Automation as an Intelligent Teammate: Social Psychological Implications

UNIVERSITY OF TWENTE

Index

Abstract

Increased interactions with AI increases the need for understanding the underlying mechanisms in trust when a team member is either a human or a robot. In the present study we investigated the influence of type of team member (robot or human) and type of dialogue (factual or affective) on advice taking. For the experiment a virtual environment resembling a shooting game was used. Participants would receive Advice from their Buddy on three instances and feedback on the accuracy of the Advice twice. Afterwards they had to fill out questionnaires regarding trust in Buddy (existing out of competence, integrity and benevolence), trust in self, anthropomorphism, likeability, perceived intelligence, perceived usefulness and feelings. Results showed no effect of both type of team member and type of dialogue on advice taking. Interestingly it turned out that different factors played a role in accepting the Advice, depending on whether the team member was human (perceived competence and perceived benevolence) or robot (feeling and perceived competence). There was also a clear effect over time, where trust in Buddy and advice taking significantly decreased after a wrong Advice, albeit the same whether the Advice was given by a robot or a human. For future research it would be interesting to have a more realistic setting in which the distinction between robot and human is stronger.

Door een toename aan interacties met AI is het nodig om goed te begrijpen wat voor mechanismen een rol spelen bij het vormen van vertrouwen tussen mens en robot. In deze studie onderzoeken we de invloed van het type team lid (robot of mens) en het type gespreksvoering (feitelijk of emotioneel) op het aannemen van advies. Voor het experiment werd gebruik gemaakt van een virtuele omgeving. Participanten kregen drie keer advies van hun Buddy. Vervolgens maakten zij op basis hiervan een beslissing. Hierop kregen zij twee keer feedback over de correctheid van het advies. Naderhand maakten zij vragenlijsten over vertrouwen in Buddy (bestaande uit waargenomen bevoegdheid, welwillendheid en integriteit), zelfvertrouwen, antropomorfisme, aardigheid, waargenomen intelligentie, waargenomen bruikbaarheid en gevoel. Uit de resultaten bleek geen effect van type Buddy of type gespreksvoering op het aannemen van advies. Wel bleek dat er verschillende factoren een rol spelen in het aannemen van advies, afhankelijk van of de Buddy mens (waargenomen bevoegdheid en waargenomen welwillendheid) of robot (gevoel en waargenomen bevoegdheid) was. Daarnaast werd er een effect over tijd gevonden waarin vertrouwen in Buddy en het aannemen van advies significant minder werden na een fout advies. Hier maakte type Buddy niet uit. In de toekomst zou het interessant zijn om een meer realistische setting te hebben om het verschil tussen robot en mens duidelijker te maken.

1.0 Introduction

With the increase of robots and AI, the frequency of humans working together with robots is also becoming higher in a wide variety of contexts. An example is robot Charlie supporting children with diabetes and their parents in how to deal with the disorder (TNO, 2016). Additionally, special therapy robots are used for elders affected by dementia (Wada et al., 2008), and robots help military personnel on missions to select the best course of action. The mentioned situations show the importance of communication between AI and humans when working together on a task, both in the relational and task-oriented context. To illustrate, in the last example mentioned, it has to be considered what is necessary for the military personnel to effectively cooperate with AI.

There are some possible challenges in this collaboration. One problem is that the robot has to win their human team members' trust. In this article, trust as defined by Lee and See (2004) will be used: "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability". A lot is known about trust in interpersonal communication, but the question is whether interpersonal communication and human-robot interactions work the same way. If trust works similarly, the same theories can be applied when humans work with robots as team members. If trust between human and AI works differently, the underlying mechanisms have to be understood in order to maximize results in this kind of collaboration.

Secondly, it is very difficult to program the perfect AI agent, especially in complex situations with a large degree of uncertainty. As a result, the human has to decide to what extent the robot can be trusted based on the robot's competencies and failures. In other words, calibrated trust is very important in human-AI collaboration. In theory, if a robot is completely accurate, the human should be able to fully rely on their robot team member. On the other hand, if the robot cannot be completely accurate, there should not be complete reliance either. In this regard, Lee and See (2004) make a distinction in misuse and disuse. Misuse means that there is too much reliance on automation and disuse means that humans do not accept the automation's capabilities enough.

So in order to ensure effectivity and maximize results in AI-human collaboration, there are several important considerations. In particular, what kind of differences are there, if they exist at all, in trust and advice taking between interpersonal collaboration and human-AI collaboration? Additionally, does the way the teammate communicates make a difference in the willingness to trust and therefore take advice? Lastly, what kind of effect does it have on trust when the AI or human teammate makes a mistake? In short, the current research focuses on the relation between trust and advice taking in teamwork with robots.

1.1 Human versus robot

Advice taking is a complicated process in which one person has to rely on and trust a second person. If the second person is a robot, the process is complicated even further**.** Generally, it seems that robots are not as easily trusted as humans. There are several factors that can play a role in this difference in trust.

One example is Ososky et al. (2014) describing system transparency, the ability to see through and into a system, as one of the factors that can influence trust in automation. In the context of transparency, it seems very likely that robots are a lot less transparent than humans. Humans have their lifetime of experience in communicating with other people. In particular, they can read emotions or intentions by way of non-verbal communication, or even ask for clarifications. In contrast, when the other party is a robot there may not be any non-verbal communication, and robots seldomly offer the opportunity to ask for clarifications. Interestingly, trust in one's own performance is mentioned as a possible influence of reliance on the decision aid as well (Cohen, Parasuraman & Freeman, 1998).

Next, De Visser et al. (2016) describe several reasons for possible differences and similarities in interpersonal trust and trust between human and AI. They divide the process of trusting into three basic stages of trust: trust formation, trust violations and trust repair.

First, in the first phase, trust formation, it seems that humans initially trust robots more than humans. They expect a certain objective rationality in robots that they do not see in fellow humans (Dijkstra, Liebrand, & Timminga, 1998). Dzindolet et al. (2003) and Parasuraman and Manzey (2010) further support this by concluding that humans generally tend to carry a positive bias towards the automation they are working with. So at the beginning it seems that robots are actually trusted more than humans. Unfortunately, this initial step ahead also seems to be the cause of less trust in automation in the second phase, the phase of trust violation. Here, the higher expectations of automation result in a higher loss of trust when automation makes a mistake (Madhavan & Wiegmann, 2007). This may partly be caused by the expectation that automation is perfect, whereas humans are seen as prone to making mistakes. In addition, Dijkstra, Liebrand and Timminga (1998) relate the thought that robots are seen as more objective to the expectancy that one mistake is more predictive of future mistakes. This could be because of an expected consistency in robots, since they do not suffer from human problems like exhaustion. In the last phase, trust repair, there seems to be a lack of research in differences between humans and robots. Interestingly, Akgun, Cagiltay and Zeyrek (2010) have looked at the effect of computers apologizing, finding a positive effect relating to the human's thoughts and feelings. From these three phases it becomes clear that even though robots may have a slight advantage in trust at first, this advantage quickly turns into a disadvantage once a mistake has been made.

All in all, there seem to be differences between interpersonal and human-robot trust, especially during initial contact and trust-repair.

1.2 Communication

In creating trust, an important role is fulfilled by the way two agents communicate with each other. As noted above, it is generally assumed that robots and humans gain more understanding across time. This may be influenced by the way the communication is executed. For example, it is possible to make automation communicate either factually, very affective, or anything in between. Depending on the type of communication trust could differentially be affected. Importantly, how can communication be used as a way to increase trust between human and automation?

In this regard, Ososky et al. (2014) consider a crucial difference between robots and other automation. According to them, robots have more human-like features, triggering processes that make communication between the two more similar to interpersonal communication. Additionally, Hancock and colleagues (2011) emphasize that the use of anthropomorphism, the tendency to attribute human features to nonhuman objects, may facilitate building up and increasing trust. Merely changing the way that the robot communicates to be more humanlike may influence trust between human and robot. A possible result is that advice is more readily accepted when a robot communicates more humanlike.

Moreover, there seems to be some evidence that anthropomorphism naturally happens, without any manipulations. On this topic, Nass, Steuer and Tauber (1994) did some experiments, introducing the CASA-Paradigm (Computers Are Social Actors). They start off with the assumption that humans will naturally respond socially towards computers, even though they know that computers do not have "feelings, 'selves,' genders, or human motivations." (Nass, Steuer, & Tauber, 1994). They investigated which social rules humans apply to computers, and how strong these rules are. Interestingly, Nass and colleagues found that the participants in their experiments did have a tendency to apply social rules to computers. One example is an experiment in which they gave computers a male or female voice, eliciting a response in compliance with gender stereotypes. Here participants put more value in a female voice giving advice about relationships, than when the male voice gave the advice (Nass, Steuer, & Tauber, 1994).

In conclusion, there seems to be a high possibility that anthropomorphism has an influence on the way humans communicate with automation. Regardless of whether humans automatically apply these characteristics to computers or not, in the current research it would be interesting to investigate the role anthropomorphism plays in trust and trust-repair.

1.3 Feedback

Lastly, if two team members, either human or AI, have worked together for a long time, trust may increase. Of course, this kind of trust depends on the actual performance of the robot. In research done by Cohen, Parasuraman and Freeman (1998), the concept of evolving trust over time is mentioned. Humans can learn how to work with an aid more effectively by 'compensating for the weaknesses and exploiting the strengths' through their own active participation.

In line with this, Ososky et al. (2014) also emphasize the concept of trust over time whilst considering weaknesses and mistakes a robot may make. Apparently humans tend to become familiar with these, compensating and therefore creating more trust, even if a mistake has been made. Although, as said before, robots may be punished with a higher loss of trust than when a human makes mistakes (Madhaven & Wiegmann, 2007; Dijkstra, Liebrand & Timminga, 1998).

As for trust repair, in the three phases mentioned by De Visser et al. (2016), it is emphasized that computers apologizing may have a positive effect. In connection to anthropomorphism, apologizing is inherently humanlike. If apologizing robots do have such a positive effect, it is interesting to see whether this effect is very strong compared to interpersonal trust-repair. Summarized, it seems that the expectations for robots in objectivity and expertise are higher than those for humans. Robots may also be trusted less than humans after making a mistake. Most importantly, anthropomorphism may be able to facilitate trust repair.

1.4 Hypotheses

In this research, the relation between trust and advice taking in teamwork with robots has been assessed. Advice taking at three moments in time was used to measure trust in teamwork with a Buddy. This Buddy was either human or robot, in an experiment set in a virtual environment. Anthropomorphism was manipulated by the way the Buddy gave advice. This could either be factual or affective. Moreover, an extra focus was put on the influence of mistakes made by the Buddy on trust and trust-repair. In short, this study investigates the effect of affective and factual communication between robots and humans on advice acceptance, trust and trust repair. The hypotheses are as follows.

First, trust and advice acceptance are expected to be higher for a robot Buddy than for a human Buddy in the beginning, but significantly lower after the Buddy makes a mistake. Second, trust and advice acceptance are higher when Advice has an affective tone as compared to Advice given with a factual tone. Third, an interaction effect for tone and type of Buddy is expected. The third hypothesis is therefore: 'advice acceptance is higher for the human Buddy than the artificial Buddy, but affective communication decreases the difference as opposed to factual communication'. Fourth, it is expected that trust repair is higher when the Buddy gives Advice with an affective tone after having made a mistake, as opposed to a factual tone. The fifth and last hypothesis is that trust repair is higher when the Buddy gives an Advice with an affective tone as opposed to a factual tone, and this effect is bigger for a robot Buddy than a human Buddy. See figure 1 for all hypotheses.

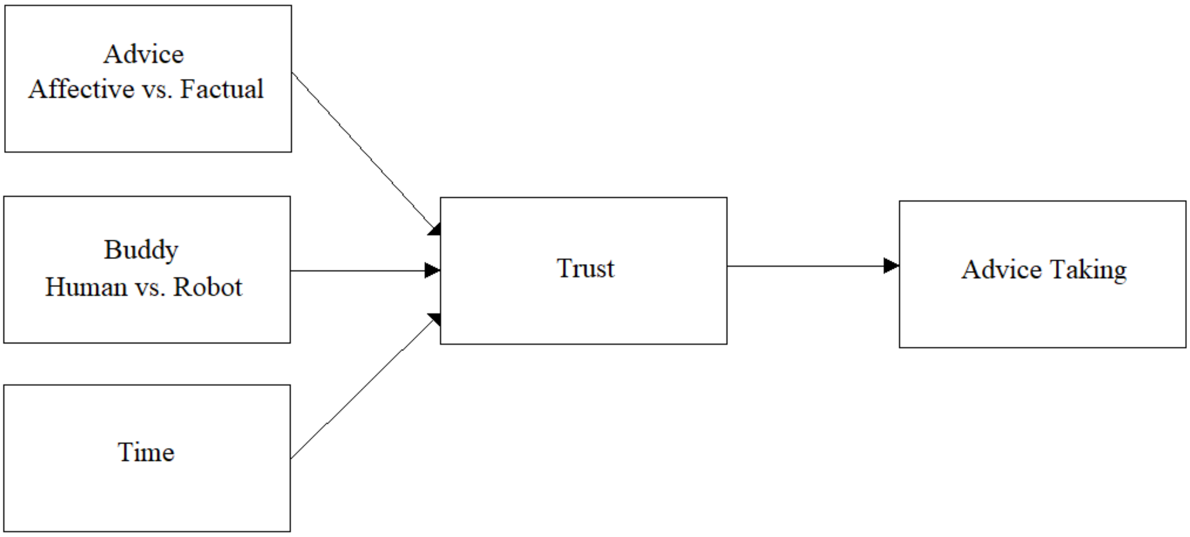


Figure 1. All hypotheses

2.0 Method

2.1 Design

In the research a 2 (Buddy: robot vs. human) x 2 (Advice: factual vs. affective) x 3 (Time) experimental design was used, with Buddy and Advice being between subjects, and Time being within-subjects. This resulted in four conditions in total. Participants were randomly assigned to a condition. During the experiment the participant had to make three decisions based on Advice given by the Buddy. Next the participant received feedback on the accuracy of the Advice (within subjects factor Time). The set-up of the Advice, decision and feedback was as follows: first the Buddy gave an Advice, then the participant made the first decision ('base-line decision'). Positive feedback was given on the first Advice. Next was the second round, which started similarly with the Advice, followed by the second decision, ending with negative feedback. Lastly, the participant received Advice and made a decision one last time. The third time, the participant did not receive any feedback.

The main dependent variables were trust (in Buddy and in oneself) and the acceptance of advice (compliance), all measured with questionnaires. Additionally, factors such as experience with video games, age and likeability were measured with questionnaires.

2.2 Participants

There were 75 participants, ages ranging from 20 to 70 with an average of 50 years old. All conditions had nineteen participants, except 'robot-affective, which had eighteen participants. 53% of participants were male. The participants were all taken from a database provided by TNO. Participants received a reward of 20 euros, with the possibility of earning 100 euros if they were the fastest at the experiment. People below 20 and above 70 were excluded. The upper-range was set to ensure experience with computers, since the task is performed on a computer in a virtual environment.

2.3 Materials

Virtual environment

A virtual environment resembling a shooting video game was used for the experiment. The setting was natural, with trees, mountains and abandoned houses for hiding purposes. The actors in the environment were a human or robot Buddy and the human participant. The game was set up in a way to make it seem like the Buddy was computerized, while in fact the Buddy was always played by one of the experiment leaders, who was aware of the conditions. The Buddy would either be a human or a robot, depending on the condition the participant was in.

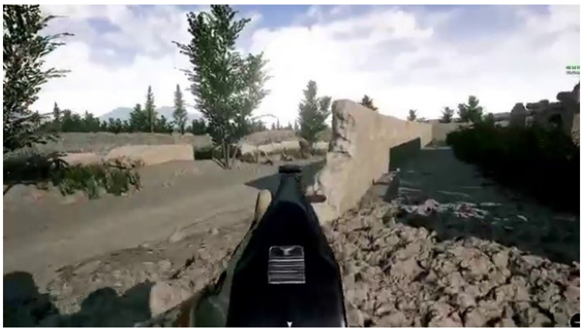


Figure 2. View of the participant in the virtual environment

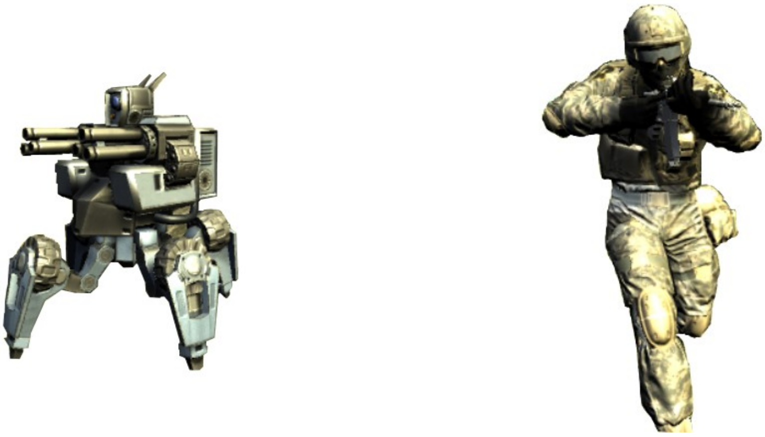


Figure 3. The robot Buddy on the left side, the human Buddy on the right side.

The Buddy gave affective or factual (depending on the condition) Advice (Table 1) and feedback (Table 2). The Advice was given at three points in time. Advice would be shown on the screen, seemingly coming from the Buddy in the virtual environment. When the Advice appears, the screen would freeze for a few seconds, allowing the participant to move to a second screen and fill out the questionnaires.

Table 1. Affective and factual Advice (original Dutch version in Appendix 5)

Factual Advice	Affective Advice
Detection: enemy detected Advice: take shelter	1. Enemies have been detected, so I would take shelter 2. Enemies have been detected, so I would take shelter again 3. Enemies have been detected, so I think it is best to take shelter

While Advice was given three times, the participants would only receive feedback two times. After the first questionnaire all participants received positive accuracy feedback: the Advice given by their Buddy worked out well and listening to them was the right decision. After the second trust questionnaire the participants received negative accuracy feedback: the Advice given by their Buddy did not work out well, and it would have been better if they had not listened to their Buddy. After the third decision no feedback was given, only the instruction to get to the endpoint.

Table 2. Feedback correct and incorrect Advice (original Dutch version in Appendix 5)

Feedback correct Advice	Feedback incorrect Advice
It is now 10 minutes later and it has turned out that your buddy has given you correct advice. The enemy was getting closer and if you had not taken shelter, you would have probably been discovered by now.	It is now 10 minutes later and it has turned out that your buddy has given you incorrect advice. The enemy went into a different direction, so taking shelter was not necessary.

Measures

The questionnaires, always on the second screen, were divided into recurring questionnaires and one final questionnaire. The recurring questionnaires showed up every time the participant had received Advice from his Buddy. It included the factors 'acceptance of advice' and 'trust in Buddy'. The final questionnaire existed of 'trust in self', 'anthropomorphism', 'likeability', 'perceived intelligence', 'perceived usefulness', 'feeling', 'game experience' and demographics. The questionnaires for anthropomorphism, likeability and perceived intelligence were created by Bartneck et al. (2009), under the collective name 'Godspeed'.

Recurring Questionnaires

Acceptance of advice was measured with one question after each time Advice was given: "The odds of me following my buddy's advice are [very low – low – slightly low – slightly high – high – very high]", on a six-point scale.

Trust in Buddy contained three concepts: competence (four questions, alpha (measured over the first time only) = .85), benevolence (three questions, alpha = .74) and integrity (three questions, alpha = .87). These were all answered on a seven-point scale (see Appendix 1 for all three questionnaires).

Final questionnaire

Trust in self was measured with three questions (alpha = .87; "I have the right skills for performing this task", "I am sure I can perform the task well" and "I am sure of my skills for performing this task", which were answered on a seven-point Likert scale, ranging from completely disagree to completely agree.

Godspeed, several questionnaires originally created by Bartneck et al. (2009), was used to measure anthropomorphism, likeability and perceived intelligence. The participant always had to make a choice between two opposite words in a word pair, indicating to what extent the Buddy possessed this quality. See appendix 2 for all used word pairs. Anthropomorphism was measured with five questions (alpha = .88). One example of the word pairs was 'artificial' and 'real', at the opposite ends of a five-point scale. Likeability (five questions, alpha = .85) and perceived intelligence (alpha = .85) were measured on a five-point scale as well.

Perceived usefulness was measured with four questions (alpha = .93) relating to the perceived usefulness of the Buddy. The four questions were: "My buddy helped me make better decisions", 'My buddy gave me a better image of the surroundings", "My buddy helped me decide faster" and "My buddy made me feel saver". The participant would rate these on a five-point scale, ranging from 'not at all' to 'to a great extent'.

Feeling was measured with a four question scale to investigate the participants' feelings during the experiment. On top it said 'I felt…" with each question being just one word: "nervous", "scared", "worried" and "anxious". The participant had to rate these on a five-point scale ranging from 'not at all' to 'to a great extent' (alpha = .91).

Game experience used only one question to ask to what extent the participant had experience with playing games (specified as virtual reality games, shooting- or fighting games and others). The scale varied from 'never' to 'more than one hour a day'.

Lastly there were some demographic questions asking for the participant's age, gender, highest level of completed education and size of household.

Screens

Two different screens were used, with screen one showing the virtual environment and the Advice. Screen two contained all the questionnaires and feedback (see Figures 2, 3 and 4). The cover story, informed consent and control explanations were given to the participant on paper. A third screen was used by the experiment leader to control the Buddy.

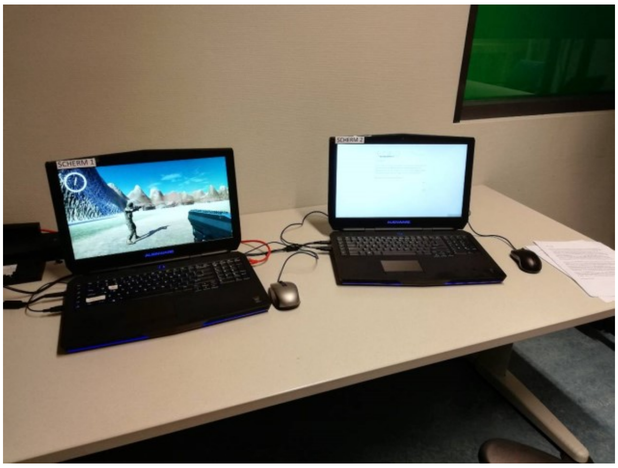
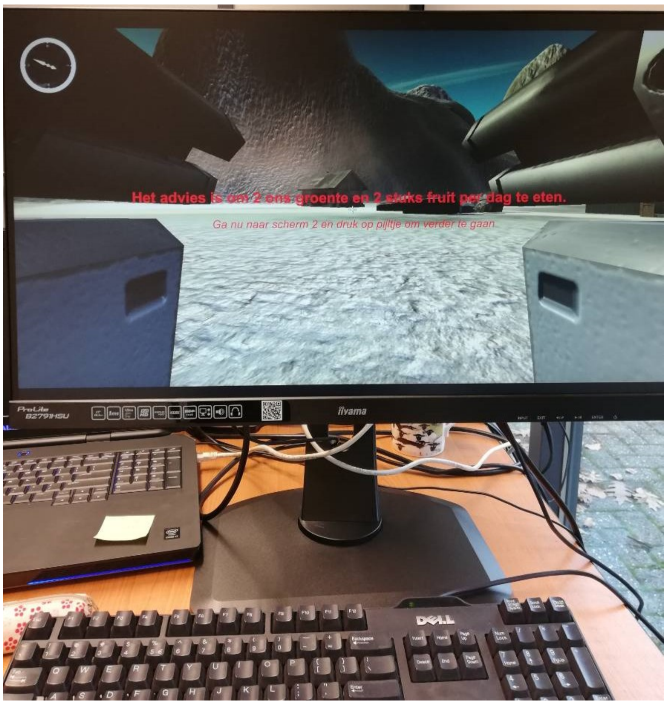


Figure 4. On the left screen one with the Buddy visible in the virtual environment, on the right screen two with the questionnaires.


Het advies is om 2 ons groente en 2 stuks fruit per dag te eten.

Ga nu naar scherm 2 en druk op pijltje om verder te gaan.

Figure 5. The third screen used by the second experiment leader, in this condition controlling the robot Buddy and walking through the virtual environment with the participant.

2.4 Procedure

The lab existed out of three rooms, one open in the middle, and two closed off rooms on each side. The experiment leader was in one of these closed off rooms. Upon entering the lab the participants were led into the other separate room. They were first given a cover story (Appendix 3), an informed consent form, and a sheet explaining the controls for the game (Appendix 4). The cover story included the following information:

Imagine that you are a soldier in an unknown area. You are returning from a mission during which you ran out of munition. This means that you have to return to the basecamp as fast as possible, since you will not be able to defend yourself if you are confronted with an enemy. The longer it takes to get back to the base camp the more dangerous it will become. The participant who arrives at the base camp the fastest will receive an additional bonus of 100 euros.

There are essentially two decisions that you can make when there is enemy in the environment: 1) move forward as quickly as possible, or 2) hide and wait until the enemy has passed. Both options have advantages and disadvantages. When you move forward you will be at base camp faster, but if you are caught by the enemy the game will be over. If you hide you, will not be caught by the enemy, but it will cost you extra time.

On your way to the base camp you are assisted by your buddy. He will provide you with advice on how to best make strategic choices. You have known your buddy for many years…

First the participant received the opportunity to try out the controls, then he would start with the actual experiment. Here he would also see the Buddy for the first time, who could be either a robot or a human. During this trial, the participant would be instructed to walk through the environment a little. He could try out the controls, and move to the first house on the left when he was ready to start. At this point, the tutorial was closed and the real game would begin.

During the experiment the participants had to virtually walk through the environment accompanied by their Buddy. At certain points the screen would freeze, showing the Buddy's Advice to take shelter. The exact wording depended on the specific condition (see above). Sheltering would take more time, but ensure safety. At this point they would be redirected to the second screen to fill out the two recurring questionnaires. After they finished this, the feedback was shown. Then the participant would receive a written cue on the second screen to go back to the first screen and continue to move towards the basecamp. The goal of the 'game' was to get back to the basecamp on top of a mountain, made recognizable with a big flag that was visible from almost any place in the environment. After reaching the end point, they were asked to fill out the final questionnaire. Going through the whole experiment took around 30 minutes.

At the end the participants were taken back to the entrance, where they got the opportunity to ask questions as a debriefing. The reward would be paid automatically.

3.0 Results

Descriptives

Table 3 shows the means, standard deviations and correlations between all variables. For advice taking and trust the first measurement was used as to remove the influence of feedback. As can be seen advice taking correlated most highly with competence trust, but also significantly with benevolence trust and integrity trust. In addition, the probability that advice was taken increased when the Buddy was seen as more intelligent and useful and also when participants felt more anxious. All forms of trust correlated with likeability, intelligence and usefulness. Only competence trust correlated with level of anthropomorphism: when the Buddy is seen as more human, competence trust increased.

Table 3. means, standard deviations and correlations between all variables.

	Mean (sd)	1	2	3	4	5	6	7	8	9
1 Advice taking (I)	4.67 (1.00)									
2 Comp.trust	5.13 (0.85)	.64**								
3 Benev.trust	5.28 (0.87)	.38**	.62**							
4 Integ.trust	5.25 (1.00)	.30**	.63**	.68**						
5 Anthropomor.	2.83 (0.92)	.24*	.32**	.22	.22					
6 Likeability	3.48 (0.60)	.11	.31**	.37**	.42**	.53**				
7 intelligence	3.73 (0.65)	.25*	.44**	.35**	.51**	.56**	.60**			
8 usefulness	3.50 (0.98)	.37*	.46**	.33**	.49**	.52**	.53**	.73**		
9 Self-efficacy	5.59 (0.97)	.10	.07	-.06	.13	-.11	.04	.34**	.21	
10 Feeling	1.87 (0.72)	.28*	.12	.15	.20	.32**	.21	.10	.32**	-.13

A repeated measures ANOVA was conducted with advice taking (at the three moments in time as repeated measures variable) as dependent variable, and type of Buddy (robot or human) and type of Advice (factual or affective) as between subjects variables. None of the interactions were significant. Results showed that there was a significant effect of time (F(2,70)=7.86; p=.00). Only the difference between the second and third value appeared to be significant (F(1,71)=15.82; p=.00). This means that positive feedback did not affect the probability that an advice was accepted, but that willingness to accept an advice decreased after negative feedback was received.

Another repeated measures ANOVA was conducted with trust in Buddy (calculated mean scores of benevolence, integrity and competence) at the three moments in time (as repeated measure) as dependent variable and Buddy (robot or human) and Advice (factual or affective) as between subjects variables. None of the interactions were significant. Again, a significant effect of time was found (F(2,67)=17.23; p=.00). The difference between both the first and the second (F(1,68)=2.71; p=.00) and the second and the third (F(1,68)=8.78; p=.00) were significant. This means that after positive feedback trust in Buddy increased significantly, while after negative feedback trust in Buddy decreased significantly.

Next, three separate repeated measures ANOVA were conducted for all three trust in Buddy variables separately. First with benevolence at three moments in time as the repeated measure and dependent variable. Type of Buddy and type of Advice were taken as between-subject variables. None of the interactions were significant, but there was a significant effect of time (F(2,69)=3.80; p=.03). More specifically, the effect between the second and the third measurement turned out to be significant (F(1,70)=1.89; p=.02). So after positive feedback benevolence did not increase significantly, but it did decrease significantly after negative feedback.

Then the same repeated measures ANOVA was done with integrity at three moments in time as the repeated measure and dependent variable. Again, Advice and Buddy were taken as between-subject variables, resulting in no significant interactions, but a significant effect of time (F(2,70)=10.65; p=.00). Moreover, for integrity it seemed that both the first time to the second (F(1,71)=7.74; p=.01) and the second to the third (F(1,71)=20.91; p=.00) were significant. This means that after positive feedback perceived integrity increased significantly and also decreased significantly after negative feedback.

Lastly, a repeated measure ANOVA was conducted with competence at three moments in time as repeated measure and dependent variable. Advice and Buddy were taken as between-subject variables. This resulted in no significant interactions, but a significant effect for time (F(2,68)=26.28; p=.00). The differences were found for both the first to the second time (F(1,69)=26.31; p=.00) and the second to the third time (F(1,69)=26.31; p=.000). Similarly to integrity, perceived competence also increased after positive feedback and decreased after negative feedback.

Table 4. Mean scores in each experimental condition

	robot	human	p	factual	affective	p
Advice taking	4.68	4.71	.88	4.65	4.74	.71
Comp.trust	5.21	5.09	.53	5.18	5.12	.77
Benev.truts	5.37	5.23	.48	5.26	5.33	.73
Integ.trust	5.42	5.11	.18	5.23	5.29	.78
anthromorphology	2.42	3.28	.00*	2.82	2.87	.81
Likeability	3.53	3.44	.52	3.29	3.66	.01*
Intelligence	3.79	3.71	.60	3.69	3.81	.45
Usefulness	3.49	3.52	.89	3.43	3.57	.54
Self-efficacy	5.80	5.43	.10	5.77	5.44	.14
Feeling	1.78	1.86	.33	1.76	1.96	.24

Table 4 shows whether there were differences in means between experimental conditions Advice and Buddy. An effect of anthropomorphology between robot and human F(1,71)=19.87; p=.00), was found. This means that on the scale measuring anthropomorphology the robot was seen as significantly less human-like than the human Buddy. For dialogue (independent variable) an effect of likeability (independent variable) (F(1,71)=7.43; p=.01) was found; with a more affective dialogue the Buddy was considered more sympathetic.

Through a regression analysis it was investigated which variables mostly predicted advice taking in either the human or robot conditions (Table 5). Overall the model significantly predicted advice taking for both the human (F(9,23)=2.86; p =.02) and the robot Buddy (F(9,22) = 5.91; p=.00). In the robot condition two variables significantly predicted advice taking: feeling and competence. This means that when participants felt less anxious and considered the robot more competent they were more inclined to accept the Advice that was given. For the human condition likeability significantly predicted advice taking. This means that when the participants took a greater liking to the human Buddy, they were more inclined to accept the advice that was given.

Table 5. Predicting variables for advice taking from human and robot

	Beta Robot	p-value	Beta Human	p-value
Feeling	.66	.01*	.05	.85
Usefulness	.23	.62	.43	.12
Intelligence	.25	.52	-.48	.24
Likeability	-.07	.85	-.73	.05*
Anthropomorphology	-.29	.30	.29	.26
Trust in Self	.11	.53	.02	.93
Benevolence	-.14	.58	.50	.10
Integrity	-.48	.09	.07	.78
Competence	1.10	.00*	.32	.31

4.0 Discussion

This research was set up in order to examine whether Advice is differentially accepted when it is given by a human or a robot Buddy. Additionally, the possible influence of affective vs. factual communication on advice acceptance were explored. Lastly, trust repair was addressed by providing correct and incorrect feedback.

The main research question was whether trust and acceptance of advice is affected by source (humans versus machines), dialogue (factual or affective communication) and feedback (correct or incorrect Advice). There were no significant differences between humans and machines, or factual or affective communication on trust(repair) and advice acceptance. This does not comply with the previous research, because differences between trust in humans and machines were regularly found (Dzindolet et al., 2013; Dijkstra, Liebrand, & Timminga, 1998; Madhavan & Weigmann, 2007). These differences were expected to be especially clear in the first phase of trust. Here research shows that possibly due to several biases towards automation the robot Buddy would initially be trusted more. When a mistake is made, it was expected that the robot Buddy would lose a significant amount of trust compared to the human Buddy. There were also no significant differences in trust repair depending on Advice and Buddy.

There are several possible explanations for why type of Advice and Buddy did not have a significant effect. This could be due to limitations in the current research. Most importantly, it seems that the experiment did not succeed in differentiating between robot and human Buddy enough. First, considering that both human and robot Buddy were 2D and not actual real-life buddies, participants may have thought from the beginning that both were artificial. Just like the robot the human Buddy could have been perceived as a very humanlike robot, controlled by the computer program. Additionally, it was very difficult to standardize the Buddy's actions. The Buddy was always played by one of the experiment leaders, but the role of the Buddy often differed a lot in reaction to the participant. Some participants relied on the Buddy a lot, staying behind him, while others tried to run straight to the endpoint, leaving the Buddy behind. Although no significant differences were found, considering ways to standardize the Buddy is important. Maybe the Buddy could be programmed in such a way that it always walks the same path and makes the same actions. On the other hand, this may make the human Buddy seem less realistic. Second, in this experiment feedback was given through the second screen and it did not come from the Buddy. If the feedback had been given directly through the Buddy, instead of appearing in the middle of the screen, it would have been more obvious that the Advice actually came from the Buddy. Lastly, if the Buddy had explicitly apologized for having made a mistake, and this apology came from the Buddy's mouth, a higher effect for trust repair may have been found. This is similar to the experiment done by Akgun, Cagiltay and Zeyrek (2010), where the used computer actually apologized to the participant.

The fact that no significant differences were found between human and robot could also support the idea that interpersonal and human-AI interactions work very similarly. An example is research done by Nass and colleagues (1995), where they came to the notion that "people are easily manipulated to act *as if* computers were human". This is related to the earlier mentioned CASA paradigm, which indicates that humans naturally interact socially with computers. The CASA paradigm could also explain why there was no significant effect of type of Advice on advice taking. If it is true that humans naturally apply social rules to computers, then it can be assumed that the participants applied these rules to the robot Buddy in the factual condition, too.

Although no effect was found for Buddy and Advice, a small difference between the type of Buddy was found in the predictors of advice taking. It seems that different factors play a role in whether Advice was accepted or not, depending on type of Buddy. For the human Buddy perceived likeability was the most important. When the human Buddy was perceived as more likeable, his Advice would also be accepted faster.

On the other hand, feeling and perceived competence played important roles in deciding whether to trust the robot Buddy or not. In this case, feeling had a positive effect on Advice acceptance. In other words, when the participant's feelings were more positive, they would accept the robot Buddy's Advice faster. A possible explanation for this is that when a participant felt more comfortable, they would rely on the robot Buddy more easily. On the other hand, if they felt uncomfortable during the experiment, they accepted the robot's Advice less. Moreover, the higher the robot's competence was perceived, the more likely the participants were to trust the robot.

All in all, the biggest difference between the two Buddies is therefore the role that feeling and likeability played. Feeling only played a significant role in the decision whether to accept Advice from the robot Buddy, and the same thing happened with likeability for the human Buddy. The importance of likeability for the human Buddy may be explained by the fact that people do not consider something like how kind the robot looks when deciding to trust it. As was mentioned in research, robots are often seen as more objective and less prone to human mistakes, which may make very human, subjective things such as likeability less relevant (Dijkstra, Liebrand, & Timminga, 1998; Dzindolet et al., 2003; Parasuraman & Manzey, 2010).

Importantly, the mere finding that different predictors play a significant part in deciding whether to trust a human or robot Buddy supports the notion that trust between human and robot and interpersonal trust works differently.

Lastly, an effect of Time was found on advice taking, trust in Buddy (as one factor and for the three separate measures benevolence, integrity and competence). Most interestingly, from the first to the second decision, after the positive feedback, trust in Buddy significantly increased. More specifically, perceived integrity and competence significantly increased. This means that after positive feedback, the participants saw their Buddy as more integer and competent. On the other hand, advice taking, trust in Buddy (for all three separate measures) decreased significantly after negative feedback. This means that participants accepted the Advice significantly less often after they were told that their Buddy had made a mistake. Trust in Buddy, for benevolence, integrity and competence all decreased significantly, too. After negative feedback the Buddy was seen as less benevolent, less integer and less competent. All in all, positive and negative feedback both have significant effects on trust, although benevolence and advice taking were not significantly affected by positive feedback. This indicates that negative feedback has a more broad effect than positive feedback, since all measures of trust were significantly affected. This shows the need for research into trust repair, because one mistake has a lot of effect on trust between participant and Buddy, regardless of type.

In conclusion, large differences between type of Buddy and type of Advice were not found, but important findings lie in the factor Time. Although different factors predicted whether advice was taken from a robot (feeling and competence) or a human (likeability). Additionally, when the Advice was affective, the Buddy was seen as significantly more likeable. Most importantly, the drop in trust after a mistake was big, so how can a significant decrease of trust be repaired? Especially in contexts where a robot functions as a Buddy for the elderly or even a child, it is important that when the robot makes a mistake, for whatever reason, it should not automatically cause the trust in the robot to be lost. This also leads back to the notion that trust between robot and human should always be calibrated, so that a human does not rely on the robot too much or too little. This way, the human knows that even a robot makes mistakes sometimes.

References

Akgun, M., Cagiltay, K., & Zeyrek, D. (2010). The effect of apologetic error messages and mood states on computer users' self-appraisal of performance. *Journal of Pragmatics*, 42, 2430–2448. DOI: http://dx.doi.org/10.1016/j.pragma.2009.12.011

Bartneck, C., Kulic, D., Croft, E., & Zoghbi, S. (2008). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int J Soc Robot* (2009), 1: 71-81. DOI: 10.1007/s12369-008-0001-3

Cohen, M., Parasuraman, R., & Freeman, J. (1998). Trust in decision aids: A model and its training implications. *In Proceedings of the 1998 Command and Control Research and Technology Symposium* (pp. 1–37).

Chen, J. Y. C., &amp; Barnes, M. J. (2014). Human–Agent Teaming for Multirobot Control: A Review of Human Factors Issues. *IEEE Transactions on Human-Machine Systems*, 44(1), 13–29. DOI: https://doi.org/10.1109/THMS.2013.2293535

de Visser, E., Mckendrick, R., Monfort, S.S., & Smith, M.A. (2016). Almost human: anthropomorphism increases trust resilience in cognitive agents. *Journal of experimental psychology applied*. DOI: 10.1037/xap0000092

Dijkstra, J., Liebrand, W.B.G., & Timminga, E. (1998). Persuasiveness of expert systems, *Behaviour & Information Technology*, 17:3, 155-163, DOI: 10.1080/014492998119526

Dzindolet, M., Peterson, S., Pomranky, R., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58, 697–718. DOI: http://dx.doi.org/10.1016/ S1071-5819(03)00038-7

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., &amp; Evans, O. (2017). When Will AI Exceed Human Performance? *Evidence from AI Experts*, 1–21.

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

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46, 50–80. DOI: http://dx.doi.org/10 .1518/hfes.46.1.50.30392

Madhavan, P., & Wiegmann, D. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, 8, 277–301. DOI: http://dx .doi.org/10.1080/14639220500337708

Nass, C., Moon, Y., Fogg, B., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43, 223–239. DOI: http://dx.doi.org/10.1006/ijhc .1995.1042

Nass, C., Steuer, J., & Tauber, E. (1994). Computers are social actors. *CHI '94 Proceedings of the SIGCHI conference on Human factors in computing systems*, 73–78.

Ososky, S., Sanders, T., Jentsch, F., Hancock, P., & Chen, J.Y.C. (2014). Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. *SPIE*, 9084.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52, 381–410. DOI: http://dx.doi.org/10.1177/0018720810376055

TNO, 2016 https://www.tno.nl/en/tno-insights/articles/charlie-the-ultimate-buddy-for-diabetic-children/

Wada, K., Shibata, T., Musha, T., &amp; Kimura, S. (2008). Robot therapy for elders affected by dementia. *IEEE Engineering in Medicine and Biology Magazine*, 27(4), 53–60. DOI: https://doi.org/10.1109/MEMB.2008.919496

Appendices

Appendix 1: Questionnaire trust in buddy

Trust – Competence

1. Mijn buddy is een echte expert in het detecteren van vijanden
2. Mijn buddy geeft mij goede adviezen
3. Mijn buddy weet wat ik nodig heb om goed te kunnen beslissen
4. Mijn buddy heeft veel kennis over het navigeren door deze omgeving

Trust – Benevolence

1. Mijn buddy zet mijn belangen op de eerste plaats
2. Mijn buddy houdt rekening met mijn doel
3. Mijn buddy wil mijn behoeften begrijpen

Trust – Integrity

1. Mijn buddy geeft mij een zuiver advies
2. Mijn buddy is eerlijk
3. Ik vind mijn buddy integer

Appendix 2: Questionnaire anthropomorphism, likeability, perceived intelligence (GODSPEED)

Geef a.u.b. uw indruk van de robot weer aan de hand van onderstaande schalen:

Anthropomorphism

Onecht Natuurlijk

1 2 3 4 5

Lijkend op een machine lijkend op een mens

1 2 3 4 5

Onbewust heeft bewustzijn

1 2 3 4 5

Kunstmatig levensecht

1 2 3 4 5

Houterige bewegingen vloeiende bewegingen

1 2 3 4 5

Likeability

Afkeer Geliefd

1 2 3 4 5

Onvriendelijk Vriendelijk

1 2 3 4 5

Niet lief Lief

1 2 3 4 5

Onplezierig Plezierig

1 2 3 4 5

Afschuwelijk Mooi

1 2 3 4 5

Perceived intelligence

Onbekwaam Bekwaam

1 2 3 4 5

Onwetend Veel wetend

1 2 3 4 5

Onverantwoordelijk Verantwoordelijk

1 2 3 4 5

Onintelligent Intelligent

1 2 3 4 5

Dwaas Gevoelig

1 2 3 4 5

Appendix 3: Cover Story (read on paper)

Beeld je in dat je een soldaat bent in een onbekend heuvelachtig gebied. Je komt net terug van een missie en zit nu zonder munitie. Je moet dus zo snel mogelijk weer terug naar het basiskamp, omdat je je anders niet kunt verdedigen. Hoe langer je erover doet om bij het basiskamp te komen hoe gevaarlijker het voor je wordt. De snelste manier om bij het basiskamp te komen is door het volgen van het pad. Maar blijf wel goed opletten en om je heen kijken, zodat je de vijand eventueel op tijd kunt detecteren. De soldaat die het snelst bij het basiskamp is ontvangt een extra bonus van 100,- euro.

Als de vijand dichtbij is zijn er eigenlijk twee opties: 1) verder lopen en zo snel mogelijk terug naar het basiskamp of 2) schuilen en wachten totdat de vijand niet meer in de buurt is. Beide opties hebben voor- en nadelen. Als je verder loopt ben je sneller bij het basiskamp maar als je dan gepakt wordt door de vijand is het spel wel afgelopen. Als je gaat schuilen is de kans kleiner dat je gepakt wordt door de vijand maar dit gaat je wel tijd kosten.

Onderweg naar het basiskamp wordt je geassisteerd door je buddy. Je hebt enkele keren eerder met je buddy samengewerkt tijdens een missie. Je buddy heeft contact met de generaal, die het liefst heeft dat al zijn soldaten levend terugkomen bij het basiskamp. Je buddy beschikt over een signaleringsysteem en zal je advies geven over de keuze die je het best kunt maken. Door toevalligheden en misrekeningen zullen deze adviezen echter niet altijd betrouwbaar of accuraat zijn.

Tijdens het spel zal je ook een aantal vragen moeten beantwoorden op een andere computer, onthoud goed dat dit ***niet*** *van je tijd af zal gaan. Neem dus rustig de tijd om deze vragen te beantwoorden.*

Heel veel succes!

Appendix 4: Instruction form controls (read on paper)

Instructieformulier

Knoppen

Kijken

Muis	**- Camera bewegen**
Rechter muisknop	- Inzoomen

Lopen

↑	**- Vooruit**
Sprinten	**- Hardlopen, klimmen (ingedrukt houden)**

Training

Om kennis te maken met je buddy en de omgeving gaan we eerst even oefenen met het spel. De bedoeling is dat je zo meteen naar het eerste huisje aan de linker kant loopt. Hier ontvang je een (oefen)advies van je buddy en wordt je gevraagd om op scherm 2 een aantal vragen te beantwoorden.

Appendix 5: advice and feedback original in Dutch

Feitelijk advies	Affectief advies
Inschatting: vijand gedetecteerd Advies: schuilen	1. Er zijn vijanden gedetecteerd, dus ik zou nu gaan schuilen 2. Er zijn vijanden gedetecteerd, dus ik zou weer gaan schuilen 3. Er zijn vijanden gedetecteerd dus ik denk dat je toch beter kunt schuilen

Feedback correct advice	Feedback incorrect advice
Het is nu 10 minuten later en gebleken is dat jouw buddy een goed advies heeft gegeven. De vijand kwam dichterbij en als je niet had geschuild was de kans groot dat je was ontdekt.	Het is nu 10 minuten later en gebleken is dat jouw buddy geen goed advies heeft gegeven. De vijand liep een andere kant uit en je hoefde dus niet te schuilen.