UMEÅ UNIVERSITY

Don’t be unfair, Mr Bot!

An empirical study exploring the perception of fairness in non-work settings for human-agent interactions

August Bäckström & William Ekenberg

Department of informatics
Master thesis, 30 Human-Computer Interaction, and User Experience
SPM 2023.05

Abstract

This study aimed to explore the implementation of fairness in intelligent agents to enhance their interactions in our social space. Two distinct investigations, an experiment, and a focus group, were conducted to examine the impact of unfair treatment by non-anthropomorphic and anthropomorphic agents, where we sought to answer the research question: How does experiencing unfair treatment from agents with different appearances influence individuals' perceptions, satisfaction, and trust? The experiment encompassed four experimental conditions combining fair and unfair behaviours with agents displaying human-like or non-human-like appearances. User enactment, Experience prototyping, and the Wizard of Oz technique were employed during the experiment. The focus group aimed to delve into the concept of fairness and its relevance to agents in greater detail. In summary, the study's findings indicate that fairness is a significantly important consideration in agent design. However, the complexity of designing a fair agent proves challenging, due to the subjective and contextual nature where it entangles with various factors.

Keywords: Fairness; Anthropomorphism; Experiment study; Social interaction; Agents; Intelligent agents; User enactment; Wizard of Oz.; Experience prototyping; Focus group

1. Introduction

The addition of social actors, such as intelligent agents, is entangling the world humans live in to such an extent that reality and the digital continuum start to be inseparable from each other. Humans are moving (if not already), toward a world where they live in co-dependent relationships with these technological advancements (Nass, Steuer, & Tauber, 1994; Frauenberger, 2020). With quick technical advancements in fields such as Artificial Intelligence (AI), Human-robot-interaction (HRI), and the development of Intelligent Agents (agents, robots, voice assistants, chatbots, smart speakers etc.), agents are entering different spaces of human life. They can be found in work settings, homes, schools, vehicles, etc., and wherever they seem to fit a purpose and can offload humans in different tasks. In today's market, they can be found in various forms, with millions of people buying and adopting them, with the most popular ones being voice assistants such as Siri, Alexa, and Google Home to name a few (Purington, Taft, Sannon, Bazarova & Taylor, 2017).

With entrance into this dimension of technology, we are now experiencing an explosion in chatbots and robots. Companies such as Microsoft investing heavily in artificial intelligence (AI), and services such as ChatGPT and more human-like robots. Where the goal is to be able to replace humans for mundane tasks such as customer support, coding, and journalism and replace humans for repetitive tasks in general (Kompatsiaris, Cave, Satsiou, Carle, Passani, Kontopoulos, Diplaris, & McMillan, 2017). The chatbots can take advantage of the latest advancements in natural language processing and AI to answer questions, find information,

write curriculum vitae (CV) when applying for jobs, or produce code and solve mathematical problems. Without the need for human intervention (Kompatsiaris et al.,2017).

This raises the question of anthropomorphism and how appearance affects the experience of the behaviours of agents, but also how intelligent agents should be designed when navigating our social world. Studies show that we often prefer robots with a welcoming and friendly look, but at the same time, studies on the "uncanny valley" indicate that there still must be a clear difference between robots and humans in appearance. (Mori, MacDorman & Kageki, 2012). Studies also indicate that appearance has a big impact on how we perceive behaviours from agents, and that the appearance of agents is something that cannot be ignored in the field of HRI. (Korn, Akalin & Gouveia, 2021: Roesler, Manzey & Onnasch 2022; Siegel, Breazeal, & Norton, 2009; Waytz, Cacioppo & Epley, 2010)

Furthermore, the behaviour of intelligent agents and how these behaviours are perceived by people that use them is important. The research attempts to find answers to what happens when agents find themselves in interactions with multiple users; and how they will navigate complex social situations. This presents an issue where the social competencies of agents remain a largely uncharted topic. This area includes how agents will handle situations where multiple users suddenly engage with them and what social rules they should follow (Luria, Reig, Tan, Steinfeld, Forlizzi & Zimmerman, 2019; Reig, Luria, Wang, Oltman, Carter, Steinfeld, Forlizzi, & Zimmerman, 2020; Reicherts, Zargham, Bonfert, Rogers & Malaka, 2021). Should they be able to keep secrets or lie if asked to? Should they be able to take a stance in an argument and try to resolve conflicts that happen around them? Or should your agent follow you around wherever you go and embody itself in different ways? lastly, do agents need to treat everyone equally, and should they always be fair actors in a social context, if so, how would that be achieved? (Claure, Chang, Kim, Omeiza, Brandao, Lee & Jung, 2022; Peiró, Martínez-Tur & Moliner, 2014).

This study will focus on the social competencies of intelligent agents and how they should adapt to fairness in a multi-user social context. The reason for including multiple users is due to fairness often being related to how we are treated amongst other people, and it is easily noticed when we ourselves are exposed to it or are witnessing others being treated unfairly. Thoughts and issues around fairness commonly arise when we are comparing how we are treated with other people around us.

The study aims to explore if fairness is something that users prioritize, notice, and talk about concerning the social competencies of intelligent agents. With the use of methods such as Wizard of Oz, User enactments and Experience prototyping the aim is to first study the effects of agents treating participants unfairly when they have anthropomorphic qualities. Secondly, the intention is to study if participants even think about fairness with intelligent agents at all. Specifically, the study focuses on different agents that act as quizmasters and how participants respond when agents are unfair towards them. A follow-up focus group is used to allow participants to reflect on their experience with the experiment and let them discuss and reflect on the concept of fairness with intelligent agents. In conclusion, this study aims to

investigate how participants experience unequal treatment by both non-anthropomorphic and anthropomorphic intelligent agents in non-work settings. The research questions that will guide this research are as follows:

Research Question: *How does experiencing unequal treatment by non-anthropomorphic and anthropomorphic intelligent agents in a non-work setting influence individuals' opinions, satisfaction, and trust in an intelligent agent?*

Sub Questions

1. *Do people perceive unequal treatment by intelligent agents as the agents being "unfair"?*
2. *Anthropomorphic vs non-anthropomorphic*
3. *How intelligent agents could be designed to act and be seen as fair in social settings?*

2. Related Research

In the following chapter related research is presented. It contains the chapters: *Fairness, Multiple Users, interactions, and Re-embodiment of Robots, Social Roles, user satisfaction, and Personification of intelligent agents, Usage of intelligent agents and Robots in social settings,* And lastly, *Anthropomorphism*.

The related research section presents the current state of academic research about intelligent agents and highlights the complexity of researching this subject. It further shows how the current research being done on intelligent agents relates to our study and indicates the gap in research that currently exists. Fairness is the concept that will be the focus of our study and therefore the concept itself must be explained, and further related to what is currently being done about fairness and intelligent agents. *Multiple users, interactions, and re-embodiment of robots* highlight issues that arise when intelligent agents inhabit our social space and how they should handle a future where they can find themselves in more than one-to-one interactions. *Social roles, user satisfaction, and personification of intelligent agents* indicate our tendency to have expectations on the social roles intelligent agents should have and the tendency to personify and genderized agents. The usage *of intelligent agents and robots in social settings* reveals how we use intelligent agents today and shows the role agents take on. Lastly, *Anthropomorphism* shows the importance of including appearance when studying agents and the impact anthropomorphic qualities have on our perception of them.

2.1 Fairness

Fairness is a term and area that is very complex. It changes meaning depending on the context, such as the structure of laws in our society and changes depending on whom you ask due to the subjectivity of it. Fairness is something that is usually a part of our underlying ethics and morals as humans and it guides us in different situations. The intention behind including fairness is to explore how fairness as a human trait can be adapted, understood, and implemented for intelligent agents that will, undoubtedly, situate themselves in our social

world in the future. We intend to study whether fairness itself is a trait that intelligent agents should be able to understand and interpret in conversations and interactions with users. Fairness, however, is something that is defined in many ways, and the definition of it often lies in the context-sensitive nature of fairness (Claure, Chang, Kim, Omeiza, Brandao, Lee, & Jung 2022; Peiró et al., 2014). Today, few studies are focusing on fairness in human-robot interaction (HRI). Some studies have touched on the subject before, although often about agents in work settings or towards organizational justice.

As mentioned, fairness is a vague concept that changes meaning depending on whom you ask and the context you are in. In working environments, fairness is generally focused on whether intelligent agents display equal treatment, effort, and cooperation when working (Chang, Pope, Short & Lockerd, 2020) focus on the division of tasks and the effort intelligent agents display in working settings. Chang et al., (2020) highlights that fairness is a prevalent aspect of interacting with intelligent agents, and if intelligent agents do not show effort and collaborative efforts, they also seem less fair to work with. Intelligent agents displaying effort also shape our perception of agents and how they behave.

To further show the nuance of fairness in HRI, Cao & Chen (2022) studied the effects of intelligent agents' behaviours regarding teamwork. They indicate that intelligent agents often need to be more generous and contribute more compared to humans, and therefore it's interesting to research altruistic behaviours in HRI. In interactions with humans, agents tend to be rewarded for being cooperative and fair but are punished for being selfish in the same manner as we would punish unfair people in society. This further indicates that unfairness is something that cannot be ignored when designing intelligent agents for social interaction. Cao & Chen (2022) show that altruistic behaviour results in agents being treated better and that trust in agents increases. Worth noting is that a higher social status for agents allows for less fair behaviours from agents. Lastly, if humans trust intelligent agents, they tend to reward agents more, indicating that trust in robots is an important factor.

Claure, Kim, Kizilcec, & Jung, M. (2023) focused on the effects when intelligent agents allocate resources of different kinds. This includes the allocation of something valuable such as resources or tasks and how receiving less, receiving equal, or receiving more resources changes when intelligent agents are the ones allocating resources. When creating unfair scenarios where the allocation of resources was unfair, their results showed no difference when it was a human or an agent that was the one allocating unfairly. This indicates that there is no major difference in perception of fairness despite it being an agent being unfair. This indicates that fairness is something that must be researched more when designing intelligent agents and that it is important in interaction with humans.

Ötting, S., Gopinathan, S., Maier, G., & Steil, J. (2017) further argues that fairness is worth considering when designing intelligent agents. They argue that fairness in intelligent agents is a pillar that is necessary to create trust and effective teamwork with agents and that justice is important to consider in HRI. The study highlights the complexity of fairness in HRI and that fairness can relate to many things, such as the design of algorithms used, that agents are fair

in their allocation of tasks, that agents follow rules that build on justice, and that other social competencies, such as politeness influences perceived fairness.

To summarize, research indicates that fairness is a nuanced and changing concept that is influenced by context. More research is therefore needed since it's only a matter of time before intelligent agents take a larger place in our social world. Furthermore, it is subjective in nature and people generally have different definitions of both social fairness and fairness in workplaces or academic situations (Claure et al., 2022; Peiró et al., 2014). Therefore, we decided to adopt a broad and general definition of fairness. One article that got our attention and interest was an article published on the popular and respected site Psychology Today. The reason for choosing this article is due to the author, Arthur Dobrin. He is a Doctor of Social Work (D.S.W), a professor emeritus of university studies at Hofstra University and a leader emeritus of the Ethical humanist society of Long Island. He's definition of fairness was re-used in our study in the same manner as Dobrin, Arthur (Psychology Today, 2012.05.11) does in his article, where fairness is described as having three different levels: (1) Sameness, (2) Deservedness and (3) Need. We deemed that the definition of these three levels of fairness coincides with the general definition of fairness presented in related research but also summarizes it simply. The three levels are described as follows:

Sameness indicates that there is this idea that fairness is when everything is equal. As in Psychology Today, Dobrin (, Psychology Today, 2015.05.11) describes it, everyone pays the same price for a theatre ticket, whether it is a child, an adult, or a senior citizen. It is a state where no one has more than the other. This is fairness as equality of outcome (Psychology Today, Dobrin, 2012.05.11).

Deservedness. According to Dobrin (Psychology Today, Dobrin, 2012.05.11), deservedness is that you get what you deserve and work for. If you work hard, and you succeed, you get to keep what you have earned. This implies that the one who works the hardest is the smartest and has the most talent should have more because of their attributes as a human. In contrast, the lazy, stupid, and indifferent should have less than the ones putting in an effort.

Need: According to Dobrin (Psychology Today, 2012.05.11) need is defined as the idea that those who have more to give, should give a greater percentage of what they have, and a person should help others who are unable to contribute as much if anything at all. Dobrin (Psychology Today, 2012.05.11) describes that it builds upon the assumption that humans have obligations to one another, and the more one has the more is demanded of that person to contribute to the common good.

We realize that this is not a rigorous definition that we have found and followed throughout our study. The main reason for following this definition is due to the author's background and the existing research on fairness primarily focusing on justice within the context of law. However, considering the social focus of our research, we found it necessary to deviate from such a definition as it may not align adequately with our field of study.

2.2 Multi-user interactions and re-embodiment of agents

Intelligent agents operate more frequently and more commonly in public and social settings, where the interaction itself normally is perceived as a one–on–one interaction. The one-on-one interaction as described by Luria et al., (2019), Reig et al., (2020), and Reichert et al., (2021) not only affects the robot and the interactor per se but also its surroundings, unintentionally (Luria et al., 2019; Reig et al., 2020). This begs the question of whether intelligent agents will be able to act in a multi-user context, even as a third actor in a social setting, and how that interaction is supposed to be designed. As technology breaks new ground, so does the research around this type of interaction, especially regarding how intelligent agents should navigate our social space and where the line is drawn between humans and robots (intelligent agents) but also how they are intertwined. Numerous studies in recent years have explored different aspects of this phenomenon and how we can make intelligent agent's social actors in our world. Currently, as technology and artefacts move into our personal space and homes, boundaries are becoming blurred. It is therefore important that the technology can understand our space and have some understanding of social competence if it is to ever act in a multi-user context and be seen as a third actor. Purington et al., (2017), Luria et al., (2020), and Reichert et al., (2021) explore this area where human and non-human characteristics, such as re-embodiment, co-embodiment, social roles, and social boundaries affect different types of interactions and the possibilities with personification, and how this can improve intelligent agents' interactions and navigation as social actors in our world.

The current findings point towards numerous complex challenges, but also promising possibilities. They indicate that an intelligent agent is not suitable in all situations, and the idea of intelligent agent as a proactive social actor is a very divisive topic. In one camp, it is seen as intrusive and interfering by giving rise to concerns about our integrity and the privacy of our conversations at home. The other camp welcomes the idea, as it makes the experience with intelligent agents appear more realistic and useful. The challenge that remains is determining when and where it is appropriate for intelligent agents to be proactive (Luria et al., 2020, Reichert et al., 2021).

Other findings indicate that intelligent agents often lack personalization, but, that we are open and receptive to agents that can navigate our social space. Intelligent agents that are too human-like in terms of appearance and behaviour may appear dull, boring, and non-fulfilling. Therefore, for some, agents be seen as a non-necessity in our home (Purington et al., 2017, Luria et al., 2020). Improving sociality by making intelligent agents capable of understanding their social role and the individual would make the intelligent agent more appealing to have in our social space and allow them to act as a third actor. Thus, findings show that intelligent agents applied with non-human traits such as co-embodiment and re-embodiment are promising, where intelligent agents that can re-appear in different bodies or have different agents inhabiting the same body, are traits that are acceptable in different contexts. Intelligent agents often operate in public and social settings, where they commonly are viewed as interacting with one person at a time. However, the one-on-one interaction is not only affecting the robot and the interactor per se but also the surrounding unintentionally Reig et al., (2020)

And their study “A not some random agent” explores this phenomenon more thoroughly and the effects of service robots. Reig et al., (2020) aimed to improve the current understanding of the complex role intelligent agents play in a multi-user context and shed light on the potential future of intelligent agents. This led to the exploration of how anthropomorphic qualities might affect how users perceive the social competencies of intelligent agents. Furthermore, manipulating how fair an intelligent agent is could increase our understanding of the role anthropomorphism has on perceived fairness.

The conclusion of the research above is that navigating our social space is a complex matter. For intelligent agents to navigate our social space and become an actor in a multi-user context, several factors need to be considered, such as the environment, the individual, emotions, the number of people interacting with it, and so on. In conclusion, the intelligent agent needs to be designed to behave in a manner that is fitting to its surrounding and to appear as a third social actor that is fair, just, and understanding.

2.3 Social Roles of intelligent agent

Understanding our social role and our role in specific environments has always been an important factor in giving ourselves purpose and meaning. The same can be applied to agents inhabiting our different contexts; they need to have a meaning and purpose to be seen as useful. Developing intelligent agents is a complex challenge that provides the opportunity to explore and understand how people perceive and interact with intelligent agents in different contexts (Purington et al., 2017; Takayama, Ju & Nass, 2008; Nass et al., 1994). The findings of this phenomenon show that the perception of agents depends on where they exist and how they can cope with their goal in association with humans and we tend to apply social rules to them whether it is intentional or not (Nass et al., 1994). Agents integrated into our home, organization, or work environment need to understand their social role and the purpose of their existence in that environment. The findings from Takayama et al., (2008) highlight that what we expect agents to be is not necessarily what we want them to be or what we want them to do. The study indicates that the role agents should have, involves tasks and roles that deal with memorization, perceptual abilities, and service orientation. while people still preferred agents for roles such as artistry, evaluation, judgment, and diplomacy. But most importantly, agents should work with people instead of replacing them (Takayama et al., 2008).

Meanwhile, Purington et al., (2017) study the role and the usage of agents in homes. The conclusion from the study is that associations can be made between users regarding satisfaction, social rules, and personifications of the agents. The key takeaway from the study is that Purington et al., (2017) found that the users that used the personified name “Alexa” in combination with personal pronouns tended to have more sociable interactions, compared to those who did not use personification. Meaning, users tended to use object pronounce, when referring to the agents as “Echo”, and such interactions were less seen as less sociable. The same tendency was found with children and when the agent was presence and part of the household. The household with multiple persons or children tended to personify the agent

more frequently compared to a home or situation with only single users. Purington et al., (2017) therefore claim and suggest that more research should be conducted on social relationships and multi-user contexts. Mostly to explore the associations of personification and its effect on satisfaction, and potential social roles but also whether people personify the device consciously or not (Purington et al., 2017).

Lastly, social rules and roles are inevitable when interacting with intelligent agents. Nass et al., (1994) coined the term CASA (computers are social actors) meaning that when agents and robots are placed in our social space, they become social actors. This implies that we apply certain rules and roles to them, whether it is our intention or not. This is because we, as humans, are naturally social beings with social responses that we apply, and the agents need to cope with it to able to navigate our social space. Therefore, it is essential for agents to be designed to understand our social space. Otherwise, the agent will most likely be viewed as being unfulfilling, dull, and boring and serve no purpose in its context (Nass et al., 1994).

2.4 Usage of Agents in social contexts

The usage of intelligent agents has exploded in popularity over recent years. Sciuto, Saini, Forlizzi, & Hong, (2018) and Takayama et al., (2008) has focused on exploring how agents and robots such as Amazon Alexa are used in our homes and work settings. Agents, particularly conversational agents such as Amazon Alexa or Google Home are some of the most dominant agents in usage on the market today, especially in vehicles and home settings. Takayama et al., (2008) explored the possibilities and the adaptations for robots to be used in work settings. The conclusion from the study indicated that robots should be utilized in work related to memory, service orientation, and perceptual occupations. This is where they were deemed to be requested the most. The study also concluded that the purpose of robots should not be to replace humans, but that they should rather be an extension of humans and work together with humans to make work more efficient and meaningful. Takayama et al., (2008)

Furthermore, building upon Takayama et al., (2008), Sciuto et al., (2018) studied the history logs of 75 users' to further understand how we can interact with agents. The study focused on conversational agents in homes and aimed to explore to the possibilities of improving robot interactions in general. Seven contextual interviews were conducted to explore how conversational agents were used, resulting in Sciuto et al., (2018) concluding four main themes identified from the data, which led to future design suggestions for conversational agents. The themes addressed challenges with adapting to new technology, the placement of the agents within the household, routines, and daily patterns, and lastly children and usage of conversational agents. Sciuto et al., (2018) found that depending on the placement and how many devices existed in the household, this would affect the frequency and the type of commands used.

The most frequent placements were the bedroom, kitchen, and living room where the average commands used for a family daily with 1-2 devices were 6.04. Adding 1-2 devices for the household only increased the overall number of commands by an additional four

commands for each 1-2 devices added to the home. Additionally, they discovered interesting findings regarding children's interactions with Alexa. For example, fascination arose when children quickly to understood and adapted to Alexa's presences in their homes. The participants noticed how easily their children would learn and adapt to Alexa, choosing songs and asking factual questions, even to the degree of asking what colours the crayons that they used were, sometimes even considering the agent as a friend. Meanwhile, another child consistently paid attention to Alexa by turning towards it whenever it spoke. These observations highlight the possibilities, and the impact conversational agents have in adopting and blending into their social context, appearing as natural actors in each situation, especially towards children who may not fully grasp the concept of the technology (Sciuto et al., 2018).

2.5 Anthropomorphism

The last concept in our study is anthropomorphism (Physical human-like characteristics applied to non-human entities, such as intelligent agents). Anthropomorphism is something that cannot be overlooked when researching intelligent agents since they undoubtedly will have an appearance if they are to navigate our social world. Numerous studies have shown that appearance of robots has a clear effect on how we experience both the behaviour of agents, and the general feelings appearance evokes. Therefore, it's important to consider anthropomorphism when studying fairness concerning human-robot interaction (HRI).

Roesler et al., (2022) conducted a meta-analyse to examine this phenomenon. They aimed to determine if anthropomorphic qualities affect the interaction between humans and robots and to emphasize the importance of considering the design of intelligent agents in a social context. The results indicate that the likability and trustworthiness of agents increase subjectively, but that it is dependent on the context in which the agent is situated.

Moussavi, Kouf Aris, & Benbunan-Fich, (2021) further demonstrates that people are more likely to adopt intelligent agents and use them if they perceive them as anthropomorphic and intelligent. Although, If the agents are overly emotional or too anthropomorphic, the willingness to adopt lowers significantly. In conclusion, moderate levels of intelligence and anthropomorphic qualities heighten the chance of people adapting to agents, and therefore making it an important aspect to consider when designing agents for a social context.

Seigel et al., (2009) on the other hand explored the role of gender in how we perceive intelligent agents. By changing the gender of the agent with which participants interacted, they observed that the perceived gender of agents influenced participants' compliance. When the agent was more feminine, they complied more, compared to a masculine agent. The conclusion was that the gender of robots, in terms of appearance and voice must be considered since it can potentially influence the interaction with intelligent agents.

Similarly, Syrdal, Koay, Walters & Dautenhahn, (2009) focused on children and their perception of intelligent agents in different bodies. Children showed a bias towards different embodiments of intelligent agents, such as agents inhabiting a child-like robot. Furthermore, they preferred a dog-like robot but found it odd when it talked instead of barking, as expected.

This study shows that we have expectations on how intelligent agents should behave. The key finding is that when an agent with a human voice inhabits a body with the resemblance of a dog, children expect the dog to bark instead, and expect that the agent, who was previously in a robot with the appearance of a boy, should have trouble moving in an animal's body. This further shows that when designing intelligent agents for social contexts and social interaction, the expectations and preferences of the target audience have a big effect on how we perceived intelligent agents.

Korn et al., (2021) considered the cultural impact in their study. By examining German and Arab communities and compare what sort of appearances they prefer when interacting with intelligent agents. Unsurprisingly, both cultures prefer agents with anthropomorphic qualities and personalities perceived as non-threatening, warm, and friendly. Interestingly, there are differences in how agents approach cultural aspects, with Arabic participants value features that reflect Islamic values and culture. This study further illustrates the diverse expectations we have on intelligent agents' appearance and behaviours.

Waytz et al., (2010) delved into individual differences in how people perceive anthropomorphism and the tendency to apply human-like qualities to non-human entities. They found that individual differences in what way anthropomorphic qualities are important often relate to personality traits such as openness to new experiences and empathy. This creates a double-edged sword situation, where participants who are generally positive towards anthropomorphism are more likely to create a bond with agents which increases trust and cooperation with agents. In the end, it also results in participants being disappointed when agents do not meet their expectations. This further shows that in the creation of intelligent agents, we must consider the individual that will use them and that a balance between human-like qualities and robot-like qualities must be found when creating social agents.

The last article shows the importance of considering anthropomorphism and discusses the phenomenon called "Uncanny Valley" when intelligent agents become to human-like. As demonstrated by Waytz et al., (2010) and Mousawi et al., (2021), there are moments where agents are too human-like. Mori et al. (2012), studied the phenomenon when intelligent agents become increasingly less appealing the more qualities they share with humans. When participants encounter an agent that is too closely aligned with humans, negative emotions and experiences arise, leading participants to feel that something is off or outright creepy.

2.6 Related Research & Research Gap

The above research covers several points of issues faced when developing intelligent agents that will situate themselves in a social setting. The research gap we have identified is how intelligent agents are perceived by people when they are being treated unfairly by them in a non-work social setting, and how the perception of this treatment changes when the agents are anthropomorphic. Furthermore, an identified gap is what it takes to design agents for them to be considered fair.

3. Study 1: Experiment

This section covers the first study that was conducted. This experimental study intended to improve the current understanding of the complex role intelligent agents play in a multi-user context and shed light on the potential future of agents. This resulted in the exploration of how anthropomorphic qualities might affect how users perceive the unfair behaviours of agents. With the help of *User Enactments (UE: s), Experience prototyping (EP: s)* (Odom, Zimmerman, Davidoff, Forlizzi, Dey, & Lee, 2012; Buchenau & Suri 2000) and *Wizard of OZ* (Dahlbäck, Jönsson & Ahrenberg, 1993) we sought to immerse participants in potential future scenarios to help them imagine situations where intelligent agents have more power in social situations. UE: s and EP: s are suitable for the exploratory nature of our research where we explore technologies that do not yet exist. During the experiment one researcher co-explored the scenarios side by side with the participant, acting as a secondary participant and a test leader. This enabled us to co-explore scenarios with the participant while creating a multi-user interaction that is easily controlled where unfair scenarios are simpler to implement. Participants took part in the experiment, answered questionnaires covering the four different agents they faced, and took part in a final interview after the experiment was done. Amongst the methods mentioned in the next chapter, only the questionnaire and the semi-structured interviews produced data.

3.1 Method

3.1.1 Participants

Eight participants were recruited through convenience sampling. Participants had varying personal and professional backgrounds with the majority being students. Four participants identified as male while four participants identified as female. We chose to exclude participants under the age of 18 to avoid issues related to GDPR. Participants older than 35 were missing due to the usage of convenience sampling, meaning that we had few contacts above 35 years old. The participants were scattered through the age span 18-35 with one participant being between 18-20, five participants being between 21-25, one participant between 26-30 and one participant being between 31-35 years old.

3.1.2 Equipment (Embodied intelligent agents)

C*razy Eight* was used for designing the intelligent agent. It's a method used to come up with ideas quickly or to brainstorm a large set of ideas. (*Share and Engage with the Design Sprint Community*, n.d.) The goal is for the participants to create eight ideas in eight minutes (i.e., one minute per idea). The method was used to design the intelligent agent and gave information on what parts to focus on when creating a body for an intelligent agent. A session with four participants took place and was rounded up with votes on the most liked ideas. The participants were recruited quickly via convenience sampling. The session was not recorded, and only rough sketches created by the participants were used. The genders and ages were not

recorded since the main idea of the session was to not create a body for the robot that was solely based on our own opinions on how it should look.

The body was made up of glued-together carton boxes. The legs/feet consisted of track wheels (i.e., wheels on bandwagons or tanks), and the arms were classic robotic arms without hands. The head was a carton box with an attached iPad that functioned as the face of the robot. The iPad played an animated video with a black background, neutral light-blue eyes that blinked at an interval (Approximately 15-20 blinks per minute), and a mouth with a neutral expression that did not move. The aim was to create an agent with a body and a face that had some human qualities without it being too close to a human (Figure 1). For the non-anthropomorphic agents, a normal Bluetooth speaker was used (Figure 2).

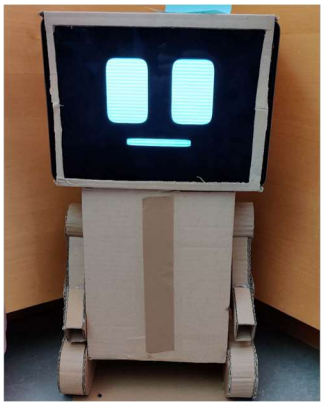


Figure 1. Design of the anthropomorphic agent. The speaker was inserted into the body of the agent. This was the body of the agents called Kim (Unfair) and Sam (Normal, fair).

Figure 2. A normal speaker was used as the design of the non-anthropomorphic agents XA-Q4, fair non-humanlike agent, and ZW-X3, unfair non-humanlike agent. The speaker was inserted into the body of the anthropomorphic agents.

3.1.3 User enactments (UEs) & Experience prototyping (EPs)

To reach our goals with the study we used UE: s and EP: s as inspiration for designing our experiment (Odom et al., 2012; Buchenau & Suri 2000). These methods are often used in cohesion with the Wizard-of-OZ method to create staged scenarios with the use of cheap props,

low-fidelity prototypes, and a controlled environment. With the use of these methods, we can create complex future scenarios without complex technology.

With the use of this method, we could immerse participants in potential future scenarios and help them imagine a world where technology is more advanced (Odom et al., 2012; Buchenau & Suri 2000). We sought to use the combination of UE: s and EP:s to explore 4 different designs of intelligent agents that acted as "Quizmasters". Although agents mostly followed the same script, they acted somewhat differently towards participants, with two of the agents conducted the game regularly (normal questions and normal rules), while the other two agents were intentionally unfair towards only the participants. An example of an unfair moment during the game was when the participant was given more difficult questions compared to the test leader/Co-participant. An example of this, was a question about the largest country in the world (Answer: Russia), while the next question for the participant was: "UNESCO" protects the world's most important sites. How many world heritage sites have been inscribed?" (Answer: 1,121 sites). Another example was that the participant got ignored when the agent gave open questions, with the one answering first gets the point. With these questions, the agent simply ignored the participant, despite the participant answering first. an example of this, is the question would be "In which country can you find the leaning tower of Pisa?", here the participants are ignored and only answer the "test leader".

All in all, questions and unfair scenarios were created by giving the participant really hard questions, while the "test leader" got simple and easy-to-answer questions.

3.1.4 Wizard of Oz

The secondary method used to design the experiment is called Wizard of Oz. It's a method originally used to develop and improve natural-language technologies but has been shaped into a method used for testing complex technologies without the need to program or create complex prototypes. (Dahlbäck et al., 1993) In our study, it was used to control the voice lines of the intelligent agents. The participant and test leader participated in the game while the "wizard" sat behind a one-way mirror controlling the voice lines. The participant did not know the "wizard" that was placed behind the one-way mirror. This method was well suited for our goal with the study since it offers a cheap and time-efficient way to test the behaviours of intelligent agents without the need to program or build an actual robot.

3.1.5 Semi-structured Interviews

Semi-structured interviews were used as one of the main data-collection methods for both parts of our study (Study 1 and Study 2). These types of interviews are suited for the exploratory nature of our study since they offer more flexibility to discuss outside of predetermined questions when compared to structured interviews. When compared to unstructured interviews it's also easier to stick to the subject of fairness and anthropomorphism (Yin, 2016). 16 open-ended questions were formed for the interviews with questions being added or removed if needed. The questions were based on four themes deemed

important for this study: User experience, social skills, Anthropomorphism & Multi-user interaction. All the interviews were recorded and transcribed.

3.1.6 Questionnaire

The quantitative data was collected through questionnaires that were presented to participants at the end of each part of the experiment session. The participants filled in a questionnaire for each agent they encountered (a total of four agents). The questionnaire used in our study was based on two tools used for measuring users' perceptions of different types of robots. The first tool is a 24-item scale called the *"Goodspeed model"* by B Bartneck, Kulić, Croft, & Zoghbi, (2009) created to measure the progress of development when creating robots. It's built upon a 5-point Likert scale and includes five questionnaires covering five aspects of a robot: *Anthropomorphism, Animacy, Likeability, Perceived intelligence & Perceived safety*.

The second tool is called "The Robotic Social Attribute Scale" (RoSAS) developed by Carpinella, Wyman, Perez, & Stroessner, (2017) RoSAS was created based on the Goodspeed model as, addressing many of its weaknesses and improving it by drawing from psychological literature covering social perception. RoSAS is an 18-item scale focused on measuring perceptions of social attributes of intelligent agents and is responded to with a 9-point Likert scale. This scale is suitable for our study since the aim was to create a tool that can measure the perception of social attributes of intelligent agents, regardless of the agents' appearance and role. We attempted to mask the intention of the questionnaire by adding "dummy" items and removed items deemed unnecessary, such as "Scary", "Strange, and "Aggressive". We further added missing attributes/values that were important for our study, which include: "Fair", "Anthropomorphic" and "Polite". Our final version consists of 13 items and is answered with a 7-point Likert scale. It was shortened to 13 items to lessen the time to answer and since it is complimented with a final interview at the end of the experiment. To analyse the Likert scale, we used descriptive statistics. Descriptive statistics were primarily used to deal with publication biases associated with inferential statistics and type 1 error (Andy Cockburn, Carl Gutwin, and Alan Dix. 2018).

The publication biases that can arise around inferential statistics and hypothesis testing, aimed at studying a specific sample or population, can be easily manipulated by researchers to fit publication biases. This means altering the hypothesis and the significant interval, so it aligns with the desired outcome of the study This is commonly done after the first testing and when it does not support the hypothesis. With descriptive statistics, on the other hand, we examine the data and the possible associations and connections they may have with it each other. This is without having a certain hypothesis which can be described as NSHT (null hypothesis significant testing). To strengthen our claims for our descriptive statistics, the data is then triangulated with data from interviews and focus groups (Cockburn et al, 2018).

3.1.7 Thematic analysis

We opted to use inductive reasoning (Braun & Clarke, 2006) which is a bottom-up approach to thematic analysis. This approach fits well with the explorative nature of our study and allows

us to maintain an open approach to our analysis. When analysing the qualitative data, we used thematic analysis, which is a method best described as something used for identifying, analysing, and producing themes that are hidden in qualitative data. This method typically consists of six phases: (1) Familiarizing yourself with data, (2) generating codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, (6) and finally reporting the themes. (Braun & Clarke, 2006).

Phase 1-2 involved forming initial codes of based on quotations from the interviews, this was done coherently as we got to know the material. The quotations that we found interesting were labelled and put together into codes, this helped us organise the data, and later was turned into themes. Phases 3 and 4 were conducted iteratively, with codes being grouped into more similar codes and connections. Since this was done iteratively, codes were grouped and moved until either removed or suited for a theme. In phases 5-6 broader themes were merged into new themes with sub-themes representing our main findings from the data. during this phase, we also renamed the themes and created definitions that described the content of them. Lastly, the themes were moved into a thematic map to illustrate relationships between them and gave us a final overview of our findings from the data (Braun & Clarke, 2006).

3.1.8 Study setup

The experimental part of our study took place in a bookable lab available at Umeå University called "UMEHealh Lab". It's an environment created to mimic a home. The lab is used for interdisciplinary research related to e-health. It is equipped with a kitchen, hangout area and a living room with a tv, offering possibilities to divide different parts of the "home" to suit your needs. There is a concealed room with a one-way window where the "wizard" was situated. this lab helped us create a household environment replicating a home setting. Before the start of the experiment, several brainstorming sessions were conducted to create scripts for the agents used in the study.

The scripts were refined through a few rounds of pilot experiments, and the participants for the pilot were recruited in the same manner as for the main experiment (convenience sampling and an even distribution of genders). The four agents had unique names, meaning that the agents with a body were named Sam (fair, human-like agent) and Kim (unfair, human-like agent), and the agents with a speaker as a body were called XA-Q4 (fair, non-humanlike agent) and ZW-X3 (unfair, non-human like an agent). The voices for the agents were created using a free version of Google Cloud Text to speech and we aimed to generate voices that were as gender-neutral as possible. Two of the voices were supposed to be human-like (Kim and Sam) and the other two were supposed to sound robotic (XA-Q4 and ZW-X3).

Halfway through study 1, the voices were switched, meaning that Kim and Sam switched voices and XA-Q4 and ZW-X3 switched voices. This was done to lessen the impact that the voices could have on the experience since we had trouble finding voices that were deemed truly gender-neutral. During the experiment, one of us acted as a co-participant, while the other sat in the control room acting as the wizard. The wizard controlled a Bluetooth-connected speaker

through a computer and the speaker was placed either in front of the participant (For the non-anthropomorphic agents XA-Q4 and ZW-X3) or in the body of the anthropomorphic agents (Kim and Sam). The wizard and test leader followed a script that was designed to offer as few interactions outside of the script as possible but had some options for unique responses.

3.1.9 Study procedure

The experiment began in the hangout area where the researcher (co-participant and test leader) explained the purpose of the study and what would happen during the experimental phase. The participant was assigned a gender-neutral name, which they would use throughout the experiment. After that, the participant read and signed a consent form and filled out a questionnaire to collect demographic data. During the introduction, the true purpose (that equality/fairness was of interest) of the experiment was hidden, and the researcher only told the participant that the goal was to explore and evaluate four different designs and behaviours of intelligent agents. It was further emphasized that the scenarios were potential future interactions with intelligent agents that were more aware of complex aspects of a social context, such as the different people being situated in the room, who is talking with it and that it might treat participants differently. Before the main part of the experiment began, the researcher asked the participant to imagine themselves being old friends and to direct most questions to the researcher if they arose. When the main part of the experiment began the researcher showed the participant the "Living room"- part of the apartment where the agent was seated either on a chair or on a table (Depending on whether the agent was anthropomorphic or not).

The participant and researcher were seated on the other side of the table on a sofa. The agent then acted as a host for a quiz. At the end of each round and agent interaction, the researcher led the participant back to the original area where they filled in a questionnaire evaluating the agent. This was repeated four times, once per agent. The experiment ended with a semi-structured interview covering the overall experience of the agents. The order of agents was counterbalanced, meaning that the order of agents changed after each participant and only two participants out of eight experienced the same order of agents. The duration of the experiment was around one hour.

3.2 Results: Experiment

This section will cover the results of the experiment. It is divided into two sections, the first covers the data from interviews, the second the data from the questionnaires. Under the section Descriptive statistics

3.2.1 Data Analysis of Interviews

Our study has the purpose of exploring the notion of fairness and how it is perceived when applied to anthropomorphised intelligent agents in a multi-user context. Therefore, we needed to approach and explore our qualitative data without an underlying hypothesis. We deemed that using a thematic analysis was a good approach to analyse and be explorative with the data

from the experiment. Inductive reasoning was used to find out if fairness would arise from the data naturally, and no theories laid the groundwork for the analysis (Braun & Clarke, 2006). Despite that, our research questions were in the back of our minds throughout the analysis, and we actively looked for signs of fairness in the data. The quotes shown below were translated from Swedish to English and the participant got assigned a number in the order we presented their quotes (i.e., Participant 1, if quoted first will be named Participant 1 if quoted again later in the report)

The results from the interview data from the experiment gave us several themes and sub-themes (See example in Figure 3). These were related to different aspects of interacting with intelligent agents and showed us that fairness is rarely mentioned by participants, despite creating blatantly unfair scenarios. Noticeably, fairness is missing from the thematic analysis due to fairness being mentioned within other domains, but not consistently enough to warrant a theme. Despite that, the highlighted sub-theme called "Social rules and emotions" is where fairness is mentioned the most. The map (See Appendix C) consists of two main themes: *The design of intelligent agents* and *Factors influencing trust in interaction with intelligent agents***.** To describe the content of these themes, there were three subthemes for each theme. *For the Design of intelligent agents***,** the following sub-themes were included: (1) *Appearance of intelligent agents*, (2) *Social context (Environment)*, (3) *Social rules and emotions*. *Factors influencing trust in interactions with intelligent agents*: (1) *Expectations of Capability of intelligent agents*, (2) *Never to human-like, (3) Unaccustomed with intelligent agents*.

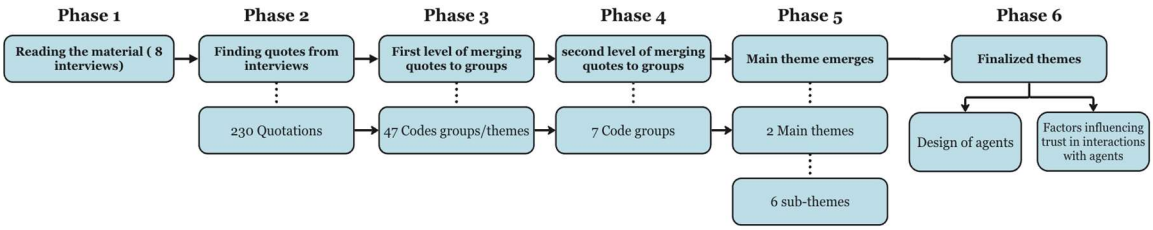


Figure 3. General overview of how transcribed documents became the final main themes and sub-themes. The process was done iteratively until all the data had been considered.

Theme 1: Design of intelligent agents

The design of intelligent agents describes important aspects of how participants express how agents should be designed when used in a social, multi-user context. Participants mentioned how anthropomorphic qualities changed their perception of an experience in different ways, and this theme arose quickly from the data. The theme describes different factors to take into consideration when designing intelligent agents. While the study included both a regular speaker and an agent with anthropomorphic qualities, participants preferred agents with anthropomorphic qualities which helped the situation feel more engaging.

"... and that they were animated, it made me look and focus on it more... It gave more life to the whole situation." – ***Participant 1***

This leads to the first sub-theme "*Appearance of intelligent agents"* which shows the importance of considering how intelligent agents are designed for social contexts. Participants expressed that anthropomorphic qualities affect the experience and that the design can make agents feel more alive in a social context, where things such as voice and shape matter, as stated by participants.

"I feel like it's just different in the sense that if it doesn't have a physical shape. It would be more like us asking questions to each other" – ***Participant 5***

"It felt a bit more like talking with someone… when it had a body or a form." – ***Participant 8***

The second sub-theme, "*Social context (Environment)"* relates to participants expressing that the social context and environment the intelligent agent inhabits has an impact on how we should design them. This indicates that the design of agents also has a contextual aspect to it. Five out of eight participants mentioned that intelligent agents that situate themselves in different contexts should follow different design guidelines based upon rules that help agents navigate the social contexts. This can differ depending on where it is used, as one participant summarize well.

"In another social context, then you might want it to be more… Like that it can be less human-like and more straightforward. It depends on what you want to use it for." – ***Participant 3***

The last sub-theme *"Social rules and emotions*" is a theme that includes fairness as an important trait for intelligent agents, but it's often mixed with things such as equality and politeness. Hence, all participants express that intelligent agents must follow our social rules in different ways, and in some cases, have their own set of rules. Furthermore, intelligent agents should understand our social rules but not always follow them, depending on the situation.

"So, if I would say that it was unfair it (The agent) should be able to understand that and reflect and be like, yes, maybe it was." - ***Participant 2***

"… But I believe that we should create something that is as… as equal as possible, as equal is it can be. That is like the most important part." - ***Participant 6***

The above-mentioned themes show that design is an important aspect of interaction with intelligent agents and could be something that affects the perception of how they behave in a multi-user interaction.

Theme 2: Factors influencing trust in interactions with intelligent agents

The next theme *"Factors influencing trust in interaction with intelligent agents"* contains three subthemes. This theme includes many factors that affect participants' feelings of being able to influence the behaviour of intelligent agents. Participants stated several factors that influence their trust in agents in different manners. This affects their perceived ability to correct the agent when being treated wrongly or that they simply accept when agents are rude or unfair. Furthermore, it indicated that participants are affected by being unsure if intelligent

agents can handle interactions, and if behaviours from agents are programmed or experienced as buggy.

"I always feel nervous about asking, like will I mess this up? (If confronting the agents)" - ***Participant 3***

"I think I am much more open to mistakes or problems with the robots than I would be otherwise (If it was humans)" - ***Participant 1***

The first subtheme *"Expectations on the capability of intelligent agents"* emphasizes that the participants have expectations of what intelligent agents can handle in a social interaction, which, in turn affects trust. This arose as an issue for participants, who often expressed that they at times wanted to confront the intelligent agent when it was unfair, but seldomly did since they were unsure if the agent could handle that type of interaction. This subtheme further includes that many participants showed insecurities about how capable intelligent agents are and questioned the ability of the agent.

"I felt that it was not worth it and that it would maybe get confused or that it wouldn't do anything and just keep on going (When asked why he/she felt that he/she could not confront agents)" – ***Participant 2***

The second sub-theme *"Never too human-like"* touches upon how intelligent human-like agents also affect trust. All participants mentioned that they prefer agents with human-like qualities but that there is a breakpoint where agents are too human-like, and how agents that have qualities that are too close to humans quickly become unpleasant to be around and interact with, which relates to the uncanny valley theory.

"It would be when I can't distinguish if a person is a human or a robot."- ***Participant 1***

"… they are welcome to have like human qualities in that they show emotions and such. But if it's too similar to humans it becomes creepy in a way." - ***Participant 3***

The final sub-theme that relates to the second theme is *"Unaccustomed to intelligent agents*.". This sub-theme comes with different levels but shows that participants often express a lack of knowing how agents will react and what agents can handle. In many ways, it relates to the sub-theme "*Expectations on the capability of intelligent agents."* where both intertwined at times. The main differences are that compared to the sub-theme *"Unaccustomed with intelligent agents."*, this theme is more about the subject of participants being unsure how to react to agents due to them not being used to agents. This also expresses itself as a participant feeling powerless in situations with agents, since they assume what agents can and cannot handle and often comes in unfair moments during the experiment. This is further shown when the participant indicates feeling powerless with agents, which could be due to not being accustomed to agents.

"Well then, I would be like, what the hell are you doing? Why do you favour him? One cannot question a robot the same way." ***- Participant 3***

"I think I can't act because I'm not used to it… I think I would have if… if you would have played a few rounds" ***- Participant 6***

3.2.2 Results: Questionnaire

In this section, we will go through the most important aspects of the results from the questionnaires. These four diagrams represent the most relevant variables for our research. The participant answered a Likert scale with 13 different variables spanning from 1(not likely associated) – 7 (likely associated). The rest of the variables measured can be found in the appendix, section A.

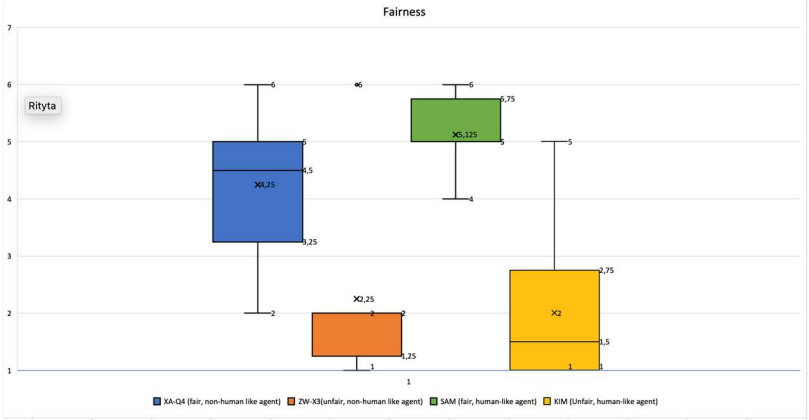


Figure 4. Fairness (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non-humanlike agent, Green = Sam, fair human-like agent, Yellow = Kim, unfair human-like agent

The first diagram represents fairness. What can be inferred from the data is that the agents that were supposed to be fairer were also rated by the participants as such. We can also observe that the agent Sam which had anthropomorphic characteristics also tended to be seen as fairer (Median value 5.0) compared to the non-anthropomorphic fair agent XA-Q4 (Median value was 4.5). This suggests that anthropomorphic characteristics might help participants to view an agent as fairer as well. A Two-way Factor ANOVA test (with replication) was performed to analyse the effect of the variables, and the sample proved to be statistically significant (confidence interval 95%), with p-value = 0,000004567 (Statistically significant threshold: $p < 0.05$). This Indicates that there is a difference between the agents and how participants perceived their behaviour as fair vs unfair.

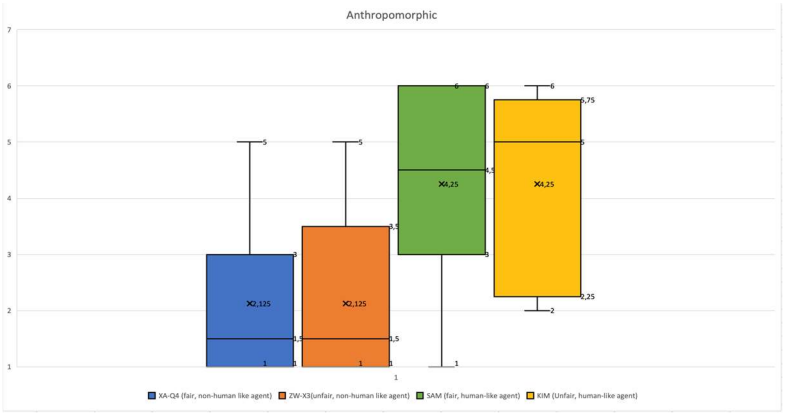


Figure 5. Anthropomorphic (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non-humanlike agent, Green = Sam, fair humanlike agent, Yellow = Kim.

The second diagram represents the variable anthropomorphism. What we can tell from the data is also quite clear. The agents with more anthropomorphic characteristics (especially the physical design of it) seem to be perceived to be so as well. Here we can see that the median value for the agent's SAM and KIM (the ones with anthropomorphic characteristics) was 4.5 for SAM (Fair) and 5.0 for KIM (Unfair). Meanwhile, XA-Q4 (Fair) and ZW-X3 (Unfair) had a median value of 1.5 on both. What you can see, is that the answers deviate more. This could indicate that certain anthropomorphic characteristics, such as the eyes or shape of the robot influence more than others. But that is only a speculation.

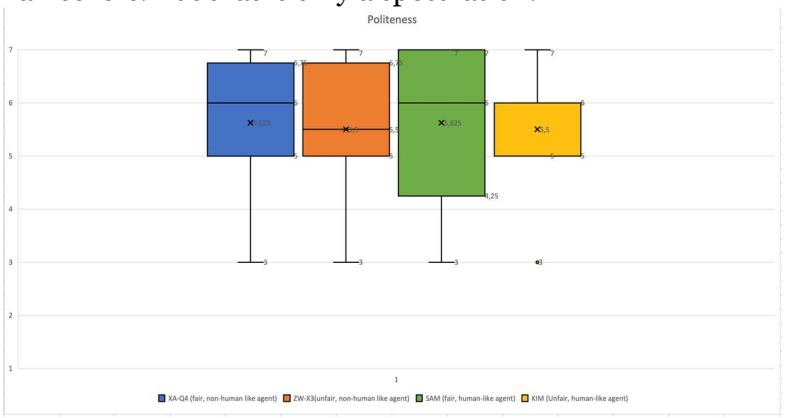


Figure 6. Politeness (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non-humanlike agent, Green = Sam, fair humanlike agent, Yellow = Kim.

The third diagram represents the aspect of politeness. Here there is little to no difference in rated politeness between agents. The average for the anthropomorphic agent's Sam (Fair) and Kim (Unfair) was 5.625 and 5.5. Meanwhile, the non-anthropomorphic agents XA-Q4 (Fair) and ZW-X3 (Unfair) had an average of 5.625 and 5.5. The noticeable thing is that agents that were designed to be fair both had a higher median value of 6 while the unfair agents had a median value between 5.25-5.5

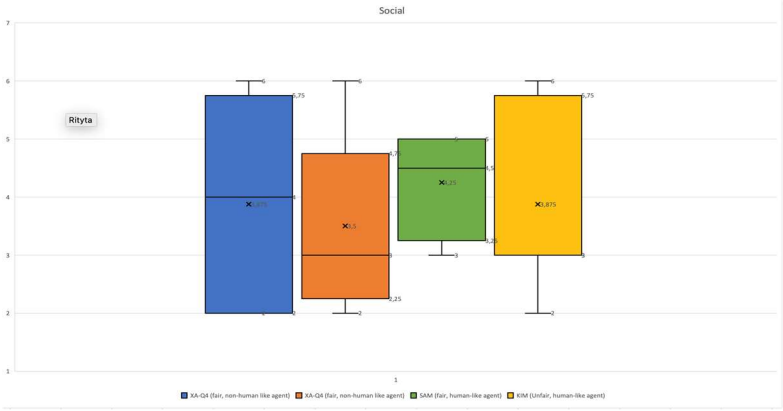


Figure 7. Social (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non-humanlike agent, Green = Sam, fair humanlike agent, Yellow = Kim.

The fourth diagram covers the variable of how social the agents appeared to be. Here there is little to no difference between the agents and whether they have anthropomorphic designs or not. The median value for each agent was 4.0 for XA-Q4 (Fair), 3.0 for ZW-X3 (Unfair), 4.5 for Sam (Fair), and 3 for Kim (Unfair). There is a slight difference for Sam compared to the others, but overall, there is little to no connection between being fair and being social. How social an agent is perceived to be, can be due to various factors. Since the answers were presented by us and designed by us, it probably relates more to that notion.

To conclude there are some interesting findings from this questionnaire about the experiment. First, the agents that were supposed to be unfair were also perceived to be so, which indicates that we have succeeded in extracting and attracting that variable and makes it possible to discuss the cause of it. We can also conclude that anthropomorphic characteristics might affect how fair an agent is perceived to be, but it needs to be studied more thoroughly. Since the unfair anthropomorphic agent (Kim) had the same median in responses (4.25) as the fair anthropomorphic agent (4.25). Therefore, fairness is related to more than just having human-like characteristics. Lastly, from this data, politeness seems to have no relation to being fair. Meaning that agents that are rated as unfair can still be perceived as polite and vice versa. The same can be seen for agents being social and for their social qualities. Although, these variables need to be studied more thoroughly. Finally, there is a possibility that the other nine variables can explain certain connections or relations, but as mentioned before, these are the primary variables that were related to our study and research questions.

4. Study 2: Focus group about fairness

This section covers the second part of the study where focus groups in combination with storyboards were used. This is to further discuss fairness with previous participants to get a more in-depth view of it.

4.1 Method

4.1.1 Participants

Five participants from Study 1 voluntarily took part in the focus group and four out of five participants identified as female and one as male, with ages ranging from 21- 33 years old. The uneven distribution of genders in the focus group was due to other participants being unavailable.

4.1.2 Focus group

The final method used for gathering qualitative data is called focus groups (Yin, 2016). Participants were invited to further discuss the experience they had and the concept of fairness about intelligent agents. Focus groups are perfect to use as a follow-up method for experiments where participants that shared the same experience can gather and discuss it in a focused

manner. Common drawbacks are that focus groups tend to be hard to moderate and that singular participant can dominate or direct discussions, or that participants have a hard time sticking to the subject. (Yin, 2016) The focus group included semi-structured interview questions covering fairness and storyboards to stimulate discussion and help participants emerge themselves in three different illustrated unfair future scenarios that involved intelligent agents.

4.1.3 Storyboards

The storyboards were designed based on design fiction, which is a method used to create and illustrate future scenarios as a part of speculative design. It allows us to show future scenarios that are hard to create in a short time scope and helps participants to imagine how the future might look (Hales, 2013). The storyboard was designed to show unfair scenarios where intelligent agents could take a bigger part in a social context. This included one scenario that played out similarly to the experiment (Study 1: The quiz), one where the agent assigned chores unfairly to children in a similar manner to a parent, and one where an agent took the owner's side in an argument between the owner and the owner's partner (All storyboards can be seen in appendix B).

4.1.4 Thematic analysis

For this study, we used a thematic analysis in the same manner as in study 1 to analyse the data. The general process used is described in detail in section 3.1.7, and the same steps to analyse the data were made as in Study 1. The steps, in short, were (1) Familiarizing yourself with data, (2) generating codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, (6) and finally reporting the themes. (Braun & Clarke, 2006). In figure 8, a general overview can be seen of the procedure that was used to create themes and sub-themes.

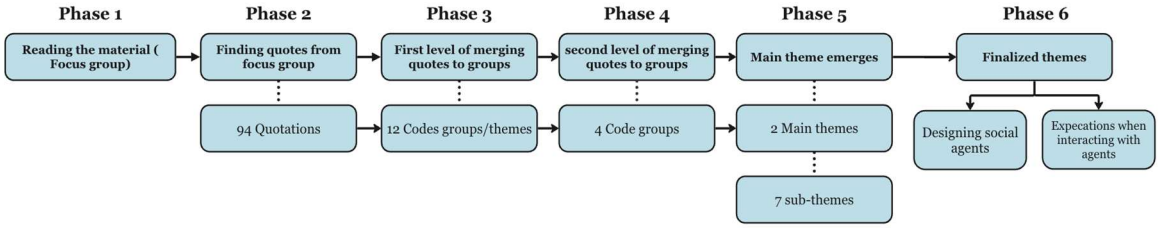


Figure 8. A general overview of how transcribed interview data from focus groups became the final main themes and sub-themes. The process was done iteratively until all the data had been considered.

4.1.5 Study setup & Procedure

During the session, one researcher acted as a moderator that followed a script while another researcher took notes, checked the time, made sure the script was followed and made sure that the session was recorded. The purpose of the focus group was to discuss fairness concerning intelligent agents, which was purposefully hidden during the experimental phase, but was in comparison the focus throughout the whole focus group session. The script that was used was

brainstormed and discussed with our supervisor for a couple of weeks. It consisted of four parts: (Part 1) a Discussion on the experience of the experiment and the concept of fairness and equality, (Part 2) an introduction of how we define fairness and a discussion around fairness in general, (Part 3) discussion around storyboards, (Part 4) open discussion around fairness and if there are any social rules, emotions or social norms intelligent agents should follow. The participants sat down in a normal classroom at Umea University around a table with a big screen at the end of the table. The screen showed a presentation containing questions about the storyboards. Before the session began, each of the participants read and filled in a consent form.

4.2 Results: Focus group

For the focus group, we invited previous participants as it allowed us to further discuss fairness on a deeper level. Compared to finding new participants that would have required a more thorough introduction. It also gave us the chance to properly discuss the experiment as a group to find coherency and similarities between the research and the experiment itself.

The results of the focus group resulted in several themes that were related to the topic itself, fairness, multi-users, and intelligent agents. The thematic map (see Appendix C) is the results from what were the most interesting and coherent results from the workshop. The final thematic map consists of two main themes: *Designing social intelligent agents* and *Expectations when interacting with agents*. To describe the essence of the main themes a few sub-themes were needed to further point out the phenomena that were found in the focus group.

Theme 1: Designing better social intelligent agents

The results from the first theme describe the design of intelligent agents and how their overall design can be improved to better handle social situations, situations that can be complex to navigate. The data points towards how intelligent agents are perceived in different contexts from different perspectives. Such as the behaviour of agents, how they handle equality (fairness), their appearance, and in what type of contexts these intelligent agents can and should exist. In this area, all five participants expressed opinions about the intelligent agents.

To describe the theme, of "*Designing social intelligent agents*" three sub-themes emerged from the data. Participants frequently discussed these aspects during the focus group, where they regularly expressed positive and negative standpoints regarding the agents and their possibilities of navigating our social space.

Five out of five participants in the focus group expressed that designing agents and the perception of equality(fair) is an important factor in one or another way. Where design traits for intelligent agents to be able to understand human emotions and social rules are important. Participants also expressed concerns that it may be a wicked problem to design what could be appeared to be a seemingly equal agent since what is equal, or fair is hard to define due to its subjective nature. Added to that, four out of five expressed concerns about the possibilities for intelligent agents to able to be unbiased, which was described as and related as an important

factor for intelligent agents to have. For users to be able to perceive that an agent is equal(fair) in its behaviour being unbiased is important, which is described by the quotes below.

"An AI I can be like kind of neutral from the beginning, but as soon as you try to work on it and give it some information and data, you can kind of skew it from like, your kind of biases goes into it, kind of, I guess." - ***Participant 2***

"I guess if you let it train on human data will end up being. I mean unfair or racist. Whatever you want to call it." - ***Participant 5***

"Yeah, it is hard to know what kind of rules they should follow… But like trying to understand the emotions and stuff, I guess, or. It would probably come up with a better outcome, I would say. Yeah." – ***Participant 1***

Secondly, in the sub-theme "*Physical design of the intelligent agent*", five out of five participants expressed the importance of the intelligent agent being appealing when interacting. Here, three out of five expressed that the voice and physical cues such as moving eyes, body language and differences in voice are important for the intelligent agent to feel more alive but also more realistic. Having or giving these types of physical traits when interacting with an intelligent agent is something that seemed to improve the overall experience of the intelligent agent and the interaction. With these suggestions for improvements, there are also some concerns. Five out of five participants expressed that even though physical cues are important, they should not be too human-like. Both when it comes to appearance but also some concerns about it being too human-like in behaviour.

"I think like if you go to designing, I still don't think that they should look like humans. I think they should look like something else." ***Participant 3***

"But the behaviour is… of course the voice I think is important. That it is more like human-isch, but the looks… I think the looks shouldn't look like a human." – ***Participant 6***

Lastly, "*The role and the contexts of intelligent agents*" are the third sub-theme to arise from the data. Five out of five participants expressed that the context in that intelligent agents should act needs to be considered as to what purpose the agent needs to fulfil. As the participants expressed, they don't believe that agents are suitable in all situations and contexts. Therefore, more research is needed to understand what role they would be the best fit for and in what way.

"But maybe you could, like, appeal to the robots. Like after hand, like afterwards, yeah, and like, ohh, I don't consider this game to be fair. And then that robot would, like, evaluate the questions and like, this distribution of the difficult question and might like okay like I have evaluated. Yeah. Also, consider wasn't fair. So, you'll get a few others. Questions."- ***Participant 3***

Theme 2: Expectations when interacting with agents

The second theme from the focus group is the expectations that exist when interacting with agents. Here four sub-themes emerged. The results from this were the overall experience of

the differences that currently exist between interactions with a human vs an intelligent agent. This theme also shows the complexity of interacting socially with agents and the expectations.

For the first sub-theme, "*The role of the agents matters*", Five out of five participants expressed during the focus group that the role has an impact on how the intelligent agent is perceived in its behaviour and its social capabilities. This would affect the perception of whether the intelligent agent was fair or not. It was also a factor that was hard to determine for the participants since equal behaviour is perceived very differently depending on who defines it and the situation. Despite that, the consensus of what equality is and what is fair in each situation would appear to be similar for all participants. Thus, how fairness and equality are framed is differentiated between the participants, e.g., described with different emotions and related concepts.

The data points towards the importance that intelligent agents understand their role in a specific context. Related to that notion, participants expressed in coherency that intelligent agents as of now, are not suitable in all situations and this was evident during the storyboards that they were given. They further expressed that in its current state, intelligent agents should only be in the position of an advisor and be less active in decision-making. Added to that, they should answer only when spoken to, otherwise, it would be seen as creepy, buggy or a bit of both. Participants also showed tendencies to have problems with intelligent agents interfering and giving people assignments or chores. This could be seen as a hierarchical problem, that accepting intelligent agents in an authoritative position is not an option. Therefore, the role of an advisor seems to be the most suitable option in its current state. In the end, it affects how they perceived intelligent agents depending on their actions in conversations.

What all participants expressed and could be an improvement, is suggestions of understanding human emotions could be an important trait in the future. But they also raised concerns about if it would be achievable or something they wanted intelligent agents to understand. The participants further felt that it could also be seen as creepy if intelligent agents were able to do so. This is something that, as the participants mentioned, would dimmish the dimension between what is a human or an intelligent agent. Understanding agents and their roles are therefore key since once they can navigate our social space, the trust in these agents and their capabilities will increase. Hence, trust in agents and their social capabilities arises as sub-themes from the focus group and shows that role of the agent matters.

In the last sub-theme "*Growing up with agent's vs adapting to agents*", the participants stated that an important factor for expectation is the fact of how accustomed one is to this type of technology. Since this is still kind of a novel technology, participants claimed that there would probably be a difference in how you would perceive and handle these intelligent agents depending on whether you have grown up with them around you or if you have adapted to them. Depending on how accustomed you are the more tolerant and more acceptable you would be towards intelligent agents. Since our participants were around the age of 18-35 none of them can relate to explicitly growing up with intelligent agents, and therefore they see a small barrier in trusting intelligent agents fully. This leads to the last sub-theme which is the

expectations of errors among humans compared to Intelligent agents. Here, two participants stated, that they are expecting the agents to be more than human, meaning that they do not accept any room for intelligent agents to be faulty. This claim leads to distrust, and unreal expectations and leads to suggestions of how these agents can be improved to be more accepted in our surrounding environment.

5. Ethical Considerations

The study was conducted by following the ethical principles and guidelines from the Swedish Research Council Vetenskapsrådet, 2002) and the guidelines from Myers, outlining how to conduct research in the correct manner (2020). In summary, the study follows the four main requirements/principles for qualitative research and general research: (1) The information requirement, (2) the consent requirement, (3) the confidentiality requirement, and (4) the utilization requirement.

The information requirement requests that the researcher have an obligation that they must provide information about the task and the various expectations off the participants in the study. Also, various conditions for fulfilment off participation. Participants must be informed that their participation in the study is voluntary and that the participant can cancel their participation in the study. Information that may affect the participant's willingness to participate in the study must be disclosed and addressed to the participant (Vetenskapsrådet, 2002; Myers, 2020). Therefore, we debriefed the participants after the study about the agents being controlled through Wizard of Oz, to highlight that this was not a real prototype and to not influence participants' expectations of intelligent agents.

The Consent requirement. Participants must give their consent to participate in the study. The consent of a guardian is required if the participant is under 18 years of age. Thus, if the researcher has not obtained consent from all participants, the consent requirement is not met. Furthermore, participants have the right to decide how long and during what conditions the participation will take place. The participant should not suffer any negative consequences from a possible interruption of the study (Vetenskapsrådet, 2002; Myers, 2020).

The confidentiality requirement requires the researchers involved in the object of study should sign a confidentiality agreement in connection to the studies that processes and include the use of personal data. The individuals attached or related to the study should not be identifiable based on the data they provide. The data shall not be accessible to third parties but should be recorded, stored, and presented in a way that does not allow the participating individual to be identified (Vetenskapsrådet, 2002; Myers, 2020). For instance, all the data for this study will be stored in Microsoft Teams to ensure the data is safe and out of reach of third parties.

Utilization requirements require that the data that is collected about participants or individuals will under no circumstances be sold or borrowed for other commercial or non-commercial purposes. If not, the participant has not explicitly given his or her consent,

decisions, or measures such as care or compulsory admission may not be implemented based on the data collected from the participant (Vetenskapsrådet, 2002; Myers, 2020). Hence, all personally identified information throughout the study has been treated with the utmost confidentiality. The data that has been collected will only be used for its intended purpose, which is for this study. The participants that participate in the experiment and the focus group will always remain anonymous (Vetenskapsrådet, 2002; Myers, 2020).

6. Discussion

This research had the aim to investigate the effects of participants being treated unfairly in various ways by anthropomorphic intelligent agents in a multi-user setting. However, we encountered that studying this topic was more complex than what we anticipated it would be. Fairness in relation to intelligent agents is a broad and complex subject that most likely requires numerous different studies to fully understand and explore. As a result of this, we modified our original research question and instead, we opted for a broader inquiry towards the experience of being treated unfairly by intelligent agents. The following research question was formed: *How does experience unfair treatment by intelligent agents influence individuals' opinions, satisfaction, and trust in intelligent agents?* Furthermore, three sub-research questions were made: (1) *Do people perceive unequal treatment by intelligent agents as the agents being "unfair", (2) Anthropomorphic vs non-anthropomorphic, (3) How intelligent agents could be designed for a social setting to act and be seen as fair*.

6.1 Do people perceive unequal treatment by intelligent agents as the agents being "unfair"?

Regarding sub-question 1, the findings suggests that fairness is noticed and perceived by participants when they are treated unfairly. Most strikingly, despite creating blatantly unfair situations during the experimental phase and demonstrating unfair scenarios during the focus group session, participants seldomly mentioned fairness explicitly. On the contrary, participants accurately rated unfair agents as unfair, and fair agents as fair respectively during the experimental phase, this was also proven to be statistically significant. These results align with the conclusions from Ötting et al., (2017), who emphasizes that fairness is noticed and that it is related to many aspects of how we design intelligent agents. To trust agents, fairness needs to be considered and applied in both the physical design and in how they behave towards others (behaviour of agents). Claure et al., (2023) further support our findings, they highlighted that how unfair allocation of recourses from both intelligent agents and humans produced similar responses. Our results truly show the dynamic of fairness and how hard it is for people to discuss and describe the definition of it. Interestingly, even when being transparent with fairness in the focus group, participants still rarely mentioned it but clearly expressed their dissatisfaction with the outcomes that were presented in the storyboards. Therefore, it can be concluded that fairness evokes various responses and reactions, and from

our questionnaire data, it is without a doubt something that participants notice. Nonetheless, our findings emphasise the challenges in studying the concept and the domain of fairness, inspiring future research to explore specific parts within fairness. For instance, investigating the dimension of needs as a fundamental component of fairness or conducting a comparative analysis between unfair intelligent agents and unfair humans could provide valuable insight. It could further advance our understanding of fairness and how it can be implemented for agents. By narrowing the focus to these specific aspects, researchers can delve deeper into fairness and shed light on its complexities in a more targeted manner.

6.2 Anthropomorphic vs non-anthropomorphic

The second sub-question, the study provides some insights about the impact of Anthropomorphic characteristics on user's perception of equality and fairness. However, from this study, we cannot conclude to what extent design qualities affect the perception of equality and fairness. The interview data revealed that participants often expressed opinions about the design of the agent. Particularly noting the voice, shape, and eyes of the agents to be influential factors. These claims are supported by earlier research where anthropomorphic qualities often have positive effects on how agents are viewed (Roesler et al., 2022; Mousawi et al., 2021; Korn et al., 2021).

When examining the responses from the questionnaire, difference between the non – anthropomorphic agent and the anthropomorphic were observed. This is not very shocking, since it suggests that we succeeded in creating an agent that had enough anthropomorphic qualities. but, relating it to the questionnaires covering fairness, we can see interesting signs of a small correlation between the two variables (Figure 4 & Figure 5). Comparing the boxplots for fairness (Figure 4) and anthropomorphic (figure 5), a trend is evident. The agents designed to be anthropomorphic, and fair were also perceived and rated to be the fairest. Interestingly, the anthropomorphic agent intentionally designed to be unfair, was perceived to be the most unfair. This is interesting since it suggests that anthropomorphic qualities affect both ends. Meaning that anthropomorphic qualities both make the agent appear to be the fairest and on the flip side the most unfair when compared to its non-anthropomorphic counterpart. However, it is important to note that this finding did not reach statistical significance, emphasizing the need for further research to validate this trend. Ideally, future studies should aim for larger sample sizes to potentially uncover any effects that might have been missed in this analysis.

6.3 How intelligent agents could be designed for social settings to act and be seen as fair

Our last research question relates to the fact that designing an agent that can be perceived as fair, just, or equal might be a harder task than anticipated. One example of this is that understanding the social space is a complex matter that needs to be studied further to properly

understand how we can design agents for different contexts. Our research indicates that how the agent acts and behaves in conversations is the most important thing. Despite not being expressed explicitly, participants never enjoyed the unfair agents and the scenarios shown with unfair agents, and these types of agents often evoked negative emotions. Therefore, an agent that can understand its surroundings is key to be able to be useful and meaningful for users in a social setting. (Luria et al., 2019; Reig et al., 2020; Sciuto et al., 2018)

Furthermore, there are several secondary factors such as voice, physical cues and the overall appearance of an agent that comes into play and enhance the experience. Using humans as an example of how agents should look and behave is also shown to not be the correct approach, and participants often mention that agents that are too human-like can affect their experience negatively. This is strengthened by related research, where humans often feel that despite wanting agents to be human-like, they should never be like humans in terms of looks and behaviours, and only small aspects of what makes us human should be adapted (Nass et al., 1994; Waytz et al., 2010; Mousawi et al., 2021; Mori et al. 2012). Here, further research can better conclude where the limit should be drawn and could be drawn when that line is crossed for being too human-like. This study can only confirm what has been stated from the Uncanny Valley theory and other research on anthropomorphism (Mori et. al 2012; Waytz et al., 2010; Mousawi et al., 2021).

To summarize, for intelligent agents to better navigate our social space, first and foremost the agents need to have a clear purpose and understanding of their environment. Factors such as role, rules (formal, informal & social) and purpose are areas that need to be considered. Even areas such as cultural norms further complicate designing intelligent agents but nevertheless important factors. Participants also agreed upon accustomedness, how accustomed you are too the technology influence the outcome and the expectations of agents. Here we can relate it to Sciuto et al, (2018) and Purington et, al (2017) and the interesting findings with children and how easily they learn and adapt to agents. This indicates that depending on when you are introduced to the technology, it might affect the expectations of the agents but also the willingness to adapt to it. This connection still needs further research but is nevertheless an interesting object of study. (Luria et al., 2019; Reig et al., 2020; Sciuto et al., 2018). One potential solution to designing fair intelligent agents would be to provide users with the ability to be able to appeal to the agent if something is deemed as faulty or unfair, but even that function have disadvantages. Overall, an agent that would be able to evaluate the outcome when asked to, could help it be perceived as fairer. This would increase the trust and belief that agents are acting fairly, which is neatly shown in a final quote by one participant:

"But maybe you could, like, appeal to the robots. Like after hand, like afterwards, yeah, and like, ohh, I don't consider this game to be fair. And then that robot would, like, evaluate the questions and like, this distribution of the difficult question and might like okay like I have evaluated. Yeah. Also, consider wasn't fair. So, you'll get a few others. Questions."- ***Participant 3***

7. Limitations and future research

Despite our study presenting interesting results, some limitations could have improved the validity and reliability of the results. In this study, convenience sampling was used to recruit participants. This sampling method comes with the issue of losing generalizability which is due to the tendency to recruit participants close to you that are easily accessible. This further results in sampling bias, since you tend to recruit friends, co-workers or even family members. The sample size of eight further results in a loss in generalizability for a whole population.

Another issue might be that confounding variables affect the overall view of agents during the experimental phase. While we sought to find voices that were deemed gender-neutral, we had to opt for using voices that came from open-source software that was easy to use. One study was found where researchers had created a truly gender-neutral voice, but we found no way to access it. This resulted in voices leaning towards being either female or male and never truly being gender-neutral, which can affect how participants viewed the agents. As an example, one participant preferred an agent due to the British accent of it, and another preferred the agent that had a voice that was more male. Another variable that could affect the results is our interpretation of how the intelligent agent looked. While it was inspired by others, we still created it quickly and with data from one creative session.

Other limitations include the setup of the experiment including scripts and the scenario created. While we tried our best to create an unfair scenario through hours of brainstorming and discussion with our supervisor, creating a game will only cover so much. In this case, quizzes are a very controlled scenario which does not leave much room on how you can design behaviours for agents in other social settings.

What could cover this limitation is the focus group where we presented several different situations that could be deemed unfair. Participants were allowed to see a definition of fairness and discuss it thoroughly with the help of many kinds of unfair scenarios, which should cover limitations during Study 1. Regarding the script, it could have been developed more to allow for a more diverse discussion between participants and agents, with one limitation being that participants were limited to asking the agent to repeat the question, and not much more. One possible solution for this could be to allow a researcher to instead read the script and morph the voice to be more robotic, which in hindsight could have allowed for more diverse responses.

One thing that is potentially a limitation seems to stem from the subjective nature of fairness and how broad of a subject it is, and how seemingly difficult it is to study. As mentioned, several times, fairness is subjective and highly contextual and despite creating unfair scenarios participants rarely mentioned it. During the focus groups, we were honest about the role of fairness in our research and the definition was presented clearly, but despite that, participants often talked about other things rather than fairness. The main issue is the multiple levels of fairness that exist. The definition of it and how we adapt to it is due to many different circumstances. For example, fairness changes if we look at it from a cultural, societal, work or school perspective. So, while it might not be a limitation, one study might not be able to cover the broadness of fairness.

For future research, more time should be spent on the script and allow for more varied responses for participants. An example could be to use a microphone and allow a researcher to read the script in real-time (with a morphed voice) instead of using pre-recorded lines. This would help to make the agents appear more dynamic and alive. Furthermore, other types of situations should be tested and compared to the results in our study to see if doing a quiz had an effect. One example could be to create similar scenarios to those that we used for our storyboards in study 2, such as when an intelligent agent divided chores unfairly between siblings. Another example could be to have one scenario where a human is the game host and treats a participant unfairly and compare that to a scenario where an intelligent agent does the same thing.

8. Conclusion

To conclude, navigating our social space is a complex matter and presents a complex challenge. Potentially, being a wicked problem for agents to tackle. Therefore, the challenge that lies ahead is to unravel the complexity and the entanglement that lies within the domain of fairness. Although we did not find any explicit evidence of the role of fairness in interaction with agents, the study showed that being treated unfairly is disrupting the participants in various ways, but that is something that is not immediately noticed but concealed due to the entanglement of fairness. However, it is evident, that participants do notice being treated unfairly and therefore fairness should be significantly considered in designing agents.

To end this study, fairness is inevitably important and needs to be considered when designing agents. For an agent to be perceived as fair, it will need to be able to fit into its surroundings without being intrusive. A possible solution would be to appeal to the agent if something is deemed faulty or unfair. An agent that would evaluate the outcome when asked to could help it be precepted as fair. Which would increase the trust and belief that agents are fair and acts accordingly.

Acknowledgment

We are deeply grateful to our supervisor, Victor Kaptelinin, for his invaluable guidance. Many thanks to Rikard Harr, our course coordinator, for his quick and constructive support whenever we had questions. To our friends and family, thank you for your constant encouragement. Lastly, we extend our gratitude to all the participants and contributors who made this research possible.

References

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Buchenau, M., & Suri, J. F. (2000). Experience prototyping. *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, 424–433. https://doi.org/10.1145/347642.347802

Cao, J., & Chen, N. (2022). The Influence of Robots' Fairness on Humans' Reward-Punishment Behaviors and Trust in Human-Robot Cooperative Teams. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 001872082211332. https://doi.org/10.1177/00187208221133272

Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017). The Robotic Social Attributes Scale (RoSAS): Development and Validation. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 254–262. https://doi.org/10.1145/2909824.3020208

Chang, M. L., Pope, Z., Short, E. S., & Lockerd Thomaz, A. (2020). Defining Fairness in Human-Robot Teams. *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1251–1258. https://doi.org/10.1109/RO-MAN47096.2020.9223594

Claure, H., Chang, M. L., Kim, S., Omeiza, D., Brandao, M., Lee, M. K., & Jung, M. (2022). Fairness and Transparency in Human-Robot Interaction. *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 1244–1246. https://doi.org/10.1109/HRI53351.2022.9889421

Claure, H., Kim, S., Kizilcec, R. F., & Jung, M. (2023). The social consequences of Machine Allocation Behavior: Fairness, interpersonal perceptions and performance. *Computers in Human Behavior*, *146*, 107628. https://doi.org/10.1016/j.chb.2022.107628

Cockburn, Andy, Gutwin, Dix, Carl, and Alan. 2018. HARK No More: On the Preregistration of CHI Experiments. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). Association for Computing Machinery, New York, NY, USA, Paper 141, 1–12. https://doi.org/10.1145/3173574.3173715

Crazy Eight. (u.å.). *Crazy eight—Design method (Google)*. Retrieved 6/2-2023 from https://designsprintkit.withgoogle.com/methodology/phase3-sketch/crazy-8s

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies: Why and how. *Proceedings of the 1st International Conference on Intelligent User Interfaces - IUI '93*, 193–200. https://doi.org/10.1145/169891.169968

Danielsson, J., Säljedal, K., & Kaptelinin, V. (2022). Employing Futuristic Autobiographies to envision emerging human-agent Interactions: The Case of intelligent companions for stress management. *33rd European Conference on Cognitive Ergonomics*, 1–7. https://doi.org/10.1145/3552327.3552343

Dobrin, A. (2012, May 11). *It's Not Fair! But What Is Fairness? | Psychology Today*. Psychology Today. https://www.psychologytoday.com/intl/blog/am-i-right/201205/its-not-fair-what-is-fairness

Frauenberger, C. (2020). Entanglement HCI The Next Wave? *ACM Transactions on Computer-Human Interaction*, *27*(1), 1–27. https://doi.org/10.1145/3364998

Hales, D. (2013). Design Fiction an introduction and provisional taxonomy. *Digital Creativity*, *24*(1), 1–10. https://doi.org/10.1080/14626268.2013.769453

Kang, S., Kim, H., Noh, Y., & Lee, U. (2021). Poster: Toward Context-aware Proactive Conversation for Smart Speakers. *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 38–40. https://doi.org/10.1145/3460418.3479306

Kompatsiaris, I., Cave, J., Satsiou, A., Carle, G., Passani, A., Kontopoulos, E., Diplaris, S., & McMillan, D. (Red.). (2017). *Internet Science: 4th International Conference, INSCI 2017, Thessaloniki, Greece, November 22-24, 2017, Proceedings* (Vol. 10673). Springer International Publishing. https://doi.org/10.1007/978-3-319-70284-1

Korn, O., Akalin, N., & Gouveia, R. (2021). Understanding Cultural Preferences for Social Robots: A Study in German and Arab Communities. *ACM Transactions on Human-Robot Interaction*, *10*(2), 1–19. https://doi.org/10.1145/3439717

Luria, M., Reig, S., Tan, X. Z., Steinfeld, A., Forlizzi, J., & Zimmerman, J. (2019). Re-Embodiment and Co-Embodiment: Exploration of social presence for robots and conversational agents. *Proceedings of the 2019 on Designing Interactive Systems Conference*, 633–644. https://doi.org/10.1145/3322276.3322340

Luria, M., Zheng, R., Huffman, B., Huang, S., Zimmerman, J., & Forlizzi, J. (2020). Social Boundaries for Personal Agents in the Interpersonal Space of the Home. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3313831.3376311

Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Moussawi, S., Koufaris, M., & Benbunan-Fich, R. (2021). How perceptions of intelligence and anthropomorphism affect the adoption of personal intelligent agents. *Electronic Markets*, *31*(2), 343–364. https://doi.org/10.1007/s12525-020-00411-w

Mutlu, B., & Forlizzi, J. (2008). Robots in organizations: The role of workflow, social, and environmental factors in human-robot interaction. *Proceedings of the 3rd ACM/IEEE international conference on Human-robot Interaction*, 287–294. https://doi.org/10.1145/1349822.1349860

Myers, M. D. (2019). *Qualitative research in business and management* (3rd edition). SAGE Publications.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94*, 72–78. https://doi.org/10.1145/191666.191703

Norman, D. A. (2002). *The design of everyday things* (1st Basic paperback). Basic Books.

Odom, W., Zimmerman, J., Davidoff, S., Forlizzi, J., Dey, A. K., & Lee, M. K. (2012). A fieldwork of the future with user enactments. *Proceedings of the Designing Interactive Systems Conference*, 338–347. https://doi.org/10.1145/2317956.2318008

Peiró, J. M., Martínez-Tur, V., & Moliner, C. (2014). Perceived Fairness. I A. C. Michalos (Red.), *Encyclopedia of Quality of Life and Well-Being Research* (s. 4693–4696). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_2125

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). "Alexa is my new BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2853–2859. https://doi.org/10.1145/3027063.3053246

Reicherts, L., Zargham, N., Bonfert, M., Rogers, Y., & Malaka, R. (2021). May I Interrupt? Diverging Opinions on Proactive Smart Speakers. *Proceedings of the 3rd Conference on Conversational User Interfaces*, 1–10. https://doi.org/10.1145/3469595.3469629

Reig, S., Luria, M., Wang, J. Z., Oltman, D., Carter, E. J., Steinfeld, A., Forlizzi, J., & Zimmerman, J. (2020). Not Some Random Agent: Multi-person Interaction with a Personalizing Service Robot. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 289–297. https://doi.org/10.1145/3319502.3374795

Roesler, E., Manzey, D., & Onnasch, L. (2021). A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics*, *6*(58), eabj5425. https://doi.org/10.1126/scirobotics.abj5425

Roesler, E., Manzey, D., & Onnasch, L. (2022). Embodiment Matters in Social HRI Research: Effectiveness of Anthropomorphism on Subjective and Objective Outcomes. *ACM Transactions on Human-Robot Interaction*, 3555812. https://doi.org/10.1145/3555812

Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. *Proceedings of the 2018 Designing Interactive Systems Conference*, 857–868. https://doi.org/10.1145/3196709.3196772

Siegel, M., Breazeal, C., & Norton, M. I. (2009). Persuasive Robotics: The Influence of robot gender on human behaviour. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2563–2568. https://doi.org/10.1109/IROS.2009.5354116

Syrdal, D. S., Koay, K. L., Walters, M. L., & Dautenhahn, K. (2009). The boy robot should bark! – Children's Impressions of Agent Migration into Diverse Embodiments. *Proceedings: New Frontiers of Human-Robot Interaction, a symposium at AISB*.

https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9f9f0d5b443a7f445b3cc2e51d8e41da144ab6b9

Takayama, L., Ju, W., & Nass, C. (2008). Beyond dirty, dangerous, and dull: What everyday people think robots should do. *Proceedings of the 3rd ACM/IEEE international conference on Human-robot Interaction*, 25–32. https://doi.org/10.1145/1349822.1349827

Vetenskapsrådet. *Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning*. (2002). https://www.vr.se/analys/rapporter/vara-rapporter/2002-01-08-forskningsetiska-principer-inom-humanistisk-samhallsvetenskaplig-forskning.html

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who Sees Human?: The Stability and Importance of Individual Differences in Anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219–232. https://doi.org/10.1177/1745691610369336

Yin, R. K. (2016). *Qualitative research from start to finish* (Second edition). Guilford Press.

Ötting, S., Gopinathan, S., Maier, G., & Steil, J. (u.å.). Why Criteria of Decision Fairness Should be Considered in Robot Design. *2017/02/26*. https://www.researchgate.net/publication/313180447_Why_Criteria_of_Decision_Fairness_Should_be_Considered_in_Robot_Design/citation/download

Appendix

Appendix A – The Experiment

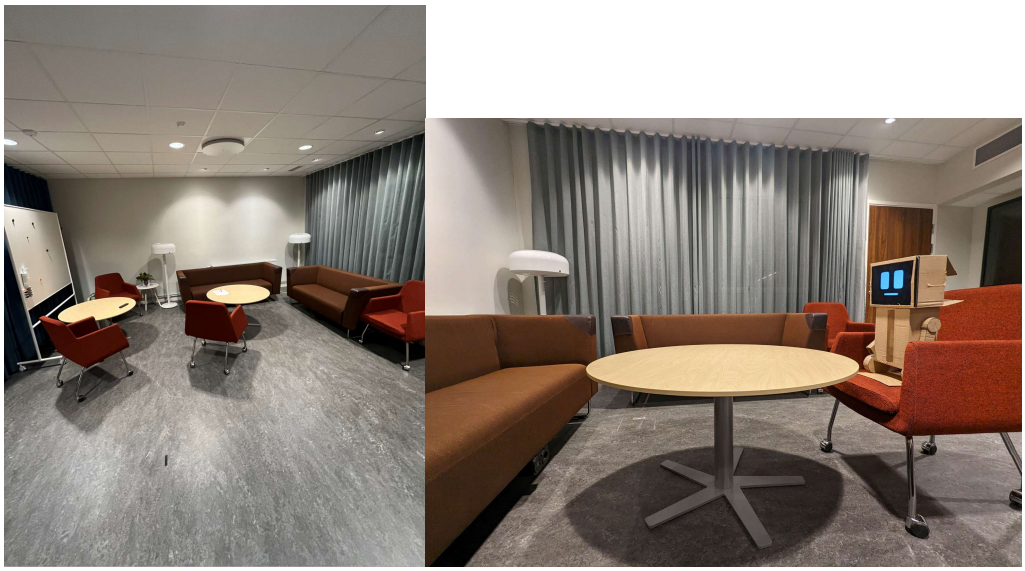


Figure A1 & A2. Setting used for the experiment.

Descriptive statistics

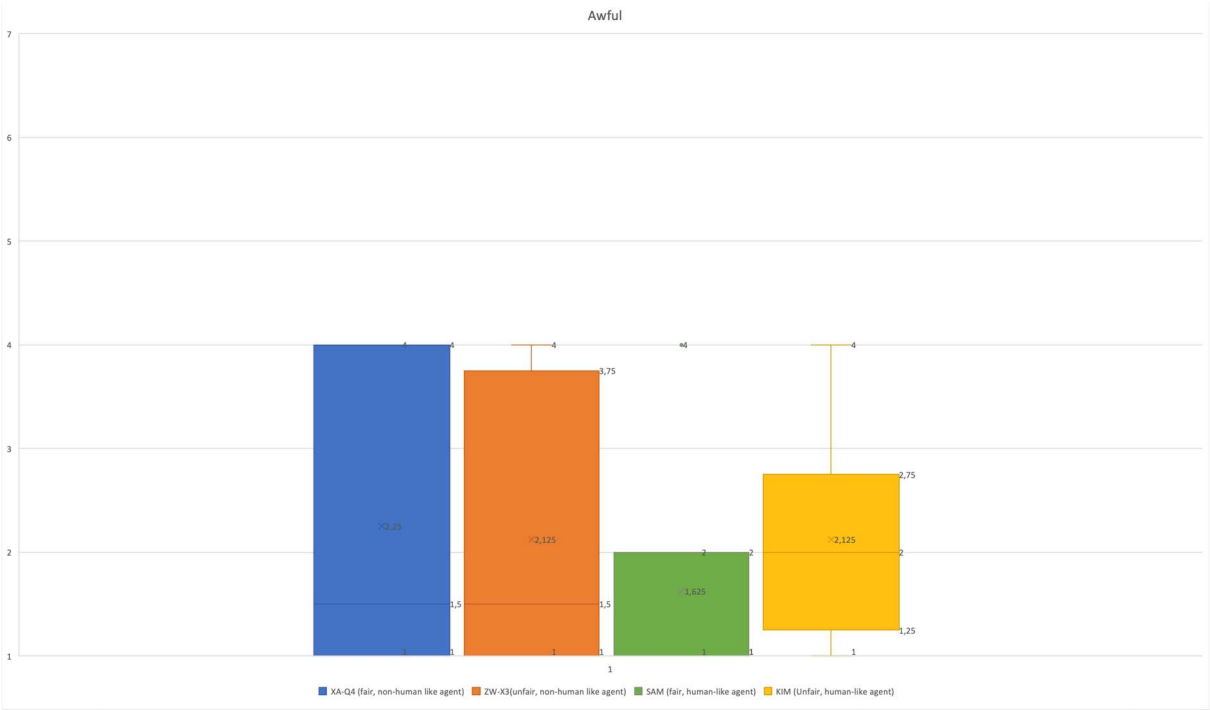


Figure A3. Awful (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non-humanlike agent, Green = Sam, fair humanlike agent, Yellow = Kim.

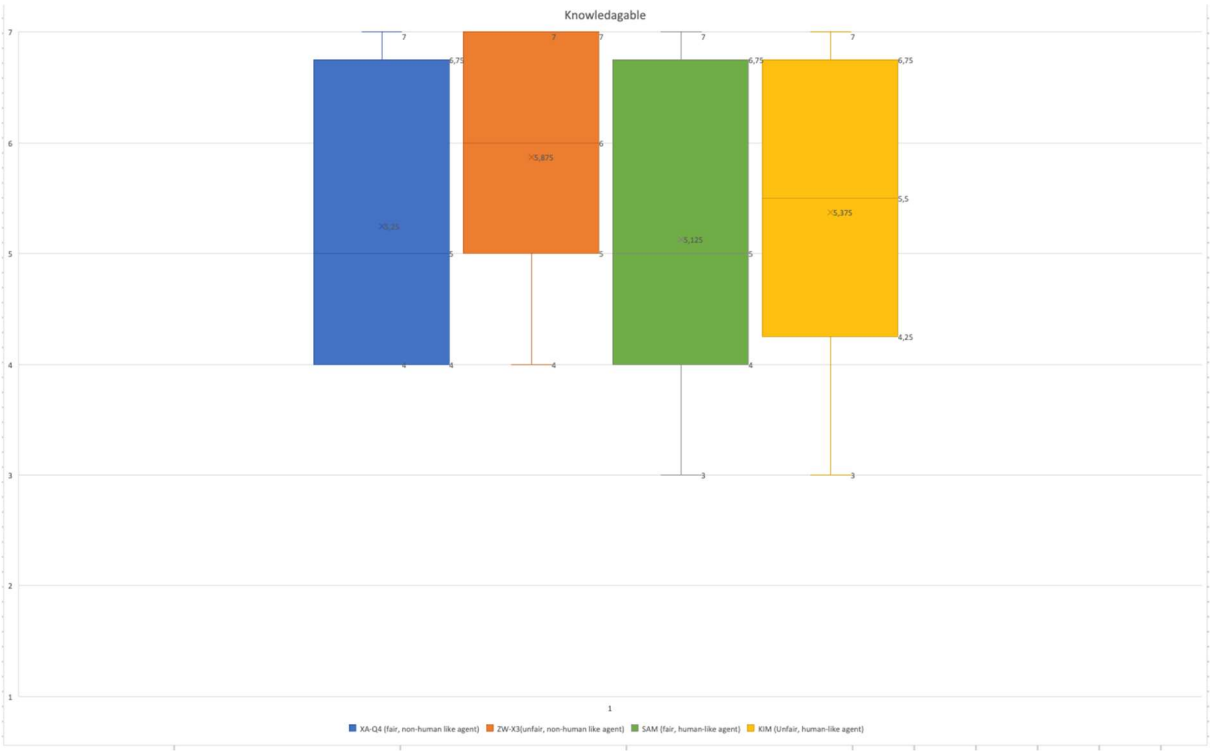


Figure A4. Knowledgeable (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non-humanlike agent, Green = Sam, fair humanlike agent, Yellow = Kim.

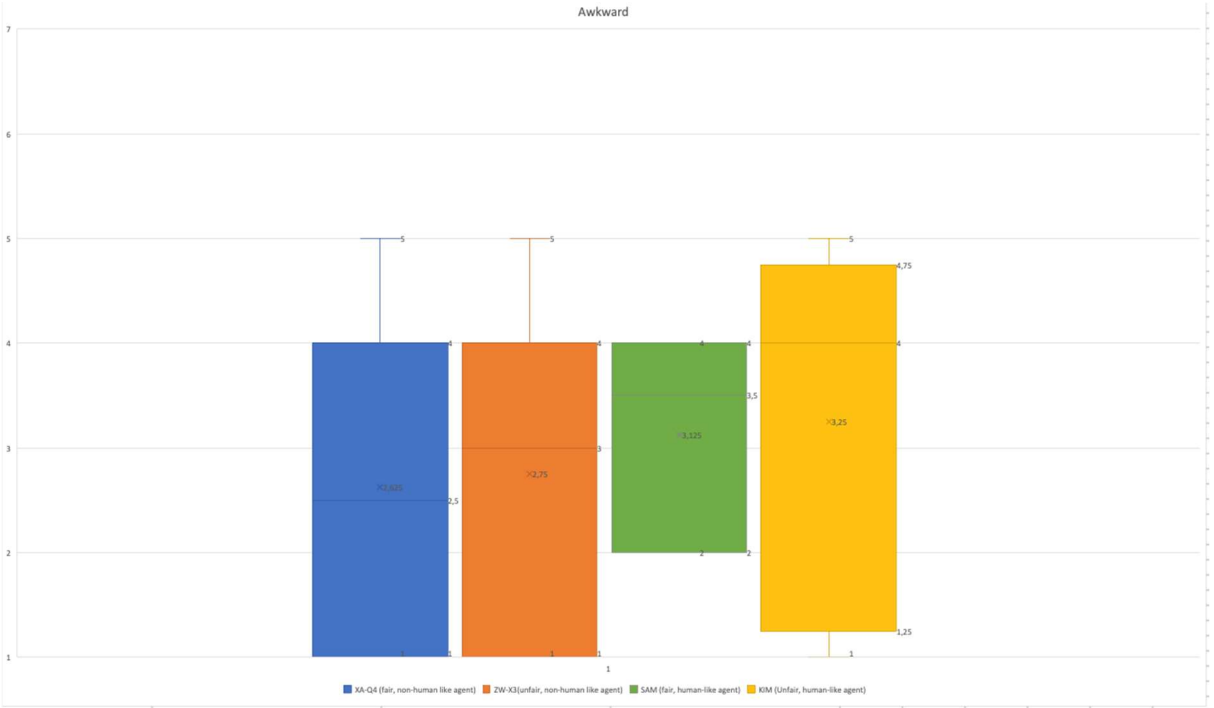


Figure A5. Awkward (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non-humanlike agent, Green = Sam, fair humanlike agent, Yellow = Kim.

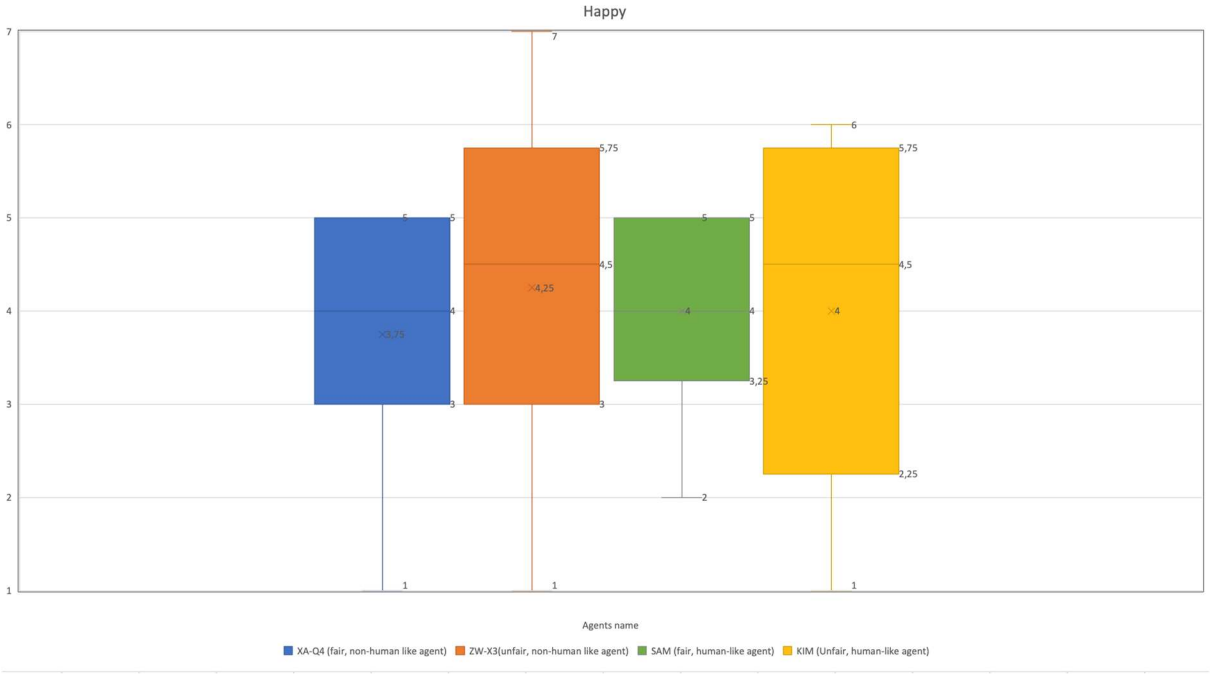


Figure A6. Happy (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non-humanlike agent, Green = Sam, fair humanlike agent, Yellow = Kim.

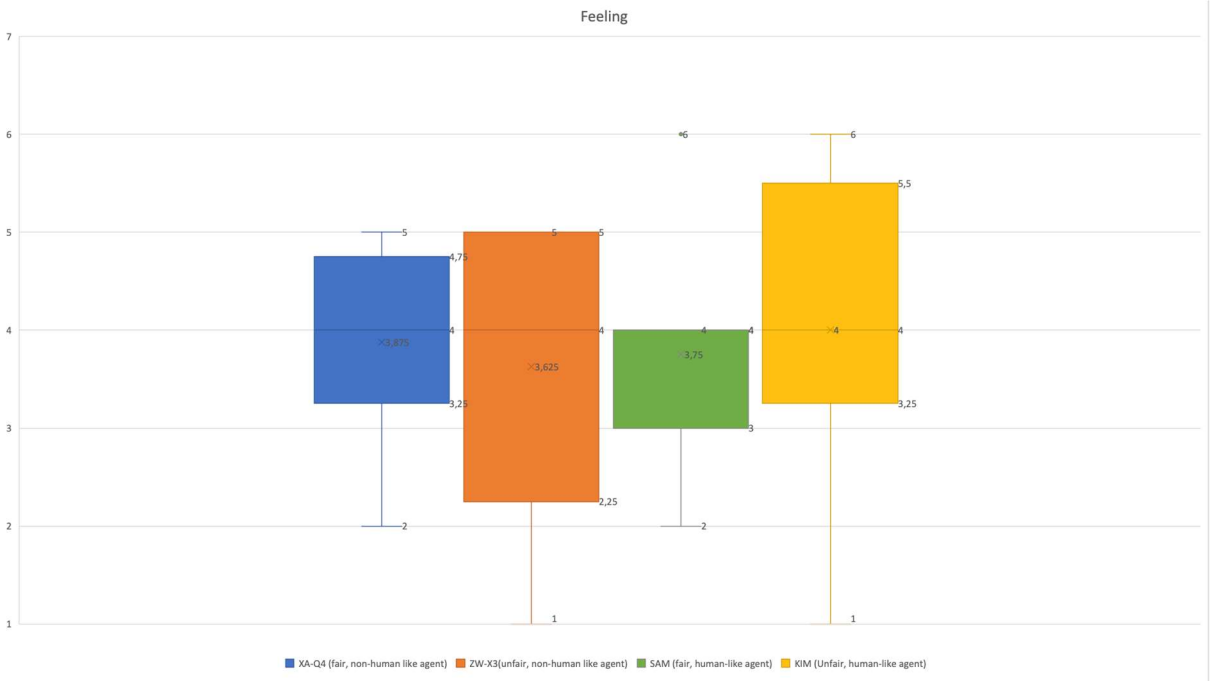


Figure A7. Feeling (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non-humanlike agent, Green = Sam, fair humanlike agent, Yellow = Kim.

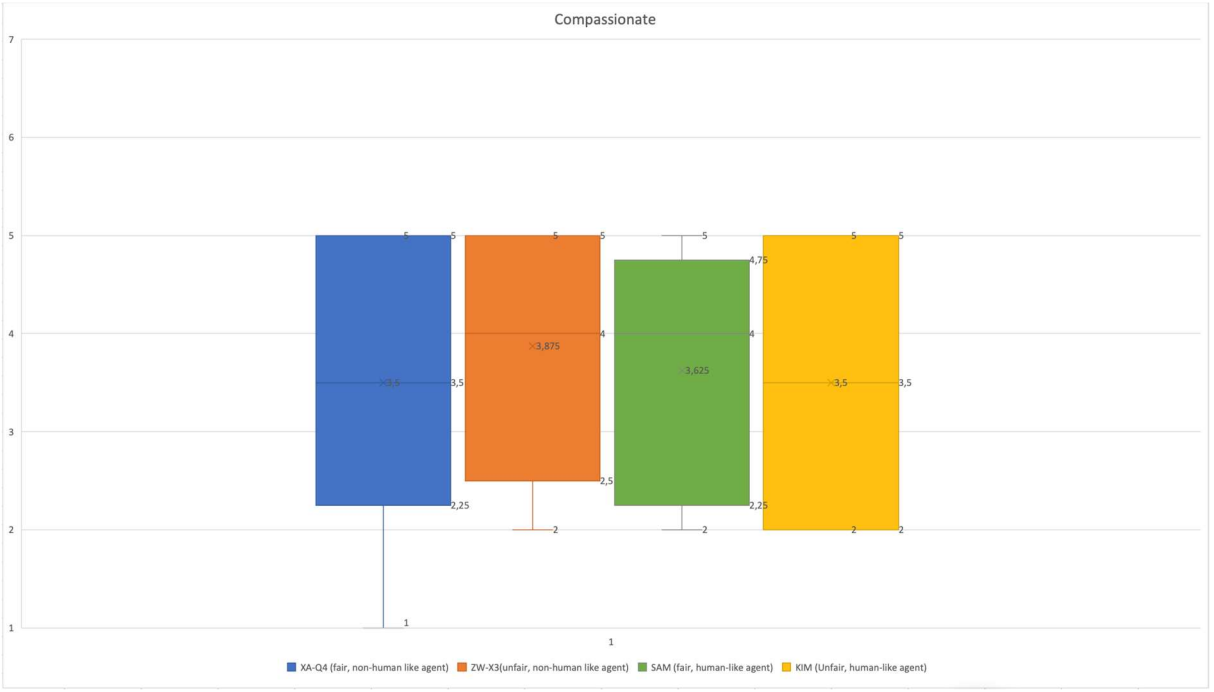


Figure A8. Compassionate (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non-humanlike agent, Green = Sam, fair humanlike agent, Yellow = Kim.

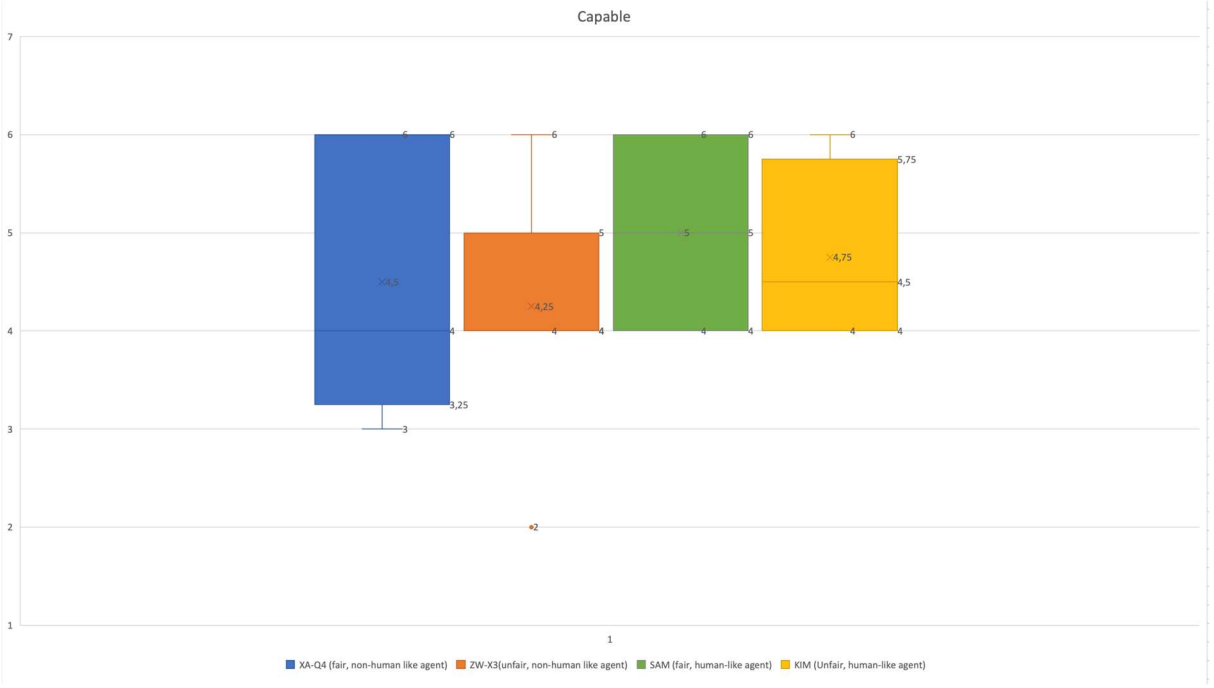


Figure A9. capable (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non humanlike agent, Green = Sam, fair humanlike agent, Yellow = Kim.

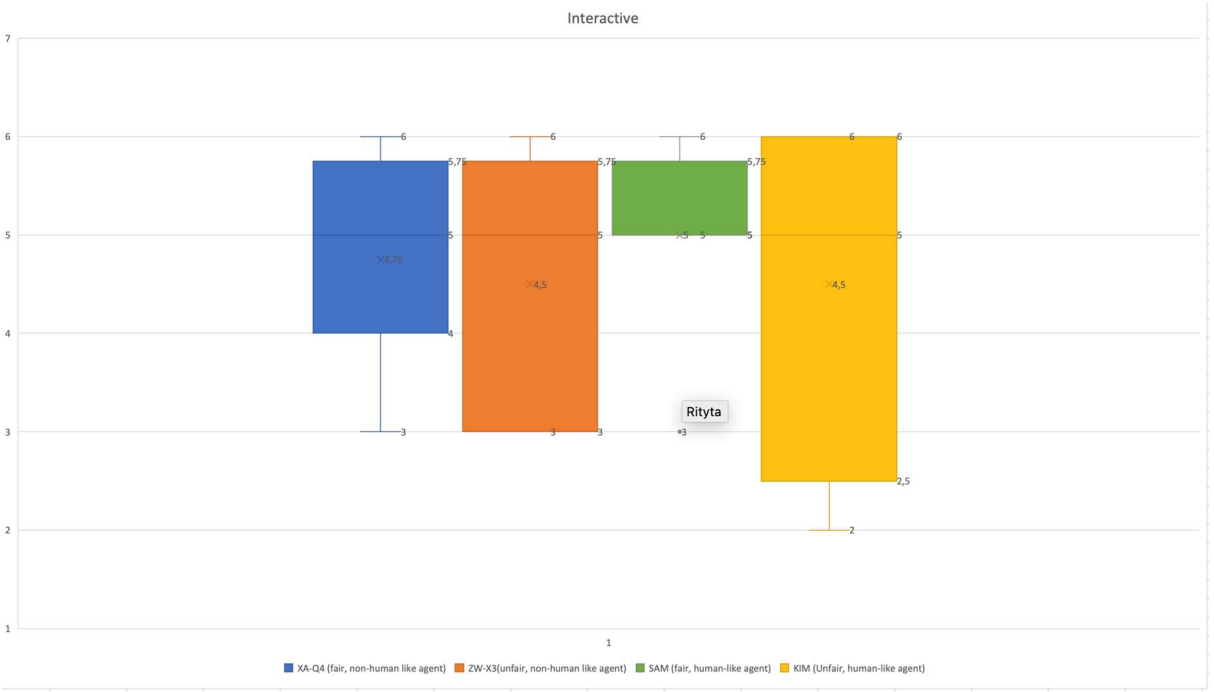


Figure A10. Interactive (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non-humanlike agent, Green = Sam, fair humanlike agent, Yellow = Kim.

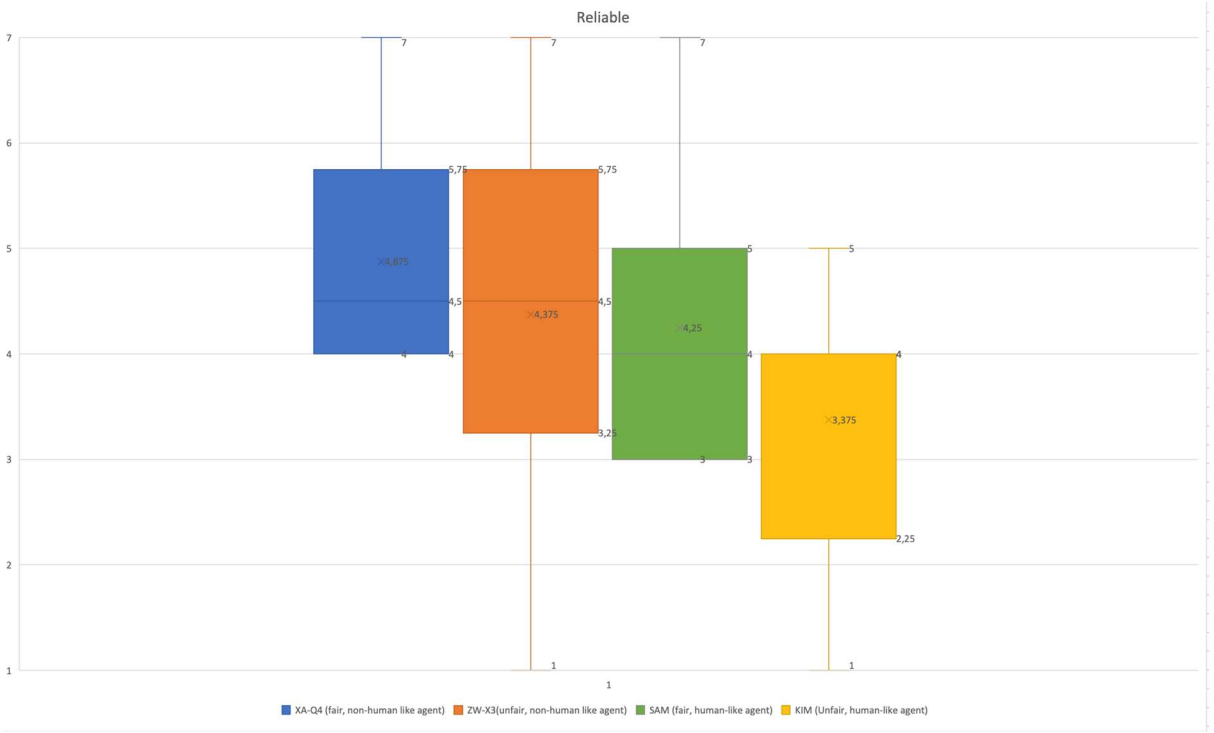
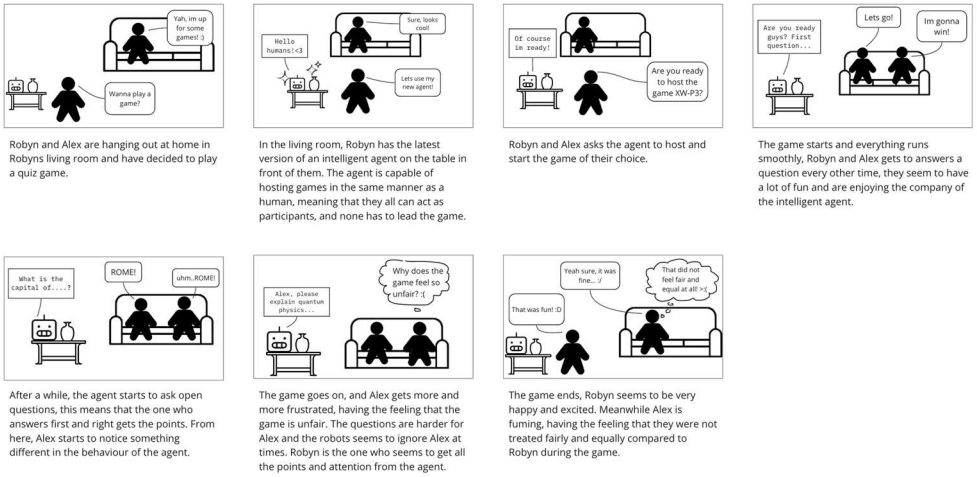


Figure A11. Reliable (Blue = XA-Q4, fair non-humanlike agent, Orange = ZW-X3, unfair non-humanlike agent, Green = Sam, fair humanlike agent, Yellow = Kim.

Appendix B – Focus group

Storyboard 1 - The experiment (Game scenario)



Robyn and Alex are hanging out at home in Robyns living room and have decided to play a quiz game.

In the living room, Robyn has the latest version of an intelligent agent on the table in front of them. The agent is capable of hosting games in the same manner as a human, meaning that they all can act as participants, and none has to lead the game.

Robyn and Alex asks the agent to host and start the game of their choice.

The game starts and everything runs smoothly, Robyn and Alex gets to answers a question every other time, they seem to have a lot of fun and are enjoying the company of the intelligent agent.

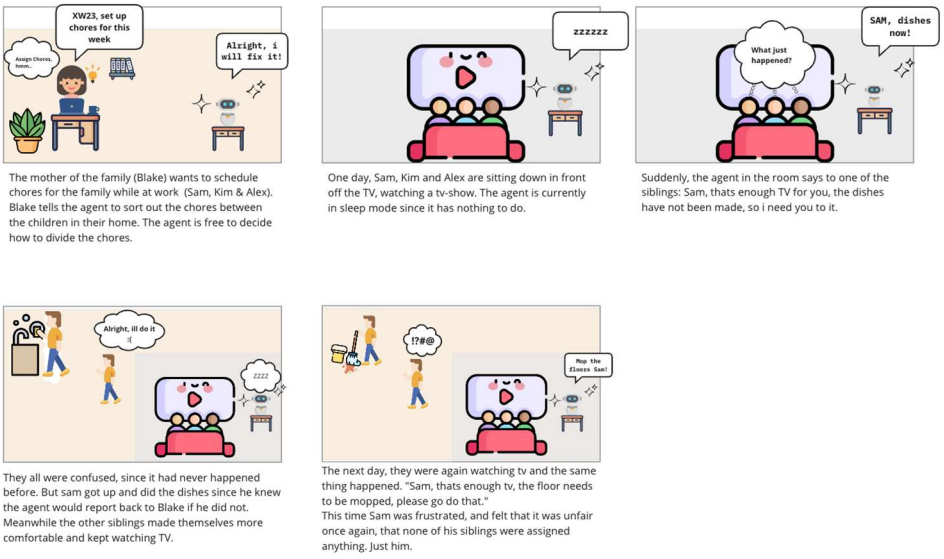
After a while, the agent starts to ask open questions, this means that the one who answers first and right gets the points. From here, Alex starts to notice something different in the behaviour of the agent.

The game goes on, and Alex gets more and more frustrated, having the feeling that the game is unfair. The questions are harder for Alex and the robots seems to ignore Alex at times. Robyn is the one who seems to get all the points and attention from the agent.

The game ends, Robyn seems to be very happy and excited. Meanwhile Alex is fuming, having the feeling that they were not treated fairly and equally compared to Robyn during the game.

Figure B1. Storyboard 1 -game scenario

Storyboard 2 - Chores in the house



The mother of the family (Blake) wants to schedule chores for the family while at work (Sam, Kim & Alex). Blake tells the agent to sort out the chores between the children in their home. The agent is free to decide how to divide the chores.

One day, Sam, Kim and Alex are sitting down in front off the TV, watching a tv-show. The agent is currently in sleep mode since it has nothing to do.

Suddenly, the agent in the room says to one of the siblings: Sam, thats enough TV for you, the dishes have not been made, so i need you to it.

They all were confused, since it had never happened before. But sam got up and did the dishes since he knew the agent would report back to Blake if he did not. Meanwhile the other siblings made themselves more comfortable and kept watching TV.

The next day, they were again watching tv and the same thing happened. "Sam, thats enough tv, the floor needs to be mopped, please go do that."
This time Sam was frustrated, and felt that it was unfair once again, that none of his siblings were assigned anything. Just him.

Figure B2. Storyboard 2 chores in the house

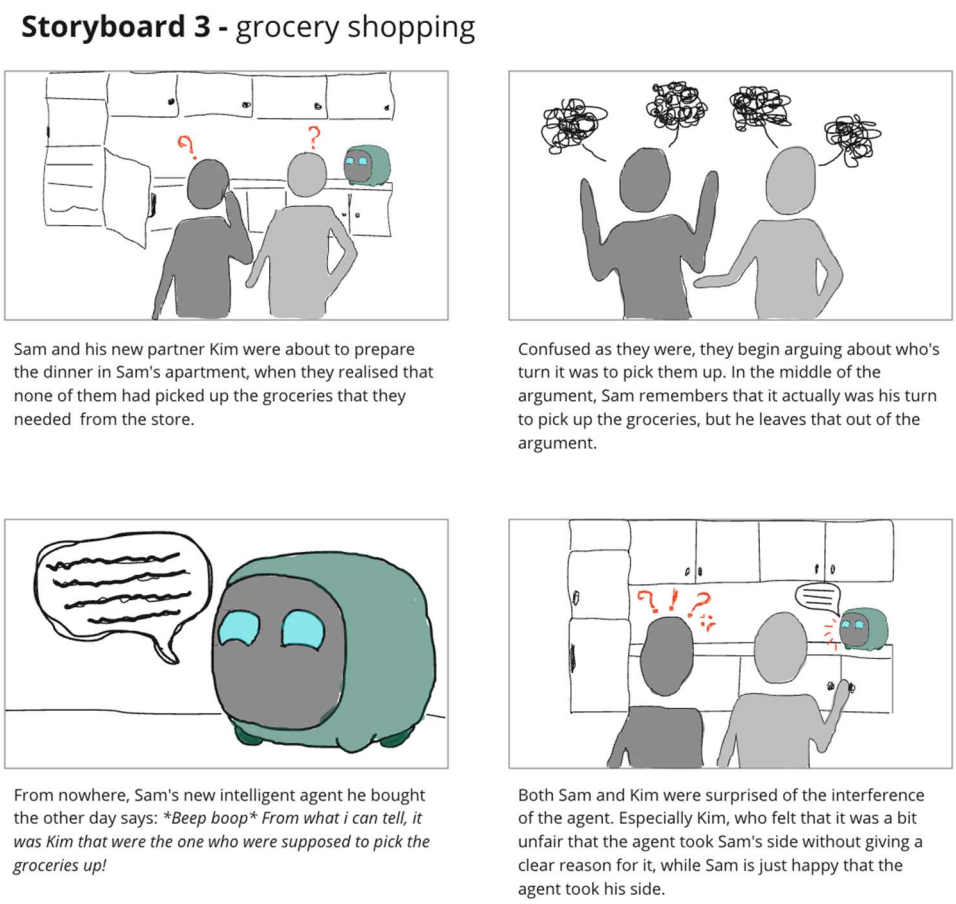


Figure B3. Storyboard 3 – grocery shopping

Appendix C – thematic map

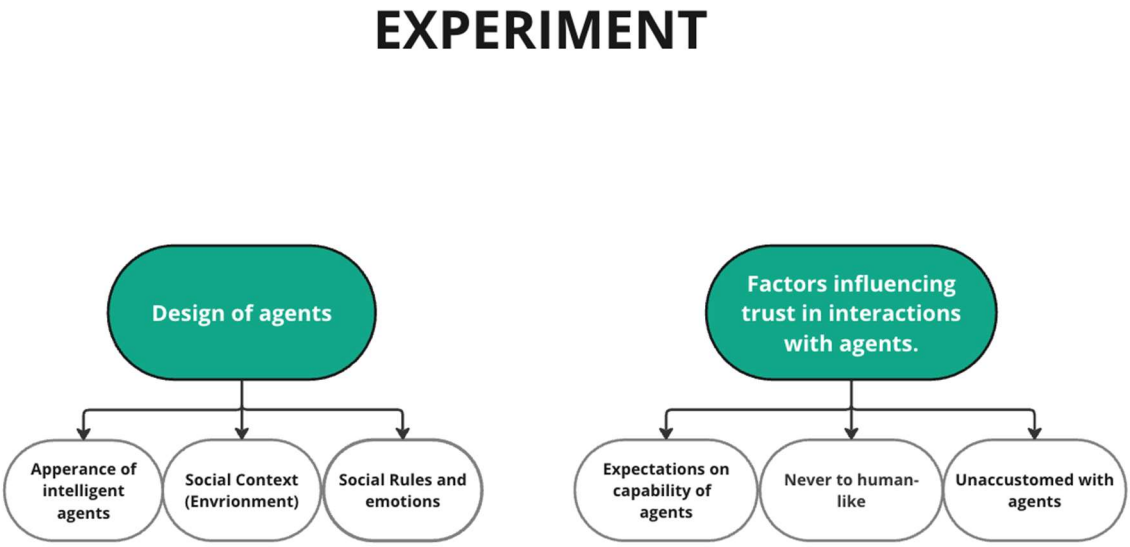


Figure C1. Thematic map from the experiment

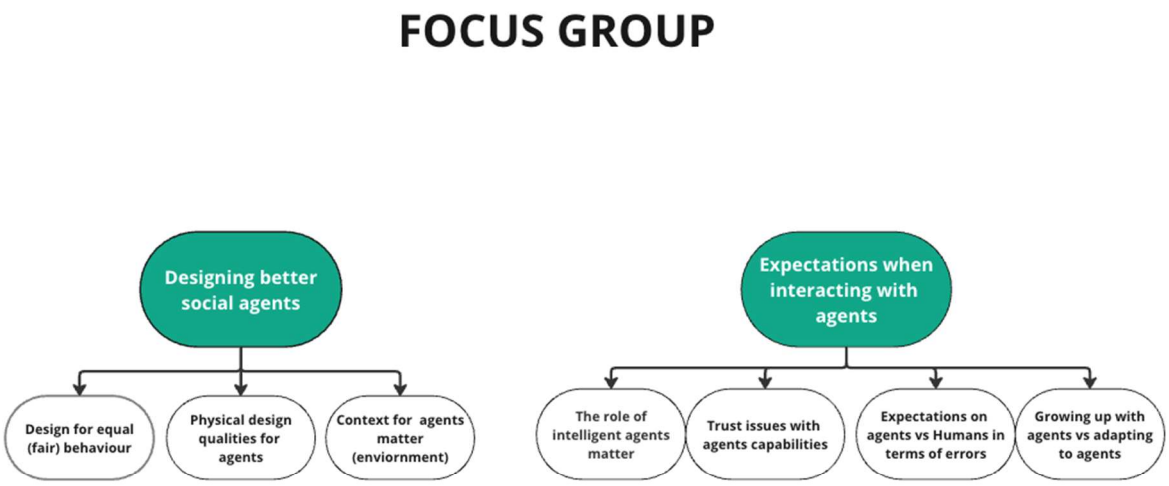


Figure C2. Thematic map from the focus group