and hair) on characteristic profiles.	1-6 Likert scale	Three-way ANOVA	The judges associated wearing glasses with intellectualism and goodness, being bald with idealism, and wearing a beard with unconventionality and goodness.
Johnston et al. (2010)	30	New Zealander	To investigate the effect of different types of smiling on attention.	1-7 Likert scale	Repeated-measures ANOVA	Enjoyment smiles are positively evaluated and are considered to have higher rates of cooperation.
Landwehr et al. (2011)	263		To investigate the effect of robot facial design on people's liking.	1000 points scale	Repeated-measures ANOVA	Perception of friendliness is associated with the robot with an upturned mouth, while aggressiveness is associated with the robot with both an upturned mouth and slanted eyes.
Kaisler and Leder (2016)	70	Austrian	To explore how eye contacting affects social and aesthetic evaluations.	1-7 Likert scale	Repeated-measures ANOVA	Direct-looking faces are considered to be more trustworthy.
Kleisner et al. (2013)	238	Czech Republic	To test whether eye color influences the perception of trustworthiness.	1-10 Likert scale	A generalized linear mixed model	Brown-eyed faces were perceived as more trustworthy and the reason lies in the facial features associated with them.
Luo et al. (2006)	183		To investigate whether or not the on-screen characters representation influence trustworthiness perception.	1-7 Likert scale	One way ANOVA and Paired t-tests	On-screen characters (OSCs) are considered to be more trustworthy in general. There is a mismatch between the expectations and capabilities of OSCs.
Ma et al. (2015)	139	Chinese	To explore how children judge trustworthiness from faces	1-3 Likert scale	Stepwise linear regressions	8-years children could use a similar inference to evaluate trustworthiness.
Maeng and Aggarwal (2018)	248		To explore the face width-to-height ratio (fWHR) can signal dominance and affect its overall evaluation	1-7 Likert scale	A linear mixed-effects analysis using lme4 and lmerTest	High fWHR robot is considered to be more dominant and liked more.

Masip et al. (2004)	324	Spanish	To examine the impact of facial maturity on impressions of truthfulness.	1-7 Likert scale	MANCOVA	Baby-face and age are perceived to be a significantly static cue to make trustworthiness evaluation.
Mathur and Reichling (2016)	334		To investigate whether human-robot interactions may be complicated by Uncanny Valley (UV).	Mean dollars wagered	Polynomial regression	The Uncanny Valley, in which imperfect human-likeness cues elicits dislike, could influence human perceptions of robots.
Okubo et al. (2013)	100	Japanese	To investigate the effect of a posed smile on people's attitudes.	response bias	Three-way ANOVA	The left–left composites were perceived to be more trustworthy when posed with a happy face.
Reed and DeScioli (2017)	218		To test whether fear expressions add credibility to a speaker's warnings of danger	1-7 Likert scale	Chi-square	Warning of danger with a fear expression is considered to be more trustworthy.
Santos and Young (2011)	Study 1: 24; Study 2: 48	UK	To investigate the importance of holistic processing in the inference of social attributes from faces.	1-7 Likert scale	Repeated-measures ANOVA	Experiment 1: internal features plays a more significant role in trustworthiness inferences. Experiment 2: different facial cues are used in different evaluations.
Sofer et al. (2015)	53	Israel	To test whether face typicality is an important factor for social perception.	1-9 Likert scale	Repeated-measures ANOVA	For a continuum of faces that vary on a typicality-attractiveness dimension, trustworthiness evaluations peak around the typical face.
Stanton and Stevens (2017)	52	Australia	To explore the relationship between gaze and trustworthiness evaluation	Mean answer change	Two-way ANOVA	People might trust the robot more on hard trials, compared with on medium trials. In addition, females are least likely to trust a robot which stared at them.
Stirrat and Perrett (2010)	62	UK	To explore the effect of fWHR on trustworthiness evaluation	The proportion of trust in the image.	A least squares regression	Wide face in men was perceived to be less trustworthy.

Verberne et al. (2015)	111	Dutch	To examine the effect of facial similarity on trust evaluation.	1-7 Likert scale	A one-way MANOVA	As the rules in human similarity, the similarity in the virtual agent would also be considered as more trustworthy.
Xu et al. (2012)	144	Chinese and Caucasian	To explore the difference in the ethnical group in trustworthiness evaluation.	1-9 Likert scale	A least squares regression	Chinese and Caucasian shared similar cues to make trustworthiness evaluation.
Zebrowitz et al. (1996)	103	US	To investigate the effect of age on trustworthiness evaluation.	1-7 Likert scale	Correlation analysis	Babyfaceness, attractiveness, facial symmetry, and large eyes had a significant impact on trustworthiness evaluation.

Appendix B. Ethical Approval

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

To	Luximon Yan (School of Design)		
From	SIU Kin Wai Michael, Chair, Departmental Research Committee		
Email	sdmsiu@	Date	10-Oct-2018

Application for Ethical Review for Teaching/Research Involving Human Subjects

I write to inform you that approval has been given to your application for human subjects ethics review of the following project for a period from 01-Sep-2018 to 31-Aug-2021:

Project Title:	Product Design and Design Psychology
Department:	School of Design
Principal Investigator:	Luximon Yan
Project Start Date:	01-Sep-2018
Reference Number:	HSEARS20181010001

You will be held responsible for the ethical approval granted for the project and the ethical conduct of the personnel involved in the project. In case the Co-PI, if any, has also obtained ethical approval for the project, the Co-PI will also assume the responsibility in respect of the ethical approval (in relation to the areas of expertise of respective Co-PI in accordance with the stipulations given by the approving authority).

You are responsible for informing the Human Subjects Ethics Sub-committee in advance of any changes in the proposal or procedures which may affect the validity of this ethical approval.

SIU Kin Wai Michael

Chair

Departmental Research Committee

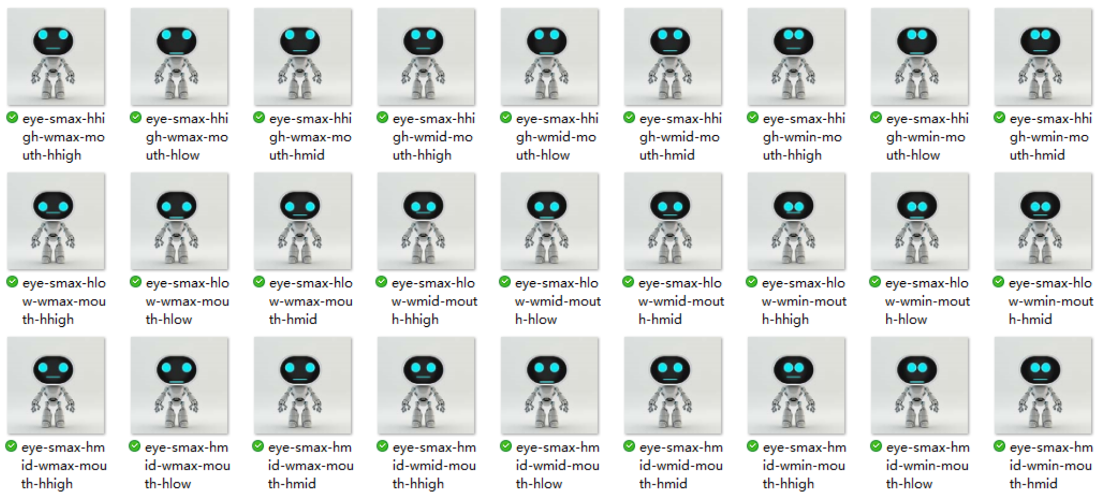
Appendix C. Original item pool

Ethics Concern	Capability	Positive Affect	Anthropomorphism
look evil	capable	friendly	uncanny
sufficient integrity	competent	conscious	an actual human face
self-awareness	perform duties	kind	natural appearance
ethics in programing	successful at the things	cute	appropriate features
sense of justice.	sufficient artificial intelligence	happy	too human-identical
stick to its program.	confident	well qualified	too close to humanity
looks fair	specialized capability	concerned welfare	a minimum of human appearance
behaviors are consistent.	satisfy users' needs	desires seem important	machine-like features
ethical concern	expect good advice	anything to hurt me	moderate lifelike
sound principles	dependable	important to me	weird face
predictable.	reliable	help me	complex detailed features
standards	autonomous	interested in welfare	neither too plain nor too weird
operates scrupulously	finish the task	put my interest first	neither too dull nor too freaky
statements	follow the advice	responsible	too boring nor too shocking
methods are clear	give me advice	supportive	balance between human and machine
keeps promises	rely on the advice	pleasant	neither too real nor too synthetic
protect human	function successfully	join our team	infantile like

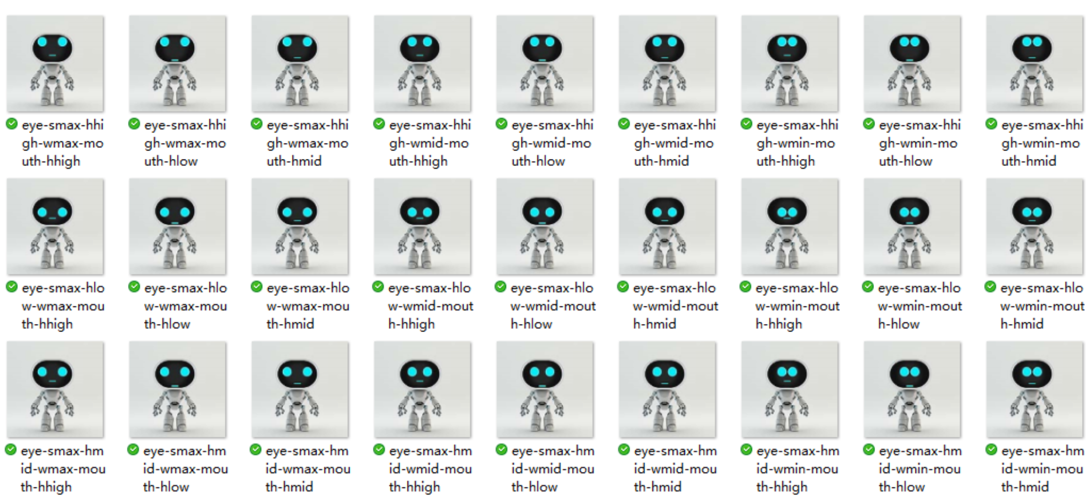
openly communicate	clearly communicate	aggressive	neither too humanoid nor too robotic
perform as instructed	frequent maintenance		neither too living nor too inanimate
obey order	better than a novice human user		
a competitor for job	provide feedback		
hacked easily	meet the need of the mission		
	provide appropriate information		

Appendix D. Robot Stimuli in Study 5

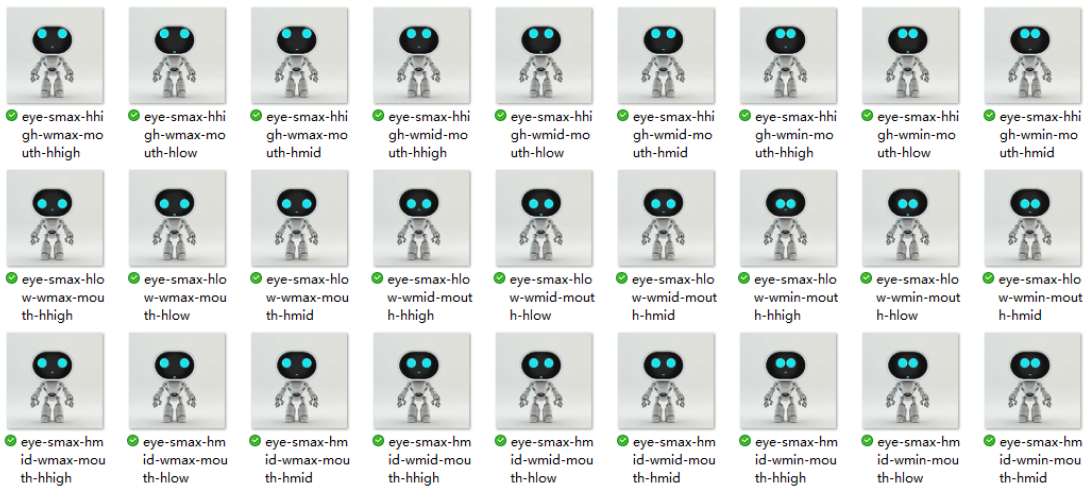
- Large eyes with a large mouth



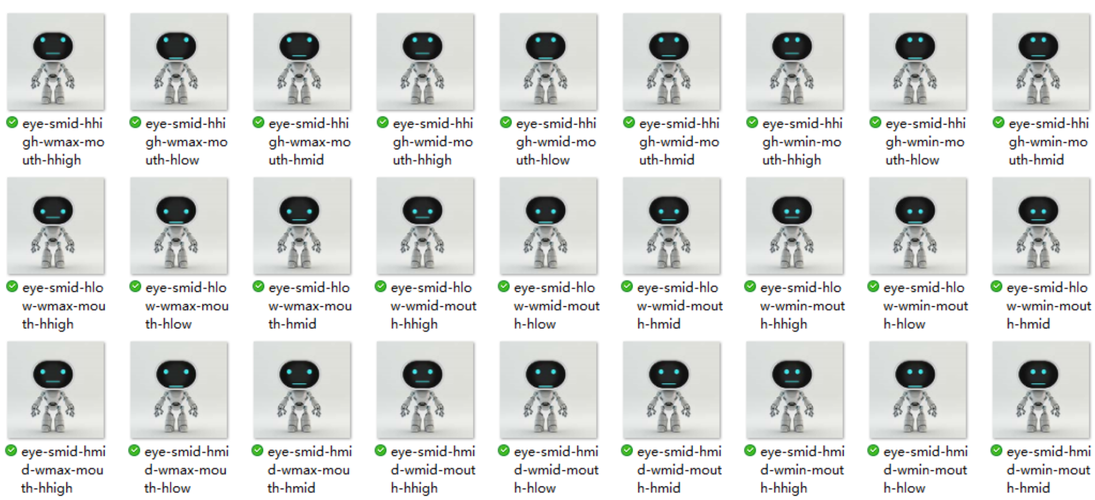
- Large eyes with a medium mouth



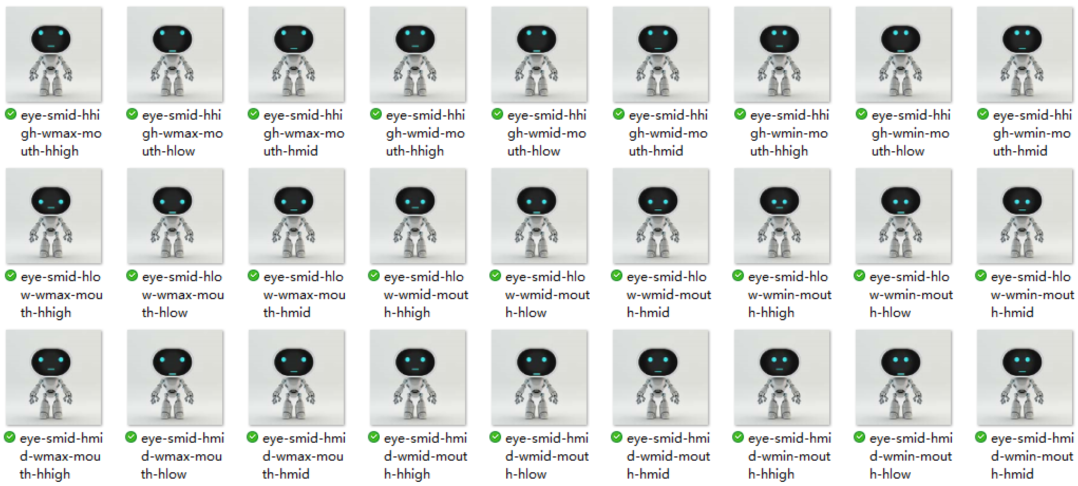
- Large eyes with a small mouth



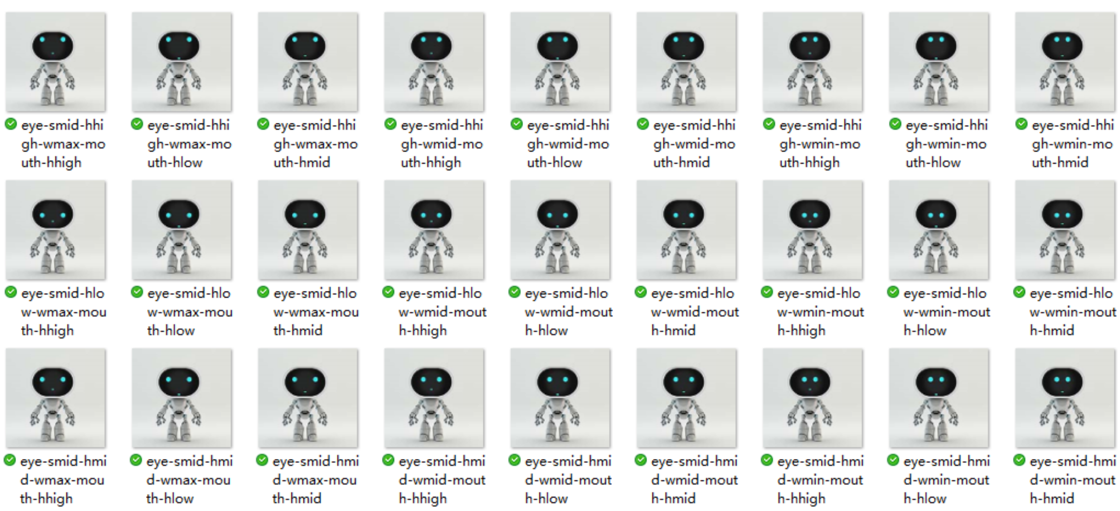
- Medium eyes with a large mouth



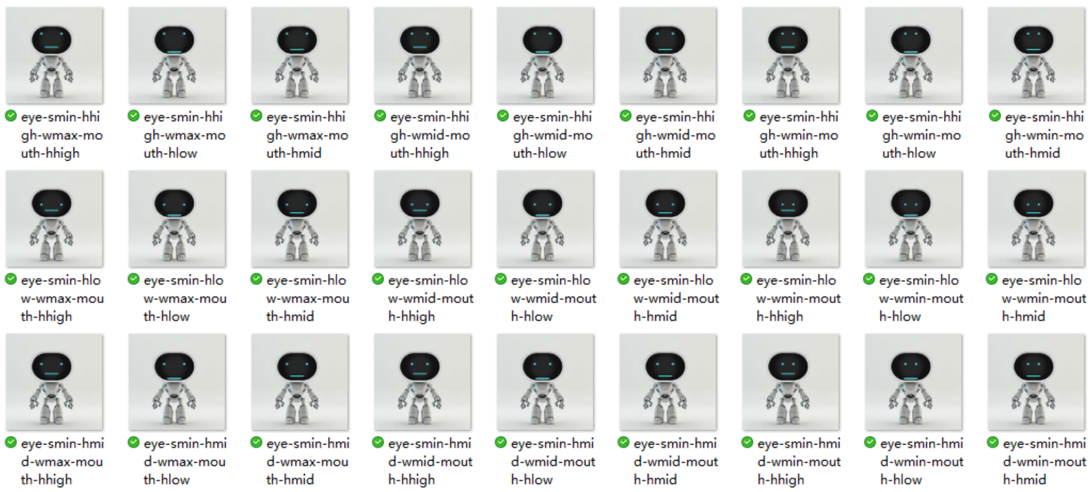
- Medium eyes with a medium mouth



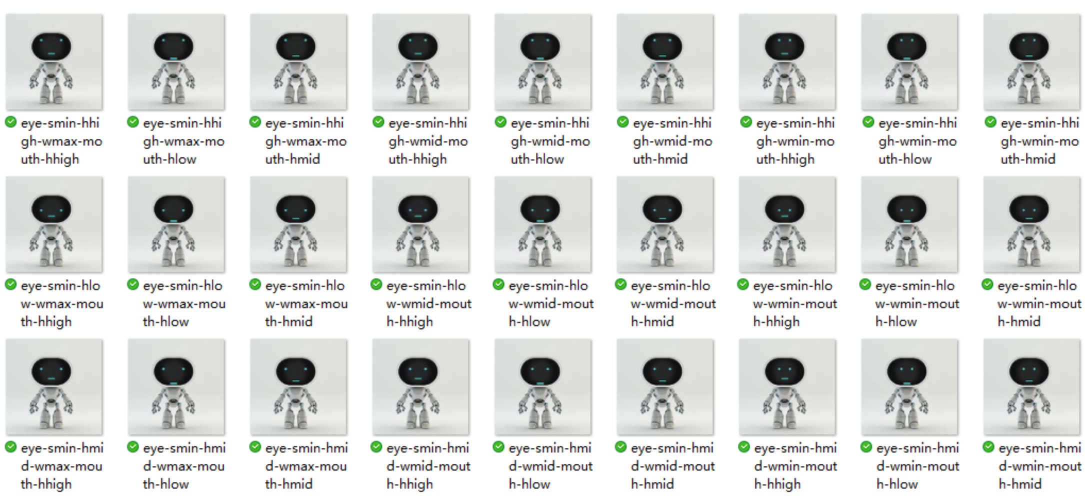
- Medium eyes with a small mouth



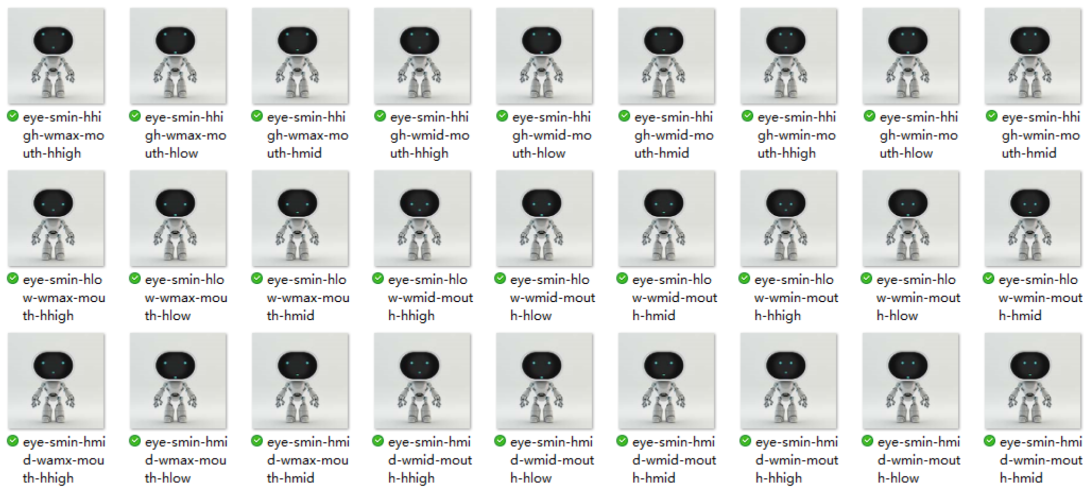
- Small eyes with a large mouth



- Small eyes with a medium mouth



- Small eyes with a small mouth



REFERENCES

Aaker, J.L., Lee, A.Y., 2006. Understanding Regulatory Fit. J. Mark. Res. 43, 15–19. https://doi.org/10.1509/jmkr.43.1.15

Aguado, L., Román, F.J., Fernández-Cahill, M., Diéguez-Risco, T., Romero-Ferreiro, V., 2011. Learning about Faces: Effects of Trustworthiness on Affective Evaluation. Span. J. Psychol. 14, 523–534. https://doi.org/10.5209/rev_sjop.2011.v14.n2.1

Aharon, I., Etcoff, N., Ariely, D., Chabris, C.F., O'Connor, E., Breiter, H.C., 2004. Beautiful Faces Have Variable Reward Value. Neuron 32, 537–551. https://doi.org/10.1016/s0896-6273(01)00491-3

Ajzen, I., 2001. Nature and Operation of Attitudes. Annu. Rev. Psychol. 52, 27–58. https://doi.org/10.1146/annurev.psych.52.1.27

Al-Qaderi, M., Rad, A., Al-Qaderi, M.K., Rad, A.B., 2018. A Multi-Modal Person Recognition System for Social Robots. Appl. Sci. 8, 387. https://doi.org/10.3390/app8030387

Ames, D.L., Fiske, S.T., Todorov, A., 2010. Impression Formation: A Focus on Others' Intents. Handb. Soc. Neurosci. 1–41. https://doi.org/10.1093/oxfordhb/9780195342161.013.0028

AMT, 2021. Amazon Mechanical Turk [WWW Document]. Mturk. URL https://www.mturk.com/ (accessed 3.30.21).

Appel, M., Izydorczyk, D., Weber, S., Mara, M., Lischetzke, T., 2020. The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. Comput. Human Behav. 102, 274–286. https://doi.org/10.1016/j.chb.2019.07.031

Arminjon, M., Chamseddine, A., Kopta, V., Paunović, A., Mohr, C., 2015. Are we modular lying cues detectors? The answer is "yes, sometimes." PLoS One 10, 1–16. https://doi.org/10.1371/journal.pone.0136418

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, in: Proceedings of the Human Factors and Ergonomics Society. pp. 303–307. https://doi.org/10.1177/1071181312561071

Avnet, T., Higgins, E.T., 2006. How regulatory fit affects value in consumer choices and opinions. J. Mark. Res. 43, 1–10. https://doi.org/10.1509/jmkr.43.1.1

Babbie, E., 2008. The basics of social research Thomson Wadsworth, Belmont, CA. Wadsworth / Cengage Learning.

Bakmazian, A., 2014. The Man Behind the Beard: Perception of Men's Trustworthiness as a Function of Facial Hair. Psychology 05, 185–191. https://doi.org/10.4236/psych.2014.53029

Bar, M., Neta, M., Linz, H., 2006. Very first impressions. Emotion 6, 269–278. https://doi.org/10.1037/1528-3542.6.2.269

Bargh, J.A., Shalev, I., 2012. The substitutability of physical and social warmth in daily life. Emotion 12, 154–62. https://doi.org/10.1037/a0023527

Bart, Y., Shankar, V., Sultan, F., Urban, G.L., 2005. Are the drivers and role of online trust the same for all web sites and consumers? a large-scale exploratory empirical study. J. Mark. 69, 133–152. https://doi.org/10.1509/jmkg.2005.69.4.133

Bartneck, C., Bleeker, T., Bun, J., Fens, P., Riet, L., 2010. The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. Paladyn, J. Behav. Robot. 1, 109–115. https://doi.org/10.2478/s13230-010-0011-3

Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., Tapp, R., 2014. Featural processing in recognition of emotional facial expressions. Cogn. Emot. 28, 416–432. https://doi.org/10.1080/02699931.2013.833500

Bellotti, V., Bly, S., 1996. Walking away from the desktop computer, in: Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work - CSCW '96. ACM Press, New York, New York, USA, pp. 209–218. https://doi.org/10.1145/240080.240256

Bergkvist, L., Rossiter, J.R., 2007. The predictive validity of multiple-item versus single-item measures of the same constructs. J. Mark. Res. 44, 175–184. https://doi.org/10.1509/jmkr.44.2.175

Berry, D.S., 1991. Attractive Faces Are not all Created Equal: Joint Effects of Facial Babyishness and Attractiveness on Social Perception. Personal. Soc. Psychol. Bull. 17, 523–531. https://doi.org/10.1177/0146167291175007

Berry, D.S., McArthur, L.Z., 1985. Some Components and Consequences of a Babyface. [Miscellaneous Article]. J. Pers.

Bhattacherjee, A., 2002. Individual trust in online firms: Scale development and initial test. J. Manag. Inf. Syst. 19, 211–241. https://doi.org/10.1080/07421222.2002.11045715

Birkás, B., Dzhelyova, M., Lábadi, B., Bereczkei, T., Perrett, D.I., 2014. Cross-cultural perception of trustworthiness: The effect of ethnicity features on evaluation of faces' observed trustworthiness across four samples. Pers. Individ. Dif. 69, 56–61.

https://doi.org/10.1016/j.paid.2014.05.012

Blijlevens, J., Carbon, C.-C., Mugge, R., Schoormans, J.P.L., 2012. Aesthetic appraisal of product designs: Independent effects of typicality and arousal. Br. J. Psychol. 103, 44–57. https://doi.org/10.1111/j.2044-8295.2011.02038.x

Blijlevens, J., Hekkert, P., Leder, H., Thurgood, C., Chen, L.L., Whitfield, T.W.A., 2017. The aesthetic pleasure in design scale: The development of a scale to measure aesthetic pleasure for designed artifacts. Psychol. Aesthetics, Creat. Arts 11, 86–98. https://doi.org/10.1037/aca0000098

Bloch, P.H., Brunel, F.F., Arnold, T.J., 2003. Individual Differences in the Centrality of Visual Product Aesthetics: Concept and Measurement. J. Consum. Res. 29, 551–565. https://doi.org/10.1086/346250

Blue Frog Robotics, 2018. Buddy Robot [WWW Document]. URL https://commons.wikimedia.org/wiki/File:BUDDY_le_robot_émotionnel.jpg (accessed 2.7.20).

Bombari, D., Schmid, P.C., Schmid Mast, M., Birri, S., Mast, F.W., Lobmaier, J.S., 2013. Emotion recognition: The role of featural and configural face information. Q. J. Exp. Psychol. 66, 2426–2442. https://doi.org/10.1080/17470218.2013.789065

Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D.I., Marlow, C., Settle, J.E., Fowler, J.H., 2012. A 61-million-person experiment in social influence and political mobilization. Nature. https://doi.org/10.1038/nature11421

Bonnefon, J.F., Hopfensitz, A., Neys, W. De, 2013. The modular nature of trustworthiness detection. J. Exp. Psychol. Gen. 142, 143–150. https://doi.org/10.1037/a0028930

Borgi, M., Cogliati-Dezza, I., Brelsford, V., Meints, K., Cirulli, F., 2014. Baby schema in human and animal faces induces cuteness perception and gaze allocation in children. Front. Psychol. https://doi.org/10.3389/fpsyg.2014.00411

Bosmans, A., Baumgartner, H., 2005. Goal-Relevant Emotional Information: When Extraneous Affect Leads to Persuasion and When It Does Not. J. Consum. Res. 32, 424–434. https://doi.org/10.1086/497554

Boynton, M.H., Richman, L.S., 2014. An online daily diary study of alcohol use using Amazon's mechanical turk. Drug Alcohol Rev. 33, 456–461. https://doi.org/10.1111/dar.12163

Braithwaite, J., Watson, D., Robert, J., Mickey, R., 2013. A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments, ….

Brañas-Garza, P., Capraro, V., Rascón-Ramírez, E., 2018. Gender differences in altruism on mechanical turk: expectations and actual behaviour. Econ. Lett. 170, 19–23. https://doi.org/10.1016/J.ECONLET.2018.05.022

Brosch, T., Sander, D., Scherer, K.R., 2007. That Baby Caught My Eye... Attention Capture by Infant Faces. Emotion 7, 685–689. https://doi.org/10.1037/1528-3542.7.3.685

Brownlow, S., 1992. Seeing is believing: Facial appearance, credibility, and attitude change. J. Nonverbal Behav. 16, 101–115. https://doi.org/10.1007/BF00990325

Buckingham, G., DeBruine, L.M., Little, A.C., Welling, L.L.M., Conway, C.A., Tiddeman, B.P., Jones, B.C., 2006. Visual adaptation to masculine and feminine faces influences generalized preferences and perceptions of trustworthiness. Evol. Hum. Behav. 27, 381–389. https://doi.org/10.1016/J.EVOLHUMBEHAV.2006.03.001

Calvo, M.G., Álvarez-Plaza, P., Fernández-Martín, A., 2017. The contribution of facial regions to judgements of happiness and trustworthiness from dynamic expressions. J. Cogn. Psychol. 29, 618–625. https://doi.org/10.1080/20445911.2017.1302450

Calvo, M.G., Krumhuber, E.G., Fernández-Martín, A., 2019. Visual attention mechanisms in happiness versus trustworthiness processing of facial expressions. Q. J. Exp. Psychol. 72, 729–741. https://doi.org/10.1177/1747021818763747

Calvo, M.G., Nummenmaa, L., 2008. Detection of Emotional Faces: Salient Physical Features Guide Effective Visual Search. J. Exp. Psychol. Gen. 137, 471–494. https://doi.org/10.1037/a0012771

Campellone, T.R., Kring, A.M., 2013. Who do you trust? The impact of facial emotion and behaviour on decision making. Cogn. Emot. 27, 603–620. https://doi.org/10.1080/02699931.2012.726608

Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M., Seabright, P., 2015. Honest signaling in trust interactions: Smiles rated as genuine induce trust and signal higher earning opportunities. Evol. Hum. Behav. 36, 8–16. https://doi.org/10.1016/j.evolhumbehav.2014.08.001

CES, 2018. BUDDY named CES 2018 BEST OF INNOVATION AWARDS Honoree - BUDDY The Emotional Robot [WWW Document]. CES. URL https://buddytherobot.com/en/news/buddy-named-ces-2018-best-innovation-awards-honoree/ (accessed 4.10.19).

Cesario, J., Corker, K.S., Jelinek, S., 2013. A self-regulatory framework for message framing. J. Exp. Soc. Psychol. 49, 238–249. https://doi.org/10.1016/j.jesp.2012.10.014

Chanseau, A., Dautenhahn, K., Koay, K.L., Salem, M., 2016. Who is in charge? Sense of control and robot anxiety in Human-Robot Interaction, in: 25th IEEE International

Symposium on Robot and Human Interactive Communication, RO-MAN 2016. Institute of Electrical and Electronics Engineers Inc., pp. 743–748. https://doi.org/10.1109/ROMAN.2016.7745202

Chen, H., Russell, R., Nakayama, K., 2010. Crossing the “uncanny valley”: Adaptation to cartoon faces can influence perception of human faces. Perception 39, 378–386. https://doi.org/10.1068/p6492

Chen, T., Xu, R., He, Y., Wang, X., 2017. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Syst. Appl. 72, 221–230. https://doi.org/10.1016/j.eswa.2016.10.065

Chiller-Glaus, S.D., Schwaninger, A., Hofer, F., Kleiner, M., Knappmeyer, B., 2011. Recognition of emotion in moving and static composite faces. Swiss J. Psychol. 70, 233–240. https://doi.org/10.1024/1421-0185/a000061

Cohen, J., 2013. Statistical Power Analysis for the Behavioral Sciences, Statistical Power Analysis for the Behavioral Sciences. https://doi.org/10.4324/9780203771587

Collins, K., 1990. The qualitative report : an online journal dedicated to qualitative research since 1990., The Qualitative Report. Nova Southeastern University, School of Social and Systematic Studies.

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. J. Appl. Psychol. 92, 909–927. https://doi.org/10.1037/0021-9010.92.4.909

Cowell, A.J., Stanney, K.M., 2005. Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. Int. J. Hum. Comput. Stud. 62, 281–306. https://doi.org/10.1016/j.ijhcs.2004.11.008

Creusen, M.E.H., Schoormans, J.P.L., 2005. The Different Roles of Product Appearance in Consumer Choice. J. Prod. Innov. Manag. 22, 63–81. https://doi.org/10.1111/j.0737-6782.2005.00103.x

Crowe, E., Higgins, E.T., 1997. Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. Organ. Behav. Hum. Decis. Process. 69, 117–132. https://doi.org/10.1006/obhd.1996.2675

Darwin, G.H., 1872. Development in dress, Macmillan’s Magazine.

De Rie, R., 2016. Empathy with and projecting feelings onto robots from schemas about humans. Int. J. Psychol. 51.

Deal, S.B., Lendvay, T.S., Haque, M.I., Brand, T., Comstock, B., Warren, J., Alseidi, A.,

2016. Crowd-sourced assessment of technical skills: an opportunity for improvement in the assessment of laparoscopic surgical skills. Am. J. Surg. 211, 398–404. https://doi.org/10.1016/j.amjsurg.2015.09.005

Dehn, D.M., Van Mulken, S., 2000. Impact of animated interface agents: a review of empirical research. Int. J. Hum. Comput. Stud. 52, 1–22. https://doi.org/10.1006/ijhc.1999.0325

Del Río, N.G., González-González, C.S., Toledo-Delgado, P.A., Muñoz-Cruz, V., García-Peñalvo, F., 2020. Health promotion for childhood obesity: An approach based on self-tracking of data. Sensors (Switzerland) 20, 1–28. https://doi.org/10.3390/s20133778

Delgado-Ballester, E., 2004. Applicability of a brand trust scale across product categories. Eur. J. Mark. 38, 573–592. https://doi.org/10.1108/03090560410529222

Deng, L., 2014. Deep Learning: Methods and Applications. Found. Trends® Signal Process. 7, 197–387. https://doi.org/10.1561/2000000039

Desor, J.A., Beauchamp, G.K., 1974. The human capacity to transmit olfactory information. Percept. Psychophys. 16, 551–556. https://doi.org/10.3758/BF03198586

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv.

Dijk, C., Koenig, B., Ketelaar, T., de Jong, P.J., 2011. Saved by the blush: Being trusted despite defecting. Emotion 11, 313–319. https://doi.org/10.1037/a0022774

Ding, Y., Cao, Y., Duffy, V.G., Wang, Y., Zhang, X., 2020. Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning. Ergonomics 63, 896–908. https://doi.org/10.1080/00140139.2020.1759699

Dong, Y., Liu, Y., Jia, Y., Li, Y., Li, C., 2018. Effects of facial expression and facial gender on judgment of trustworthiness: The modulating effect of cooperative and competitive settings. Front. Psychol. 9, 1–5. https://doi.org/10.3389/fpsyg.2018.02022

Dotsch, R., Hassin, R.R., Todorov, A., 2017. Statistical learning shapes face evaluation. Nat. Hum. Behav. 1, 1–6. https://doi.org/10.1038/s41562-016-0001

Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W.S., Wilson, J.R., van der Doelen, B., 2012. A strategy for human factors/ergonomics: developing the discipline and profession. Ergonomics 55, 377–395. https://doi.org/10.1080/00140139.2012.661087

Dunn, J.R., Schweitzer, M.E., 2005. Feeling and believing: The influence of emotion on trust. J. Pers. Soc. Psychol. 88, 736–748. https://doi.org/10.1037/0022-3514.88.5.736

Dzhelyova, M., Perrett, D.I., Jentzsch, I., 2012. Temporal dynamics of trustworthiness perception. Brain Res. 1435, 81–90. https://doi.org/10.1016/j.brainres.2011.11.043

Elfenbein, H.A., Ambady, N., 2002. On the universality and cultural specificity of emotion recognition: A meta-analysis. Psychol. Bull. 128, 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Ellingsen, E.F., Drevsjø, S., Volden, F., Watten, R.G., 2019. Extraversion and focus of attention on facial emotions: an experimental eye-tracking study. Curr. Issues Personal. Psychol. 7, 91–97. https://doi.org/10.5114/cipp.2019.85413

Engell, A.D., Todorov, A., Haxby, J. V., 2010. Common neural mechanisms for the evaluation of facial trustworthiness and emotional expressions as revealed by behavioral adaptation. Perception 39, 931–941. https://doi.org/10.1068/p6633

Epley, N., Waytz, A., Cacioppo, J.T., 2007. On Seeing Human: A Three-Factor Theory of Anthropomorphism. Psychol. Rev. 114, 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Etcoff, N.L., 1994. Beauty and the beholder. Nature. https://doi.org/10.1038/368186a0

Etcoff, N.L., Stock, S., Haley, L.E., Vickery, S.A., House, D.M., 2011. Cosmetics as a feature of the extended human phenotype: Modulation of the perception of biologically important facial signals. PLoS One 6, 1–9. https://doi.org/10.1371/journal.pone.0025656

Ewe, S.Y., Gul, F.A., Lee, C.K.C., Yang, C.Y., 2018. The role of regulatory focus and information in investment choice: some evidence using visual cues to frame regulatory focus. J. Behav. Financ. 19, 89–100. https://doi.org/10.1080/15427560.2017.1373283

Falck, A.-C., Rosenqvist, M., 2012. What are the obstacles and needs of proactive ergonomics measures at early product development stages? – An interview study in five Swedish companies. Int. J. Ind. Ergon. 42, 406–415. https://doi.org/10.1016/J.ERGON.2012.05.002

Farmer, H., McKay, R., Tsakiris, M., 2014. Trust in me: trustworthy others are seen as more physically similar to the self. Psychol. Sci. 25, 290–2. https://doi.org/10.1177/0956797613494852

Fasola, J., Mataric, M.J., 2015. Evaluation of a spatial language interpretation framework for natural human-robot interaction with older adults, in: 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE, pp. 301–308. https://doi.org/10.1109/ROMAN.2015.7333611

Fernández-Rodicio, E., Castro-González, Á., Alonso-Martín, F., Maroto-Gómez, M., Salichs, M.Á., 2020. Modelling Multimodal Dialogues for Social Robots Using

Communicative Acts. Sensors 20, 3440. https://doi.org/10.3390/s20123440

Ferstl, Y., Kokkinara, E., Mcdonnell, R., 2017. Facial Features of Non-player Creatures Can Influence Moral Decisions in Video Games. ACM Trans. Appl. Percept. 15, 1–12. https://doi.org/10.1145/3129561

Ferstl, Y., Kokkinara, E., McDonnell, R., 2016. Do I trust you, abstract creature?: A study on personality perception of abstract virtual faces. Proc. ACM Symp. Appl. Perception, SAP 2016 39–43. https://doi.org/10.1145/2931002.2931014

Fink, B., Grammer, K., Thornhill, R., 2001. Human (Homo sapiens) facial attractiveness in relation to skin texture and color. J. Comp. Psychol. 115, 92–99. https://doi.org/10.1037/0735-7036.115.1.92

Foo, Y.Z., Simmons, L.W., Rhodes, G., 2017. Predictors of facial attractiveness and health in humans. Sci. Rep. 7, 1–12. https://doi.org/10.1038/srep39731

Fornell, C., Larcker, D.F., 2006. Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. J. Mark. Res. 18, 382. https://doi.org/10.2307/3150980

Förster, F., Förster, Frank, 2019. Enactivism and Robotic Language Acquisition: A Report from the Frontier. Philosophies 4, 11. https://doi.org/10.3390/philosophies4010011

Fortunati, L., 2015. Social Robots from a Human Perspective, Social Robots from a Human Perspective. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-15672-9

Fosch-Villaronga, E., Lutz, C., Tamò-Larrieux, A., 2019. Gathering Expert Opinions for Social Robots' Ethical, Legal, and Societal Concerns: Findings from Four International Workshops. Int. J. Soc. Robot. 1–18. https://doi.org/10.1007/s12369-019-00605-z

Fox, W.S., Denzin, N.K., 1979. The Research Act: A Theoretical Introduction to Sociological MethodsSociological Methods: A Sourcebook, Contemporary Sociology. Taylor and Francis. https://doi.org/10.2307/2065439

Frey, J.H., Fontana, A., 1991. The group interview in social research. Soc. Sci. J. 28, 175–187. https://doi.org/10.1016/0362-3319(91)90003-M

Friedenberg, J., 2001. Lateral Feature Displacement and Perceived Facial Attractiveness. Psychol. Rep. 88, 295–305. https://doi.org/10.2466/pr0.2001.88.1.295

Frith, C., 2009. Role of facial expressions in social interactions. Philos. Trans. R. Soc. B Biol. Sci. 364, 3453–3458. https://doi.org/10.1098/rstb.2009.0142

Funk, F., Todorov, A., 2013. Criminal stereotypes in the courtroom: Facial tattoos affect guilt and punishment differently. Psychol. Public Policy, Law 19, 466–478. https://doi.org/10.1037/a0034736

Galinsky, D.F., Erol, E., Atanasova, K., Bohus, M., Krause-Utz, A., Lis, S., 2020. Do I trust you when you smile? Effects of sex and emotional expression on facial trustworthiness appraisal. PLoS One 15, e0243230. https://doi.org/10.1371/journal.pone.0243230

Gantiva, C., Sotaquirá, M., Araujo, A., Cuervo, P., 2019. Cortical processing of human and emoji faces: an ERP analysis. Behav. Inf. Technol. https://doi.org/10.1080/0144929X.2019.1632933

Gardner, D.G., Cummings, L.L., Dunham, R.B., Pierce, J.L., 1998. Single-Item Versus Multiple-Item Measurement Scales: An Empirical Comparison. Educ. Psychol. Meas. 58, 898–915. https://doi.org/10.1177/0013164498058006003

Ghazali, A.S., Ham, J., Barakova, E.I., Markopoulos, P., 2018. Effects of Robot Facial Characteristics and Gender in Persuasive Human-Robot Interaction. Front. Robot. AI 5, 1–16. https://doi.org/10.3389/frobt.2018.00073

Gill, D., Garrod, O.G.B., Jack, R.E., Schyns, P.G., 2014. Facial Movements Strategically Camouflage Involuntary Social Signals of Face Morphology. Psychol. Sci. 25, 1079–1086. https://doi.org/10.1177/0956797614522274

Glocker, M.L., Langleben, D.D., Ruparel, K., Loughead, J.W., Gur, R.C., Sachser, N., 2009a. Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults. Ethology 115, 257–263. https://doi.org/10.1111/j.1439-0310.2008.01603.x

Glocker, M.L., Langleben, D.D., Ruparel, K., Loughead, J.W., Valdez, J.N., Griffin, M.D., Sachser, N., Gur, R.C., 2009b. Baby schema modulates the brain reward system in nulliparous women. Proc. Natl. Acad. Sci. U. S. A. 106, 9115–9119. https://doi.org/10.1073/pnas.0811620106

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

Gompei, T., Umemuro, H., 2018. Factors and Development of cognitive and affective trust on social robots, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Cham, pp. 45–54. https://doi.org/10.1007/978-3-030-05204-1_5

Gomulya, D., Wong, E.M., Ormiston, M.E., Boeker, W., 2017. The role of facial appearance on CEO selection after firm misconduct. J. Appl. Psychol. 102, 617–635. https://doi.org/10.1037/apl0000172

González-González, C.S., Guzmán-Franco, M.D., Infante-Moro, A., 2019a. Tangible technologies for childhood education: A systematic review. Sustain. https://doi.org/10.3390/su11102910

González-González, C.S., Herrera-González, E., Moreno-Ruiz, L., Reyes-Alonso, N., Hernández-Morales, S., Guzmán-Franco, M.D., Infante-Moro, A., 2019b. Computational thinking and down syndrome: An exploratory study using the KIBO robot. Informatics 6. https://doi.org/10.3390/informatics6020025

González-González, C.S., Infante-Moro, A., Infante-Moro, J.C., 2020. Implementation of e-proctoring in online teaching: A study about motivational factors. Sustain. 12. https://doi.org/10.3390/SU12083488

Gorn, G.J., Jiang, Y., Johar, G.V., 2008. Babyfaces, Trait Inferences, and Company Evaluations in a Public Relations Crisis. J. Consum. Res. 35, 36–49. https://doi.org/10.1086/529533

Goucher-Lambert, K., Cagan, J., 2019. Crowdsourcing inspiration: Using crowd generated inspirational stimuli to support designer ideation. Des. Stud. 61, 1–29. https://doi.org/10.1016/j.destud.2019.01.001

Gray, K., Wegner, D.M., 2012. Feeling robots and human zombies: Mind perception and the uncanny valley. Cognition 125, 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Green, P.E., Carroll, J.D., Goldberg, S.M., 1981. A General Approach to Product Design Optimization Via Conjoint Analysis. J. Mark. 45, 17. https://doi.org/10.2307/1251539

Groeneveld, R.A., Meeden, G., 1984. Measuring Skewness and Kurtosis. Stat. 33, 391. https://doi.org/10.2307/2987742

Guido, G., Peluso, A.M., 2009. When are baby-faced endorsers appropriate? Testing effects on credibility and purchase intention. J. Curr. Issues Res. Advert. 31, 67–74. https://doi.org/10.1080/10641734.2009.10505266

Guo, F., Li, M., Hu, M., Li, F., Lin, B., 2019. Distinguishing and quantifying the visual aesthetics of a product: An integrated approach of eye-tracking and EEG. Int. J. Ind. Ergon. 71, 47–56. https://doi.org/10.1016/j.ergon.2019.02.006

Gupta, R., Hur, Y.J., Lavie, N., 2016. Distracted by pleasure: Effects of positive versus negative valence on emotional capture under load. Emotion 16, 328–337. https://doi.org/10.1037/emo0000112

Gutiérrez-García, A., Calvo, M.G., 2016a. Social anxiety and trustworthiness judgments of dynamic facial expressions of emotion. J. Behav. Ther. Exp. Psychiatry 52, 119–127. https://doi.org/10.1016/j.jbtep.2016.04.003

Gutiérrez-García, A., Calvo, M.G., 2016b. Social anxiety and perception of (un)trustworthiness in smiling faces. Psychiatry Res. 244, 28–36. https://doi.org/10.1016/j.psychres.2016.07.004

Haag, A., Goronzy, S., Schaich, P., Williams, J., 2004. Emotion recognition using bio-sensors: First steps towards an automatic system, in: Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science). Springer Verlag, pp. 36–48. https://doi.org/10.1007/978-3-540-24842-2_4

Hall, C.C., Goren, A., Chaiken, S., Todorov, A., 2010. Shallow Cues With Deep Effects: Trait Judgments From Faces and Voting Decisions, in: The Political Psychology of Democratic Citizenship. https://doi.org/10.1093/acprof:oso/9780195335453.003.0004

Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y.C., De Visser, E.J., Parasuraman, R., 2011. A meta-analysis of factors affecting trust in human-robot interaction. Hum. Factors 53, 517–527. https://doi.org/10.1177/0018720811417254

Harber, P., Leroy, G., 2015. Assessing Work–Asthma Interaction With Amazon Mechanical Turk. J. Occup. Environ. Med. 57, 381–385. https://doi.org/10.1097/JOM.0000000000000360

Haselhuhn, M.P., Wong, E.M., Ormiston, M.E., 2013. Self-Fulfilling Prophecies as a Link between Men's Facial Width-to-Height Ratio and Behavior. PLoS One 8. https://doi.org/10.1371/journal.pone.0072259

Hedt, B.L., Pagano, M., 2011. Health indicators: Eliminating bias from convenience sampling estimators. Stat. Med. 30, n/a-n/a. https://doi.org/10.1002/sim.3920

Hellström, Å., Tekle, J., 1994. Person perception through facial photographs: Effects of glasses, hair, and beard on judgments of occupation and personal qualities. Eur. J. Soc. Psychol. 24, 693–705. https://doi.org/10.1002/ejsp.2420240606

Higgins, E.T., 2000. Making a Good Decison: Value Form Fit. Am. Psychol. 30, 1217–1230.

Higgins, E.T., 1998. Promotion and Prevention: Regulatory Focus as A Motivational Principle. Adv. Exp. Soc. Psychol. 30, 1–46. https://doi.org/10.1016/S0065-2601(08)60381-0

Hildebrandt, K.A., Fitzgerald, H.E., 1979. Facial feature determinants of perceived infant attractiveness. Infant Behav. Dev. 2, 329–339. https://doi.org/10.1016/S0163-6383(79)80043-0

Ho, C.C., MacDorman, K.F., 2017. Measuring the Uncanny Valley Effect: Refinements to Indices for Perceived Humanness, Attractiveness, and Eeriness. Int. J. Soc. Robot. 9, 129–139. https://doi.org/10.1007/s12369-016-0380-9

Hoff, K.A., Bashir, M., 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. Hum. Factors 57, 407–434. https://doi.org/10.1177/0018720814547570

Holliday, R., 1993. Evolutionary Biology of Aging. Michael R. Rose , The Quarterly Review of Biology. https://doi.org/10.1086/417928

Hoorn, J.F., 2018. Theory of Robot Communication : II . Befriending a Robot over Time 2502572, 1–21.

Hoorn, J.F., 2015. The Handbook of the Psychology of Communication Technology, The Handbook of the Psychology of Communication Technology. Wiley. https://doi.org/10.1002/9781118426456

Hsiao, S.-W., Huang, H.., 2002. A neural network based approach for product form design. Des. Stud. 23, 67–84. https://doi.org/10.1016/S0142-694X(01)00015-1

Hwang, J., Park, T., Hwang, W., 2013. The effects of overall robot shape on the emotions invoked in users and the perceived personalities of robot. Appl. Ergon. 44, 459–471. https://doi.org/10.1016/j.apergo.2012.10.010

Ichikawa, H., Kanazawa, S., Yamaguchi, M.K., 2011. Finding a face in a face-like object. Perception 40, 500–502. https://doi.org/10.1068/p6926

J. A. Russell, 1980. A circumplex model of affect. J. Pers. Soc. Psychol.

Jaeger, S.R., Roigard, C.M., Jin, D., Vidal, L., Ares, G., 2019. Valence, arousal and sentiment meanings of 33 facial emoji: Insights for the use of emoji in consumer research. Food Res. Int. 119, 895–907. https://doi.org/10.1016/j.foodres.2018.10.074

Jentsch, E., 1997. On the psychology of the uncanny (1906). Angelaki - J. Theor. Humanit. 2, 7–16. https://doi.org/10.1080/09697259708571910

Jiao, J., Chen, C.H., Kerr, C., 2006. Customer requirement management in product development. Concurr. Eng. Res. Appl. https://doi.org/10.1177/1063293X06068355

Jin, S.A.A., 2010. The roles of regulatory focus and medical recommendation avatars' trustworthiness in virtual environment-based E-health, Cyberpsychology, Behavior, and Social Networking. https://doi.org/10.1089/cyber.2009.0291

Johnson, P.A., Sieber, R.E., Magnien, N., Ariwi, J., 2012. Automated web harvesting to collect and analyse user-generated content for tourism. Curr. Issues Tour. https://doi.org/10.1080/13683500.2011.555528

Johnson, R.B., Onwuegbuzie, A.J., 2007. Toward a Definition of Mixed Methods Research. J. Mix. Methods Res. 1, 112–133. https://doi.org/10.1177/1558689806298224

Johnson, R.B., Onwuegbuzie, A.J., 2004. Mixed Methods Research: A Research Paradigm Whose Time Has Come. Educ. Res. 33, 14–26.

https://doi.org/10.3102/0013189X033007014

Johnston, L., Miles, L., Macrae, C.N., 2010. Why are you smiling at me? Social functions of enjoyment and non-enjoyment smiles. Br. J. Soc. Psychol. 49, 107–127. https://doi.org/10.1348/014466609X412476

Jones, B.C., Little, A.C., Penton-Voak, I.S., Tiddeman, B.P., Burt, D.M., Perrett, D.I., 2001. Facial symmetry and judgements of apparent health: Support for a “good genes” explanation of the attractiveness-symmetry relationship. Evol. Hum. Behav. 22, 417–429. https://doi.org/10.1016/S1090-5138(01)00083-6

Kaiser, H.F., 1974. An index of factorial simplicity. Psychometrika 39, 31–36. https://doi.org/10.1007/BF02291575

Kaisler, R.E., Leder, H., 2016. Trusting the Looks of Others: Gaze Effects of Faces in Social Settings. Perception 45, 875–892. https://doi.org/10.1177/0301006616643678

Kalegina, A., Schroeder, G., Allchin, A., Berlin, K., Cakmak, M., 2018. Characterizing the Design Space of Rendered Robot Faces, in: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. ACM, New York, NY, USA, pp. 96–104. https://doi.org/10.1145/3171221.3171286

Kam, C.D., Wilking, J.R., Zechmeister, E.J., 2007. Beyond the “Narrow Data Base”: Another Convenience Sample for Experimental Research. Polit. Behav. 29, 415–440. https://doi.org/10.1007/s11109-007-9037-6

Karbauskaite, R., Sakalauskas, L., Dzemyda, G., 2020. Kriging Predictor for Facial Emotion Recognition Using Numerical Proximities of Human Emotions. Inform. 31, 249–275. https://doi.org/10.15388/20-INFOR419

Kardes, F.R., Posavac, S.S., Cronley, M.L., 2004. Consumer inference: a review of processes, bases, and judgment contexts. J. Consum. Psychol. 14, 230–256. https://doi.org/10.1207/s15327663jcp1403_6

Karwowski, W., 2005. Ergonomics and human factors: the paradigms for science, engineering, design, technology and management of human-compatible systems. Ergonomics 48, 436–463. https://doi.org/10.1080/00140130400029167

Khare, R., Burger, J.D., Aberdeen, J.S., Tresner-Kirsch, D.W., Corrales, T.J., Hirchman, L., Lu, Z., 2015. Scaling drug indication curation through crowdsourcing. Database 2015. https://doi.org/10.1093/database/bav016

Kiesler, S., Goodrich, Michael A, Goodrich, M A, 2018. The Science of Human-Robot Interaction. ACM Trans. Hum.-Robot Interact. 7, 1, Artic. 9. https://doi.org/10.1145/3209701

Kim, K.J., Park, E., Shyam Sundar, S., 2013. Caregiving role in human-robot interaction:

A study of the mediating effects of perceived benefit and social presence. Comput. Human Behav. 29, 1799–1806. https://doi.org/10.1016/j.chb.2013.02.009

Kiss, T., Strunk, J., 2006. Unsupervised multilingual sentence boundary detection. Comput. Linguist. 32, 485–525. https://doi.org/10.1162/coli.2006.32.4.485

Kleisner, K., Priplatova, L., Frost, P., Flegr, J., 2013. Trustworthy-Looking Face Meets Brown Eyes. PLoS One 8, 1–7. https://doi.org/10.1371/journal.pone.0053285

Kocsor, F., Bereczkei, T., 2017. First Impressions of Strangers Rely on Generalization of Behavioral Traits Associated with previously Seen Facial Features. Curr. Psychol. 36, 385–391. https://doi.org/10.1007/s12144-016-9427-1

Kocsor, F., Kozma, L., Neria, A.L., Jones, D.N., Bereczkei, T., 2019. Arbitrary signals of trustworthiness – social judgments may rely on facial expressions even with experimentally manipulated valence. Heliyon 5, e01736. https://doi.org/10.1016/j.heliyon.2019.e01736

Kompatsiari, K., Ciardo, F., Tikhanoff, V., Metta, G., Wykowska, A., 2019. It's in the Eyes: The Engaging Role of Eye Contact in HRI. Int. J. Soc. Robot. 1–11. https://doi.org/10.1007/s12369-019-00565-4

Konijn, E.A., Hoorn, J.F., 2017. Parasocial Interaction and Beyond: Media Personae and Affective Bonding, in: The International Encyclopedia of Media Effects. Wiley, pp. 1–15. https://doi.org/10.1002/9781118783764.wbieme0071

Kramer, R.S.S., 2015. Facial width-to-height ratio in a large sample of commonwealth games athletes. Evol. Psychol. 13, 197–209. https://doi.org/10.1177/147470491501300112

Kruglanski, A.W., Higgins, E.T. (Edward T., 2008. Social psychology: handbook of basic principles, Choice Reviews Online. Guilford Press. https://doi.org/10.5860/choice.45-3496

Krumhuber, E., Manstead, A.S.R., Cosker, D., Marshall, D., Rosin, P.L., Kappas, A., 2007. Facial Dynamics as Indicators of Trustworthiness and Cooperative Behavior. Emotion 7, 730–735. https://doi.org/10.1037/1528-3542.7.4.730

Ku, J., Jang, H.J., Kim, K.U., Kim, J.H., Park, S.H., Lee, J.H., Kim, J.J., Kim, I.Y., Kim, S.I., 2005. Experimental results of affective valence and arousal to avatar's facial expressions. Cyberpsychology Behav. 8, 493–503. https://doi.org/10.1089/cpb.2005.8.493

Kurniawati, A., 2017. The role of perceived trust of physical bank towards perceived risk on internet banking acceptance from the user perspective knowledge sharing in production network view project knowledge management in Indonesian small and medium industries view project I. J. Econ. Bus. Manag. 5, 168–172.

https://doi.org/10.18178/joebm.2017.5.4.506

Ladik, D., Carrillat, F., Tadajewski, M., 2015. Belk's (1988) "possessions and the extended self" revisited. J. Hist. Res. Mark. 7, 184–207. https://doi.org/10.1108/JHRM-06-2014-0018

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R., 2019. Albert: A lite bert for self-supervised learning of language representations. arXiv.

Landwehr, J.R., McGill, A.L., Herrmann, A., 2011. It's Got the Look: The Effect of Friendly and Aggressive "Facial" Expressions on Product Liking and Sales. J. Mark. 75, 132–146. https://doi.org/10.1509/jmkg.75.3.132

Lane, R.D., Chua, P.M.L., Dolan, R.J., 1999. Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. Neuropsychologia 37, 989–997. https://doi.org/10.1016/S0028-3932(99)00017-2

Leder, H., Forster, M., Gerger, G., 2011. The Glasses Stereotype Revisited. Swiss J. Psychol. 70, 211–222. https://doi.org/10.1024/1421-0185/a000059

Lederman, L.C., 1990. Assessing educational effectiveness: The focus group interview as a technique for data collection [1]. Commun. Educ. 39, 117–127. https://doi.org/10.1080/03634529009378794

Lee, J.J., Lee, K.P., 2009. Facilitating dynamics of focus group interviews in East Asia: Evidence and tools by cross-cultural study. Int. J. Des. 3, 17–28.

Lee, Y.-K., 2013. Babyfacedness, Sex of Face Stimulus, and Social Context in Face Perception and Person Evaluation. Psychol. Rep. 112, 800–817. https://doi.org/10.2466/01.17.pr0.112.3.800-817

Linke, L., Saribay, S.A., Kleisner, K., 2016. Perceived trustworthiness is associated with position in a corporate hierarchy. Pers. Individ. Dif. 99, 22–27. https://doi.org/10.1016/j.paid.2016.04.076

Liu, C., Ham, J., Postma, E., Midden, C., Joosten, B., Goudbeek, M., 2013. Representing Affective Facial Expressions for Robots and Embodied Conversational Agents by Facial Landmarks. Int. J. Soc. Robot. 5, 619–626. https://doi.org/10.1007/s12369-013-0208-9

Liu, Q., Yang, X., Lin, L., Liu, M., Lin, H., Liu, F., Xie, Y., Lam, D.S.C., 2019. Review on centration, astigmatic axis alignment, pupil size and optical zone in SMILE. Asia-Pacific J. Ophthalmol. https://doi.org/10.1097/01.APO.0000580144.22353.46

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach.

Loken, B., Ward, J., 1990. Alternative Approaches to Understanding the Determinants of Typicality. J. Consum. Res. 17, 111. https://doi.org/10.1086/208542

Loo, R., 2002. A caveat on using single-item versus multiple-item scales. J. Manag. Psychol. 17, 268–3946. https://doi.org/10.1108/02683940210415933

Lorenz, K., 1943. Die angeborenen Formen möglicher Erfahrung. Z. Tierpsychol. 5, 235–409. https://doi.org/10.1111/j.1439-0310.1943.tb00655.x

Lougen, C., 2009. Sources: The Sage Encyclopedia of Qualitative Research Methods, Reference & User Services Quarterly. Sage Publications. https://doi.org/10.5860/rusq.49n1.101

Lovett, Matt, Bajaba, S., Lovett, Myra, Simmering, M.J., 2018. Data Quality from Crowdsourced Surveys: A Mixed Method Inquiry into Perceptions of Amazon's Mechanical Turk Masters, in: Applied Psychology. Blackwell Publishing Ltd, pp. 339–366. https://doi.org/10.1111/apps.12124

Ludolph, R., Schulz, P.J., 2015. Does regulatory fit lead to more effective health communication? Asystematic review. Soc. Sci. Med. 128, 142–150. https://doi.org/10.1016/j.socscimed.2015.01.021

Luo, J.T., McGoldrick, P., Beatty, S., Keeling, K.A., 2006. On-screen characters: Their design and influence on consumer trust. J. Serv. Mark. 20, 112–124. https://doi.org/10.1108/08876040610657048

Luria, M., Forlizzi, J., Hodgins, J., 2018. The Effects of Eye Design on the Perception of Social Robots. RO-MAN 2018 - 27th IEEE Int. Symp. Robot Hum. Interact. Commun. 1032–1037. https://doi.org/10.1109/ROMAN.2018.8525767

Lutz, C., Newlands, G., 2018. Consumer segmentation within the sharing economy: The case of Airbnb. J. Bus. Res. 88, 187–196. https://doi.org/10.1016/j.jbusres.2018.03.019

Lyons, J.B., Vo, T., Wynne, K.T., Mahoney, S., Nam, C.S., Gallimore, D., 2020. Trusting Autonomous Security Robots: The Role of Reliability and Stated Social Intent. Hum. Factors 001872082090162. https://doi.org/10.1177/0018720820901629

Ma, F., Xu, F., Luo, X., 2015. Children's and Adults' Judgments of Facial Trustworthiness: The Relationship to Facial Attractiveness. Percept. Mot. Skills 121, 179–198. https://doi.org/10.2466/27.22.pms.121c10x1

Maeng, A., Aggarwal, P., 2018. Facing dominance: Anthropomorphism and the effect of product face ratio on consumer preference. J. Consum. Res. 44, 1104–1122. https://doi.org/10.1093/jcr/ucx090

Maoz, I., 2012. The Face of the Enemy: The Effect of Press-Reported Visual Information Regarding the Facial Features of Opponent Politicians on Support for Peace. Polit.

Commun. 29, 243–256. https://doi.org/10.1080/10584609.2012.694987

Marriam-Webster, 2002. Dictionary by Merriam-Webster: America's most-trusted online dictionary [WWW Document]. Merriam Webster. URL https://www.merriam-webster.com/ (accessed 1.7.21).

Marutho, D., Hendra Handaka, S., Wijaya, E., Muljono, 2018. The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News, in: Proceedings - 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life, ISemantic 2018. Institute of Electrical and Electronics Engineers Inc., pp. 533–538. https://doi.org/10.1109/ISEMANTIC.2018.8549751

Masip, J., Garrido, E., Herrero, C., 2004. Facial appearance and impressions of 'credibility': The effects of facial babyishness and age on person perception. Int. J. Psychol. 39, 276–289. https://doi.org/10.1080/00207590444000014

Mathur, M.B., Reichling, D.B., 2016. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. Cognition 146, 22–32. https://doi.org/10.1016/j.cognition.2015.09.008

Mayer, R.C., Davis, J.H., Schoorman, F.D., 1995. An Integrative Model of Organizational Trust. Acad. Manag. Rev. 20, 709–734. https://doi.org/10.5465/amr.1995.9508080335

McGinn, C., 2019. Why Do Robots Need a Head? The Role of Social Interfaces on Service Robots. Int. J. Soc. Robot. 1–15. https://doi.org/10.1007/s12369-019-00564-5

Meeden, L.A., Blank, D.S., 2006. Introduction to developmental robotics. Conn. Sci. https://doi.org/10.1080/09540090600806631

Menne, I.M., Schwab, F., 2018. Faces of Emotion: Investigating Emotional Facial Expressions Towards a Robot. Int. J. Soc. Robot. 10, 199–209. https://doi.org/10.1007/s12369-017-0447-2

Miesler, L., 2011. Imitating Human Forms in Product Design: How Does Anthropomorphism Work, When Does It Work, and What Does It Affect.

Miesler, L., Leder, H., Herrmann, A., 2011. Isn't it cute: An evolutionary perspective of baby-schema effects in visual product designs. Int. J. Des. 5, 17–30.

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient estimation of word representations in vector space, in: 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings. International Conference on Learning Representations, ICLR.

Mollahosseini, A., Abdollahi, H., Sweeny, T.D., Cole, R., Mahoor, M.H., 2018. Role of embodiment and presence in human perception of robots' facial cues. Int. J. Hum.

Comput. Stud. 116, 25–39. https://doi.org/10.1016/J.IJHCS.2018.04.005

Montepare, J.M., Zebrowitz, L.A., 1998. Person Perception Comes of Age: The Salience and Significance of Age in Social Judgments. Adv. Exp. Soc. Psychol. 30, 93–161. https://doi.org/10.1016/S0065-2601(08)60383-4

Mortensen, K., Hughes, T.L., 2018. Comparing Amazon's Mechanical Turk Platform to Conventional Data Collection Methods in the Health and Medical Research Literature. J. Gen. Intern. Med. 33, 533–538. https://doi.org/10.1007/s11606-017-4246-0

Munuera-Aleman, J.L., Delgado-Ballester, E., Yague-Guillen, M.J., 2003. Development and Validation of a Brand Trust Scale. Int. J. Mark. Res. 45, 1–18. https://doi.org/10.1177/147078530304500103

Nabi, R.L., Walter, N., Oshidary, N., Endacott, C.G., Love-Nichols, J., Lew, Z.J., Aune, A., 2020. Can Emotions Capture the Elusive Gain-Loss Framing Effect? A Meta-Analysis. Communic. Res. 47, 1107–1130. https://doi.org/10.1177/0093650219861256

Neuendorf, K.A., 2020. The Content Analysis Guidebook, The Content Analysis Guidebook. https://doi.org/10.4135/9781071802878

Ngan, H.F.B., Yu, C.E., 2019. To smile or not to smile–an eye-tracking study on service recovery. Curr. Issues Tour. 22, 2327–2332. https://doi.org/10.1080/13683500.2018.1502260

Ngiam, J., Khosla, A., Kim, M., 2011. Multimodal deep learning. Mach. Learn. 1–9.

Oh, Y.H., Ju, D.Y., 2020. Age-related differences in fixation pattern on a companion robot. Sensors (Switzerland) 20, 1–21. https://doi.org/10.3390/s20133807

Okubo, M., Ishikawa, K., Kobayashi, A., 2018. The cheek of a cheater: Effects of posing the left and right hemiface on the perception of trustworthiness. Laterality 23, 209–227. https://doi.org/10.1080/1357650X.2017.1351449

Okubo, M., Ishikawa, K., Kobayashi, A., 2013. No trust on the left side: Hemifacial asymmetries for trustworthiness and emotional expressions. Brain Cogn. 82, 181–186. https://doi.org/10.1016/j.bandc.2013.04.004

Oosterhof, N.N., Todorov, A., 2009. Shared Perceptual Basis of Emotional Expressions and Trustworthiness Impressions From Faces. Emotion 9, 128–133. https://doi.org/10.1037/a0014520

Oosterhof, N.N., Todorov, A., 2008. The functional basis of face evaluation, Proceedings of the National Academy of Sciences of the United States of America. PNAS August. https://doi.org/10.1073/pnas.0805664105

Palinko, O., Rea, F., Sandini, G., Sciutti, A., 2015. Eye gaze tracking for a humanoid robot, in: IEEE-RAS International Conference on Humanoid Robots. pp. 318–324. https://doi.org/10.1109/HUMANOIDS.2015.7363561

Park, S., Mowen, J.C., 2007. Perceived trustworthiness of online shops. J. Consum. Behav. 6, 35–50.

Pavlov, S. V., Korenyok, V. V., Reva, N. V., Tumyalis, A. V., Loktev, K. V., Aftanas, L.I., 2015. Effects of long-term meditation practice on attentional biases towards emotional faces: An eye-tracking study. Cogn. Emot. 29, 807–815. https://doi.org/10.1080/02699931.2014.945903

Pennebaker, J.W., Chew, C.H., 1985. Behavioral Inhibition and Electrodermal Activity During Deception, Journal of Personality and Social Psychology. https://doi.org/10.1037/0022-3514.49.5.1427

Penton-Voak, I.S., Jones, B.C., Little, A.C., Baker, S., Tiddeman, B., Burt, D.M., Perrett, D.I., 2001. Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. Proc. R. Soc. B Biol. Sci. 268, 1617–1623. https://doi.org/10.1098/rspb.2001.1703

Pollmann, K., Tagalidou, N., Fronemann, N., 2019. It's in your eyes: Which facial design is best suited to let a robot express emotions? ACM Int. Conf. Proceeding Ser. 639–642. https://doi.org/10.1145/3340764.3344883

Prakash, A., Rogers, W.A., 2015. Why Some Humanoid Faces Are Perceived More Positively Than Others: Effects of Human-Likeness and Task. Int. J. Soc. Robot. 7, 309–331. https://doi.org/10.1007/s12369-014-0269-4

Ratislavová, K., Ratislav, J., 2014. Asynchronous email interview as a qualitative research method in the humanities. Hum. Aff. 24, 452–460. https://doi.org/10.2478/s13374-014-0240-y

Re, D.E., Rule, N.O., 2016. The big man has a big mouth: Mouth width correlates with perceived leadership ability and actual leadership performance. J. Exp. Soc. Psychol. 63, 86–93. https://doi.org/10.1016/j.jesp.2015.12.005

Reed, L.I., DeScioli, P., 2017. Watch out! How a fearful face adds credibility to warnings of danger. Evol. Hum. Behav. 38, 490–495. https://doi.org/10.1016/j.evolhumbehav.2017.03.003

Reeves, B., Hancock, J., Liu, X., 2020. Social robots are like real people: First impressions, attributes, and stereotyping of social robots. Technol. Mind, Behav. 1. https://doi.org/10.1037/tmb0000018

Riegelsberger, J., Sasse, M.A., McCarthy, J.D., 2004. Trust at First Sight? A Test of Users' Ability to Identify Trustworthy E-commerce Sites. People Comput. XVII — Des.

Soc. 243–259. https://doi.org/10.1007/978-1-4471-3754-2_15

Rietjens, S., 2015. Qualitative Data Analysis, in: Routledge Handbook of Research Methods in Military Studies. https://doi.org/10.4324/9780203093801.ch12

Roberts, S.C., Little, A.C., DeBruine, L.M., Petrie, M., 2017. Discrimination of Attractiveness and Health in Men's Faces: the Impact of Color Cues and Variation in Relation to Sex and Age of Rater. Adapt. Hum. Behav. Physiol. 3, 401–411. https://doi.org/10.1007/s40750-017-0081-0

Rosenthal-von der Pütten, A.M., Krämer, N.C., Maderwald, S., Brand, M., Grabenhorst, F., 2019. Neural Mechanisms for Accepting and Rejecting Artificial Social Partners in the Uncanny Valley. J. Neurosci. 39, 6555–6570. https://doi.org/10.1523/JNEUROSCI.2956-18.2019

Salem, M., Lakatos, G., Amirabdollahian, F., Dautenhahn, K., 2015. Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust, in: ACM/IEEE International Conference on Human-Robot Interaction. pp. 141–148. https://doi.org/10.1145/2696454.2696497

Salganik, M.J., Dodds, P.S., Watts, D.J., 2006. Experimental study of inequality and unpredictability in an artificial cultural market. Science (80-.). 311, 854–856. https://doi.org/10.1126/science.1121066

Samal, A., Iyengar, P.A., 1992. Automatic recognition and analysis of human faces and facial expressions: a survey. Pattern Recognit. 25, 65–77. https://doi.org/10.1016/0031-3203(92)90007-6

Sanchez, A., Vazquez, C., 2014. Looking at the eyes of happiness: Positive emotions mediate the influence of life satisfaction on attention to happy faces. J. Posit. Psychol. 9, 435–448. https://doi.org/10.1080/17439760.2014.910827

Sanchez, A., Vazquez, C., Gomez, D., Joormann, J., 2014. Gaze-fixation to happy faces predicts mood repair after a negative mood induction. Emotion 14, 85–94. https://doi.org/10.1037/a0034500

Sandry, E., 2015. Robots and communication, Robots and Communication. https://doi.org/10.1057/9781137468376

Santos, I.M., Young, A.W., 2011. Inferring social attributes from different face regions: Evidence for holistic processing. Q. J. Exp. Psychol. 64, 751–766. https://doi.org/10.1080/17470218.2010.519779

Santos, S., Almeida, I., Oliveiros, B., Castelo-Branco, M., 2016. The role of the amygdala in facial trustworthiness processing: a systematic review and meta-analyses of fMRI studies. PLoS One 11, e0167276. https://doi.org/10.1371/journal.pone.0167276

Sassenrath, C., Sassenberg, K., Ray, D.G., Scheiter, K., Jarodzka, H., 2014. A Motivational Determinant of Facial Emotion Recognition: Regulatory Focus Affects Recognition of Emotions in Faces. PLoS One 9, e112383. https://doi.org/10.1371/journal.pone.0112383

Saunderson, S., Nejat, G., 2019. How Robots Influence Humans: A Survey of Nonverbal Communication in Social Human–Robot Interaction. Int. J. Soc. Robot. 1–34. https://doi.org/10.1007/s12369-019-00523-0

Schaefer, K.E., 2016. Measuring trust in human robot interactions: Development of the "trust perception scale-HRI," in: Robust Intelligence and Trust in Autonomous Systems. Springer US, pp. 191–218. https://doi.org/10.1007/978-1-4899-7668-0_10

Sevilla, J., Kahn, B.E., 2014. The completeness heuristic: Product shape completeness influences size perceptions, preference, and consumption. J. Mark. Res. 51, 57–68. https://doi.org/10.1509/jmr.12.0153

Shank, D.B., 2016. Using Crowdsourcing Websites for Sociological Research: The Case of Amazon Mechanical Turk. Am. Sociol. 47, 47–55. https://doi.org/10.1007/s12108-015-9266-9

Sharma, Mahima, Kacker, S., Sharma, Mohit, 2016. A Brief Introduction and Review on Galvanic Skin Response. Int. J. Med. Res. Prof. 2. https://doi.org/10.21276/ijmrp.2016.2.6.003

Sim, D.Y.Y., Loo, C.K., 2015. Extensive assessment and evaluation methodologies on assistive social robots for modelling human-robot interaction - A review. Inf. Sci. (Ny). https://doi.org/10.1016/j.ins.2014.12.017

Slessor, G., Miles, L.K., Bull, R., Phillips, L.H., 2010. Age-Related Changes in Detecting Happiness: Discriminating Between Enjoyment and Nonenjoyment Smiles. Psychol. Aging 25, 246–250. https://doi.org/10.1037/a0018248

Sofer, C., Dotsch, R., Oikawa, M., Oikawa, H., Wigboldus, D.H.J., Todorov, A., 2017. For Your Local Eyes Only: Culture-Specific Face Typicality Influences Perceptions of Trustworthiness. Perception 46, 914–928. https://doi.org/10.1177/0301006617691786

Sofer, C., Dotsch, R., Wigboldus, D.H.J., Todorov, A., 2015. What Is Typical Is Good. Psychol. Sci. 26, 39–47. https://doi.org/10.1177/0956797614554955

Song, Y., 2020. Building a 'Deeper' Trust: Mapping the Facial Anthropomorphic Trustworthiness in Social Robot Design through Multidisciplinary Approaches. Des. J. 23, 1–11. https://doi.org/10.1080/14606925.2020.1766871

Song, Y., Luximon, Y., 2021. The face of trust: The effect of robot face ratio on consumer preference. Comput. Human Behav. 116, 106620. https://doi.org/10.1016/j.chb.2020.106620

Song, Y., Luximon, Y., 2020a. Trust in AI Agent: A Systematic Review of Facial Anthropomorphic Trustworthiness for Social Robot Design. Sensors 20, 5087. https://doi.org/10.3390/s20185087

Song, Y., Luximon, Y., 2020b. Hotspots and Trends in Creativity and Design Research: A Knowledge Graph Analysis. Springer, Cham, pp. 128–135. https://doi.org/10.1007/978-3-030-51626-0_15

Song, Y., Luximon, Y., 2019. Design for sustainability: The effect of lettering case on environmental concern from a green advertising perspective. Sustain. 11, 1333. https://doi.org/10.3390/su11051333

Song, Y., Luximon, Y., Luo, J., 2020. A moderated mediation analysis of the effect of lettering case and color temperature on trustworthiness perceptions and investment decisions. Int. J. Bank Mark. 38, 987–1005. https://doi.org/10.1108/IJBM-09-2019-0315

Spotts, M.F., Shoup, T.E., Hornberger, L.E., Kazmer, D.O., 2004. Design of Machine Elements, Eighth Edition. J. Mech. Des. 126, 201–201. https://doi.org/10.1115/1.1637657

Sproull, L., Subramani, M., Kiesler, S., Walker, J., Waters, K., 1996. When the Interface Is a Face. Human-Computer Interact. 11, 97–124. https://doi.org/10.1207/s15327051hci1102_1

Stanley, D.A., Sokol-Hessner, P., Banaji, M.R., Phelps, E.A., 2011. Implicit race attitudes predict trustworthiness judgments and economic trust decisions. Proc. Natl. Acad. Sci. U. S. A. 108, 7710–7715. https://doi.org/10.1073/pnas.1014345108

Stanton, C.J., Stevens, C.J., 2017. Don’t Stare at Me: The Impact of a Humanoid Robot’s Gaze upon Trust During a Cooperative Human–Robot Visual Task. Int. J. Soc. Robot. 9, 745–753. https://doi.org/10.1007/s12369-017-0422-y

Stanton, N.A., Stewart, R., Harris, D., Houghton, R.J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M.S., Linsell, M., Dymott, R., Green, D., 2006. Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. Ergonomics 49, 1288–1311. https://doi.org/10.1080/00140130600612762

Steinel, W., Van Kleef, G.A., Harinck, F., 2008. Are you talking to me?! Separating the people from the problem when expressing emotions in negotiation. J. Exp. Soc. Psychol. 44, 362–369. https://doi.org/10.1016/j.jesp.2006.12.002

Stellar, J.E., Cohen, A., Oveis, C., Keltner, D., 2015. Affective and physiological responses to the suffering of others: Compassion and vagal activity. J. Pers. Soc. Psychol. 108, 572–585. https://doi.org/10.1037/pspi0000010

Stirrat, M., Perrett, D.I., 2010. Valid Facial Cues to Cooperation and Trust. Psychol. Sci.

21, 349–354. https://doi.org/10.1177/0956797610362647

Stone, A.A., Wen, C.K.F., Schneider, S., Junghaenel, D.U., 2020. Evaluating the Effect of Daily Diary Instructional Phrases on Respondents' Recall Time Frames: Survey Experiment. J. Med. Internet Res. 22, e16105. https://doi.org/10.2196/16105

Stroessner, S.J., Benitez, J., 2019. The Social Perception of Humanoid and Non-Humanoid Robots: Effects of Gendered and Machinelike Features. Int. J. Soc. Robot. 11, 305–315. https://doi.org/10.1007/s12369-018-0502-7

Stuck, R.E., Rogers, W.A., 2018. Older Adults' Perceptions of Supporting Factors of Trust in a Robot Care Provider. J. Robot. 2018. https://doi.org/10.1155/2018/6519713

Syakur, M.A., Khotimah, B.K., Rochman, E.M.S., Satoto, B.D., 2018. Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster, in: IOP Conference Series: Materials Science and Engineering. https://doi.org/10.1088/1757-899X/336/1/012017

Tay, B., Jung, Y., Park, T., 2014. When stereotypes meet robots: The double-edge sword of robot gender and personality in human-robot interaction. Comput. Human Behav. 38, 75–84. https://doi.org/10.1016/j.chb.2014.05.014

Thatcher, J.B., Harrison, D., White Baker, E., Arsal, R.E., Roberts, N.H., 2011. The Role of Trust in Postadoption IT Exploration: An Empirical Examination of Knowledge Management Systems. IEEE Trans. Eng. Manag. 58. https://doi.org/10.1109/TEM.2009.2028320

Thompson, B., 1995. Stepwise Regression and Stepwise Discriminant Analysis Need Not Apply here: A Guidelines Editorial. Educ. Psychol. Meas. https://doi.org/10.1177/0013164495055004001

Timoshenko, A., Hauser, J.R., 2019. Identifying customer needs from user-generated content. Mark. Sci. 38, 1–20. https://doi.org/10.1287/mksc.2018.1123

Todorov, A., Baron, S.G., Oosterhof, N.N., 2008a. Evaluating face trustworthiness: A model based approach. Soc. Cogn. Affect. Neurosci. 3, 119–127. https://doi.org/10.1093/scan/nsn009

Todorov, A., Mandisodza, A.N., Goren, A., Hall, C.C., 2005. Psychology: Inferences of competence from faces predict election outcomes. Science (80-.). 308, 1623–1626. https://doi.org/10.1126/science.1110589

Todorov, A., Olivola, C.Y., Dotsch, R., Mende-Siedlecki, P., 2015. Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. Annu. Rev. Psychol. 66, 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., Said, C.P., Engell, A.D., Oosterhof, N.N., 2008b. Understanding evaluation of faces on social dimensions. Trends Cogn. Sci. 12, 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Toet, A., Bijlsma, M., Brouwer, A.M., 2017. Stress Response and Facial Trustworthiness Judgments in Civilians and Military. SAGE Open 7, 1–11. https://doi.org/10.1177/2158244017725386

Tomash, J.J., Reed, P., 2015. Using conditioning to elicit skin conductance responses to deception. Learn. Motiv. 49, 31–37. https://doi.org/10.1016/j.lmot.2015.02.002

Torres-Carrión, P. V., González-González, C.S., Toledo-Delgado, P.A., Muñoz-Cruz, V., Gil-Iranzo, R., Reyes-Alonso, N., Hernández-Morales, S., 2019. Improving cognitive visual-motor abilities in individuals with down syndrome. Sensors (Switzerland) 19. https://doi.org/10.3390/s19183984

Torresen, J., 2018. A Review of Future and Ethical Perspectives of Robotics and AI. Front. Robot. AI 4, 75. https://doi.org/10.3389/frobt.2017.00075

Tranel, D., Damasio, A., 1985. Knowledge without awareness: an autonomic index of facial recognition by prosopagnosics. Science (80-.). 228, 1453–1454. https://doi.org/10.1126/SCIENCE.4012303

Turner, A., 2015. Generation Z: Technology and Social Interest. J. Individ. Psychol. 71, 103–113. https://doi.org/10.1353/jip.2015.0021

Ulrich, K., 1992. Product design and development. Biosens. Bioelectron. 7, 85–89. https://doi.org/10.1016/0956-5663(92)90013-D

van 't Wout, M., Sanfey, A.G., 2008. Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. Cognition 108, 796–803. https://doi.org/10.1016/j.cognition.2008.07.002

Van Rompay, T.J.L., Pruyn, A.T.H., 2011. When visual product features speak the same language: Effects of shape-typeface congruence on brand perception and price expectations. J. Prod. Innov. Manag. 28, 599–610. https://doi.org/10.1111/j.1540-5885.2011.00828.x

van Rompay, T.J.L., Pruyn, A.T.H., Tieke, P., 2009. Symbolic meaning integration in design and its influence on product and brand evaluation. Int. J. Des. 3, 19–26.

Vanderborght, B., Simut, R., Saldien, J., Pop, C., Rusu, A.S., Pintea, S., Lefeber, D., David, D.O., 2012. Using the social robot probo as a social story telling agent for children with ASD. Interact. Stud. Stud. Soc. Behav. Commun. Biol. Artif. Syst. 13, 348–372. https://doi.org/10.1075/is.13.3.02van

Venturoso, L., Gabrieli, G., Truzzi, A., Azhari, A., Setoh, P., Bornstein, M.H., Esposito, G., 2019. Effects of Baby Schema and Mere Exposure on Explicit and Implicit Face Processing. Front. Psychol. 10, 2649. https://doi.org/10.3389/fpsyg.2019.02649

Verberne, F.M.F., Ham, J., Midden, C.J.H., 2015. Trusting a Virtual Driver That Looks, Acts, and Thinks Like You. Hum. Factors 57, 895–909. https://doi.org/10.1177/0018720815580749

Verhelst, M., Moons, B., 2017. Embedded Deep Neural Network Processing: Algorithmic and Processor Techniques Bring Deep Learning to IoT and Edge Devices. IEEE Solid-State Circuits Mag. 9, 55–65. https://doi.org/10.1109/MSSC.2017.2745818

Vesker, M., Bahn, D., Degé, F., Kauschke, C., Schwarzer, G., 2018. Perceiving arousal and valence in facial expressions: Differences between children and adults. Eur. J. Dev. Psychol. 15, 411–425. https://doi.org/10.1080/17405629.2017.1287073

Vrticka, P., Lordier, L., Bediou, B., Sander, D., 2012. Human Amygdala Response to Dynamic Facial Expressions of Positive and Negative Surprise. Cunningham & Brosch. https://doi.org/10.1037/a0034619.supp

Walker, M., Jiang, F., Vetter, T., Sczesny, S., 2011. Universals and cultural differences in forming personality trait judgments from faces. Soc. Psychol. Personal. Sci. 2, 609–617. https://doi.org/10.1177/1948550611402519

Walters, M.L., Syrdal, D.S., Dautenhahn, K., te Boekhorst, R., Koay, K.L., 2008. Avoiding the uncanny valley: Robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. Auton. Robots 24, 159–178. https://doi.org/10.1007/s10514-007-9058-3

Warrington, E.K., James, M., 1967. An Experimental Investigation of Facial Recognition in Patients with Unilateral Cerebral Lesions. Cortex 3, 317–326. https://doi.org/10.1016/S0010-9452(67)80020-0

Weiβ, M., Mussel, P., Hewig, J., 2019. The value of a real face: Differences between affective faces and emojis in neural processing and their social influence on decision-making. Soc. Neurosci. https://doi.org/10.1080/17470919.2019.1675758

Welker, K.M., Bird, B.M., Arnocky, S., 2016. Commentary: Facial Width-to-Height Ratio (fWHR) Is Not Associated with Adolescent Testosterone Levels. Front. Psychol. 7, 1745. https://doi.org/10.3389/fpsyg.2016.01745

Westerman, S.J., Gardner, P.H., Sutherland, E.J., White, T., Jordan, K., Watts, D., Wells, S., 2012. Product design: Preference for rounded versus angular design elements. Psychol. Mark. 29, 595–605. https://doi.org/10.1002/mar.20546

Westlund, J.K., Lee, J.J., Plummer, L., Faridi, F., Gray, J., Berlin, M., Quintus-Bosz, H., Hartmann, R., Hess, M., Dyer, S., dos Santos, K., Adalgeirsson, S.O., Gordon, G., Spaulding, S., Martinez, M., Das, M., Archie, M., Jeong, S., Breazeal, C., 2016. Tega:

A social robot, in: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, pp. 561–561. https://doi.org/10.1109/HRI.2016.7451856

Wheless, L.R., Grotz, J., 1977. the Measurement of Trust and Its Relationship To Self-Disclosure. Hum. Commun. Res. 3, 250–257. https://doi.org/10.1111/j.1468-2958.1977.tb00523.x

Willis, J., Todorov, A., 2006. First Impressions. Psychol. Sci. 17, 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Willis, M.S., Esqueda, C.W., Schacht, R.N., 2008. Social Perceptions of Individuals Missing Upper Front Teeth. Percept. Mot. Skills 106, 423–435. https://doi.org/10.2466/pms.106.2.423-435

Windhager, S., Hutzler, F., Carbon, C.-C., Oberzaucher, E., Schaefer, K., Thorstensen, T., Leder, H., Grammer, K., 2010. Laying eyes on headlights: eye movements suggest facial features in cars. Coll. Antropol. 34, 1075–80.

Winfield, A., 2019. Ethical standards in robotics and AI, Nature Electronics. https://doi.org/10.1038/s41928-019-0213-6

Woodall, W.G., Burgoon, J.K., Markel, N.N., 1980. The effects of facial-head cue combinations on interpersonal evaluations. Commun. Q. 28, 47–55. https://doi.org/10.1080/01463378009369374

Xie, Y., DeVellis, R.F., 1992. Scale Development: Theory and Applications., 16th ed, Contemporary Sociology. Sage Publications. https://doi.org/10.2307/2075704

Xiong, C., Hua, Z., Lv, K., Li, X., 2017. An improved K-means text clustering algorithm by optimizing initial cluster centers, in: Proceedings - 2016 7th International Conference on Cloud Computing and Big Data, CCBD 2016. pp. 265–268. https://doi.org/10.1109/CCBD.2016.059

Xu, F., Wu, D., Toriyama, R., Ma, F., Itakura, S., Lee, K., 2012. Similarities and differences in chinese and caucasian adults' use of facial cues for trustworthiness judgments. PLoS One 7, 1–9. https://doi.org/10.1371/journal.pone.0034859

Xu, K., 2019. First encounter with robot Alpha: How individual differences interact with vocal and kinetic cues in users' social responses. New Media Soc. 21, 2522–2547. https://doi.org/10.1177/1461444819851479

Yagoda, R.E., Gillan, D.J., 2012. You Want Me to Trust a ROBOT? The Development of a Human-Robot Interaction Trust Scale. Int. J. Soc. Robot. 4, 235–248. https://doi.org/10.1007/s12369-012-0144-0

Yan, Y., Nie, J., Huang, L., Li, Z., Cao, Q., Wei, Z., 2015. Is Your First Impression Reliable? Trustworthy Analysis Using Facial Traits in Portraits. Multimed. Model.

148–158.

Yovel, G., Duchaine, B., 2006. Specialized face perception mechanisms extract both part and spacing information: Evidence from developmental prosopagnosia. J. Cogn. Neurosci. 18, 580–593. https://doi.org/10.1162/jocn.2006.18.4.580

Yu, P.L., Balaji, M.S., Khong, K.W., 2015. Building trust in internet banking: a trustworthiness perspective. Ind. Manag. Data Syst. 115, 235–252. https://doi.org/10.1108/IMDS-09-2014-0262

Zebrowitz, L.A., Montepare, J.M., 2008. Social Psychological Face Perception: Why Appearance Matters. Soc. Personal. Psychol. Compass 2, 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x

Zebrowitz, L.A., Montepare, J.M., Strom, M.A., 2013. 10 Face and body physiognomy: nonverbal cues for trait impressions. Nonverbal Commun. https://doi.org/10.1515/9783110238150.263

Zebrowitz, L.A., Voinescu, L., Collins, M.A., 1996. "Wide-eyed" and "crooked-faced": Determinants of perceived and real honesty across the life span. Personal. Soc. Psychol. Bull. 22, 1258–1269. https://doi.org/10.1177/01461672962212006

Zhang, J., Luximon, Y., Song, Y., 2019. The role of consumers' perceived security, perceived control, interface design features, and conscientiousness in continuous use of mobile payment services. Sustain. 11. https://doi.org/10.3390/su11236843

Zhang, Y., Song, W., Tan, Z., Zhu, H., Wang, Y., Lam, C.M., Weng, Y., Hoi, S.P., Lu, H., Man Chan, B.S., Chen, J., Yi, L., 2019. Could social robots facilitate children with autism spectrum disorders in learning distrust and deception? Comput. Human Behav. 98, 140–149. https://doi.org/10.1016/j.chb.2019.04.008

Zhu, M., Yang, Y., Hsee, C.K., 2018. The Mere Urgency Effect. J. Consum. Res. 45, 673–690. https://doi.org/10.1093/jcr/ucy008