Working with an interface or a partner?

The influence of anthropomorphism on the design of an everyday working system

Master thesis at the Dutch National Police
HCI-5948436

Graduate School of Natural Sciences, Utrecht University
Human Computer Interaction

Written by:
Rosalie Wendt
Student number: 5948436

Daily supervisor:	Vincent Bodifee
First supervisor:	Christof van Nimwegen
Second supervisor:	Judith Masthoff

Utrecht, February 2023

Acknowledgements, Woord van dank

I would like to give special thanks to several people who made authoring this thesis possible. Without their help, this would have been a much more difficult journey. I am grateful for their time and effort, which played a crucial role in the success of this thesis. I appreciate everything they have done.

I wish to express my sincere gratitude to Vincent Bodifee, Bastiaan Klooster and Simone Stroomer. Their insights, expertise, and guidance have been instrumental in shaping this work. Their knowledge about research has taught me a lot and their contributions and support were invaluable. They were always there to help whenever I needed it.

I would like to convey appreciation to Edo-Jan Meijer, who helped in making this project possible. He provided me with as much support as possible. A thank you to Govert Buizert, who provided me with a police phone on very short notice.

Christof van Nimwegen played a very important role in writing this thesis as well. I am thankful for his support and guidance throughout this project. He always offered good support and guidance when it was needed.

I am very appreciative of the participants who participated in this study. Without them, this research would not be possible.

I would also like to thank Judith Masthoff for her time and effort.

Abstract

This research aimed to examine the impact of anthropomorphism, applied by adding subtle features to an interface, on the usability, credibility, trust, and task efficiency of RAPP (Registrative App Police Processes, Registratieve App Politie Processen). The study utilized three prototypes (RAPP, ARAPPV and ARAPPN-V) which incorporated no cues, verbal cues and non-verbal cues, respectively. A within-subject design was used with a sample size of $N = 13$. Results indicated that anthropomorphism improved usability in ARAPPV (verbal condition) and ARAPPN-V (non-verbal condition) and task efficiency in ARAPPV. However, there were no significant results found in terms of credibility and trust. Qualitative results suggest that the stability of the system needs to be improved to enhance credibility and trust. Participants expressed support for ARAPPV to guide less experienced users.
Overall, the research provides evidence that non-obvious anthropomorphism through verbal cues can improve usability and task efficiency, and to some extent credibility and trust, when designing an interface used in a daily work environment.

Keywords: Anthropomorphism, User Experience (UX), System Usability Scale (SUS), Credibility, Trust, Task Efficiency

Nederlandse samenvatting

Het doel van deze studie is het onderzoeken van de invloed van antropomorfisme, toegepast door subtiele kernmerken toe te voegen aan een interface, op de gebruiksvriendelijkheid, integriteit, vertrouwen en taakefficiëntie van RAPP (Registratieve App Politie Processen). Het onderzoek maakte gebruik van drie prototypes (RAPP, ARAPPV en ARAPPN-V) die respectievelijk geen cues, verbale en non-verbale cues bevatten. Er is gebruikgemaakt van een between-subject design met een steekproefgrootte van $N = 13$. Resultaten wezen uit dat antropomorfisme de gebruiksvriendelijkheid verbetert in zowel ARAPPV (verbale conditie) als ARAPPN-V (non-verbale conditie) en taakefficiëntie in ARAPPV. Er zijn echter geen significante resultaten gevonden op het gebied van integriteit en vertrouwen. Kwalitatieve resultaten suggereren dat de stabiliteit van het systeem verbeterd moet worden om integriteit en vertrouwen te verhogen. Participanten gaven aan dat ARAPPV zeer geschikt was om minder ervaren gebruikers te ondersteunen.
Samenvattend, wijst dit onderzoek uit dat niet overduidelijke (non-obvious) antropomorfistische elementen, zoals verbale cues, de gebruiksvriendelijkheid en taakefficiëntie kunnen verbeteren in een systeem dat gebruikt wordt voor dagelijkse werkzaamheden. Tot op zekere hoogte kan het ook integriteit en vertrouwen in dat systeem vergroten.

Kernwoorden: Antropomorfisme, User Experience (UX), System Usability Scale (SUS), Integriteit, Vertrouwen, Taakefficiëntie

Contents

1 Topic and motivation

A perfect interface. Completely user-friendly. That is the dream of every UX/UI team. This can be achieved by an optimal agreement between user and software. The UX/UI team of the Dutch National Police is also trying to achieve this optimal agreement (see Figure 1.1). This is done by using the following police oath which every police officer needs to take when becoming an officer:

Figure 1.1. Overview of agreement between user and software. Translated from presentation by Meijer [52]

Art. 3 sub 2 of the police Act states: *"The task of the police, subordinate to the competent authority and in accordance with the applicable legal rules, is to ensure the effective maintenance of the rule of law and to provide assistance to those who need it"* [1].

Figure (1.1) is based on the above oath. The core values of the oath are translated for practical use. For example, honesty is a key value, therefore the system should be competent. As said, if the values of the user are in agreement with the values of the

[1] Translated from original source: De politie heeft tot taak in ondergeschiktheid aan het bevoegd gezag en in overeenstemming met de geldende rechtsregels te zorgen voor de daadwerkelijke handhaving van de rechtsorde en het verlenen van hulp aan hen die deze behoeven. (Art. 3 lid 2 Politiewet) [62]

software, the system will function optimally. The past few years show investment in the creation of such a system.

1.1 History of the Operational Police Platform

To understand why the police systems need changing, we have to go back in history. The Dutch police history starts with the formation of the Republic of the Seven United Netherlands in 1581. A police organisation without clear authorizations and tasks is founded. We skip the part were Napoleon starts the creation of a police force in 1810, and move to 1945.
After World War II, the government decided to start a new police organisation to restore and control the public piece. A deviation was made between police per municipality and a police department for the entire nation (Korps Rijkspolitie). After 1994, this distinction disappeared and 25 regional police forces and a National Police Agency (Korps landelijke politiediensten, KLPD) are its replacement. Up till 2013, they were able to act independently and were able to make their own choices. From hereonend, the police department is one organisation divided into 10 regional forces, the National Unity (Landelijke Eenheid) and the Police Service Centre (Politiedienstencentrum) all in charge of one chief of police [61, 63].
In the past, every municipality operated on its own, and used its own system. In 1980 the government asked to create a coherent "automation policy" for the state police and police per municipality. Plans were made to create the system, but affairs were not in order and the actual creation of the system was left to wait. The years between 1990 and 2017 show many attempts to bring affairs in order. An investigation carried out by de Algemene Rekenkamer, experts in the inspection of public order and safety, showed that usability, continuity and stability of the systems were inadequate [55, 70]. This lead to the creation of a new program in 2011, called Aanvalsprogramma (AVP). The program aimed to realise a useful, flexible, affordable and good national police system for the future [17, 63]. The start of AVP was caused by a disruption of the IT service. Because of this, the police IT services were limited for 8 weeks. The police officers criticised the key systems for enforcement (BVH) and detection (BVO).
In 2017, the AVP program was shut down. Modernised infrastructures resulted in improved stability and continuity. The new National Unity (Landelijke Eenheid) was supported by a national system and the first steps were taken to renew the operational systems. However, more progress was needed to continue this innovation [17]. Therefore, Program Innovative Registration (Programma Vernieuwend Registreren, PVR) was founded in 2018 [63].
The main goal of PVR is to support officers and detectives in the registration and retrieval of data in a way that is unambiguous and user-friendly. This is done by redesigning the existing systems (see figure 1.3), but the current systems have aged in

technique and functionality and can therefore, not be used for further development [33, 34, 79]. This resulted in the creation of the Operational Police Platform (OPP): all registration functionality integrated into one platform [34, 79] (see figure 1.3 for more clarification).

Old police systems, called Legacy systems, that will be integrated:

- Basic Provision for Enforcement (Basis Voorziening Handhaving, BVH): used by the standard teams for daily work. For example, incident registration and taking actions accordingly.
- Amazone: registration of target groups, like gangs, that cause nuisance.
- Police report via Intranet (Aangifte Via Intranet, AVI): to deal with the most common police reports in a guided and structured way.
- Detainees Module (Arrestantenmodule): Is a part of BVH and used to support the logistical tasks of detainee management.
- Better Detection by Managing Business (Betere Opsporing door Sturing op Zaken, BOSZ): is used to manage criminal investigations. Police officers and prosecution (OM) can control this.
- Summ-IT: the most important national detection system for detectives.
- Basic Facility Detection (Basis Voorziening Opsporing, BVO-bruto): used to register information of informants.
- National Coordination file Traces (Landelijk Coördinatiebestand Sporen, LCS): used to handle forensic investigations (trace evidence) and coordination of corresponding traces.
- National Track Following (Landelijk Spoor Volgen, LSV): supports the LCS system. It provides an overview of the data processed by LCS.
- Technical Detective Information System (Technische Recherche Informatie Systeem, TRIS): used in forensic investigations.
- Custody (Beslag): is used to register and deal with confiscated goods.
- Detection System (OpsporingsSysteem, OPS): application used in the past to identify people nationally.

PVR will add, amongst other things, to the improvement of usability. The main frustrations of the current systems, like signing in multiple times, working with lots of systems that differ in interface as well as usage and entering data multiple times, will

be resolved [34]. The result, a new operational platform with low administrative expenses, increase in usage, quality of data and work environment [33, 79]. Another advantage resulting from this OPP is that the design will be in accordance with the Law Police data (Wet Politiegegevens, Wpg). The current systems do not account for this law, because they were designed before it was implemented [79].

Main targets of PVR [63, 79]:

1. Improve usability. Administrative tasks are reduced by means of effective and easy support in recording and retrieving data. This has a positive effect on the work environment of police officers.

2. Working on smartphones, tablets, and laptops, next to a set computer on a desk, independent of time and place. By means of this improvement pictures, audio, and video can be combined with text and documents. Police officers are enabled to take pictures, videos etc. on sight, at the location of the crime, and are enabled to share those with colleagues.

3. Storing information only once, unambiguous, and qualitative reliable. Data is collected at a central location, and errors can no longer arise as a result of having to re-enter the same data in other systems.

4. Removing current application boundaries, which facilitates the workflow and the related information streaming process. Polices officers are no longer confronted with multiple, self-functioning systems that all add a piece to finalize the total puzzle called the administrative process.

5. Lower maintenance costs and faster implementation of changes necessary to do police work. This can, for example, be changes based on new law enforcement and new technological possibilities [34].

6. Aid in saving time while doing police work.

This research will focus on improving usability (1) and additionally aid in saving time while doing police work (6).

1.2 RAPP

Implementing the legacy systems to the OPP to realise these goals is a step by step process. The systems run 24/7, therefore shutting them down to update is not an option. To achieve the above targets, the systems must be replaced one by one, starting with BVH [33]. RAPP (Registrative App Police Processes, Registratieve App Politie Processen) will replace BVH.

This research will focus on RAPP. To understand why this is important for the implementation of the other systems, we first need to zoom in on the commonality between the legacy systems. This starts with the most common crime, namely shoplifting.
When such an incident occurs, an officers has to go over a series of steps. The officers must for example register his own perception, record a declaration, question the suspect, send out the file, close the case etc. For every step, a new ICT-functionality is designed. This is called a "building block". The steps-by-step process to register a shoplifting incident has a lot in common with, for example the registration of a violent crime [33].
When looking at all the legacy systems, steps as notifying, reporting and identifying appear everywhere. These features are, as said, implemented in different building blocks, forming "generic building blocks". Next to this, each system also has specific functionality: coded in "specific building blocks" [79]. One by one the legacy systems will be implemented onto the OPP. To visualise this, figure 1.2 provides a schematic overview. When most of the generic blocks are created, it will speed up the process for the rest of the implementation, since only the specific ones will be left. Resulting in an OPP built out of generic building blocks and specific ones (see figure 1.3). It will also improve maintainability and technological progress, because only one block can be updated without touching the others [60].

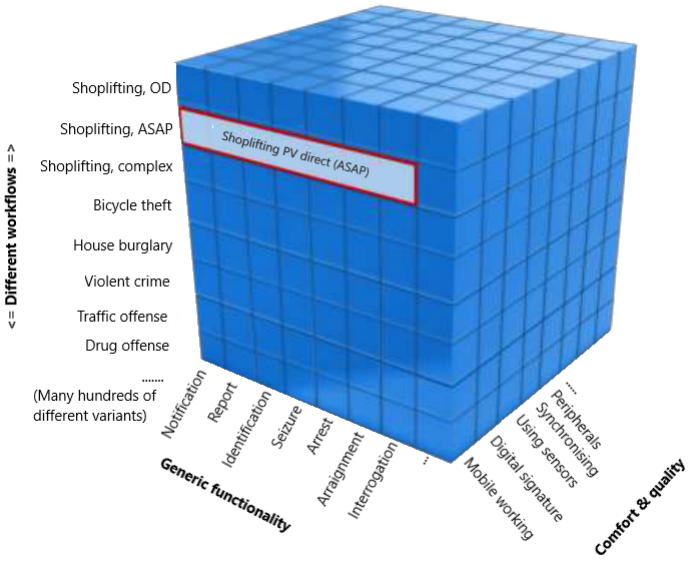


Figure 1.2. A schematic overview of the different boxes/building blocks used to create the whole OPP. Translated from parliamentary document [33]

Thus, shoplifting is the first process that is implemented on the OPP by means of RAPP. In RAPP all the building blocks come together. Lets again use the example of shoplifting: a notification appears in the RAPP app, because a report is filed by a victim. The police officer on duty "picks up" the notification and fills out the step-by-step process: register own perception, recording declarations, questioning suspect,

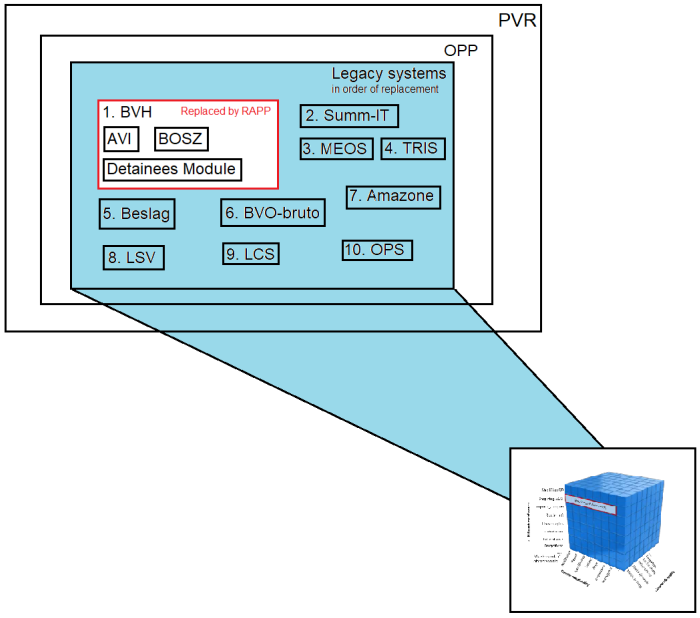


Figure 1.3. Clarification of PVR: PVR will create the OPP, which consists of the replaced legacy systems. The numbers 1-10 represent the order in which the systems will be replaced. First of all, RAPP will replace BVH, AVI, BOSZ and Detainees Module.

sending file, closing case etc.

When RAPP has provided the example of correctly handling a shoplifting incident, the new systems replacing the old legacy systems (e.g. the registration of a violent crime) can benefit from that example. A correct design of RAPP is, therefore the start of a well implemented OPP. The current interface design of RAPP can be ambiguous in some cases and it is not yet as user-friendly as it should be. This study will hence focus on improving certain element of the current design of the RAPP interface.

The RAPP interface is designed to support its users. All the features that are important for a user are translated into system features, as explained by figure 1.1. When a police officer wants to fulfil his duties, the system should provide support. When the user wants to save energy, the system should be simple. When the user searches for patterns, the system should use patterns etc. To achieve this goal, the police UX/UI team has designed contextual frames for the user and contextual flexibility for the system. To relieve an officer from mental pressure for example, the software should have a clear role and to fulfil a task, the software should be focused on the user (be personal). Research says that for systems to function in a good way, they should be social and cooperative [45], as if the user is interacting with a human.

This research focuses on this human aspect, by means of triggering anthropomorphism in RAPP, with the aim of improving the cooperation between RAPP and the police officer. Research shows that anthropomorphism has a positive effect on people's attitude towards technology. Using a product while anthropomorphising is experienced as relatively pleasant. Therefore, the desire to use it again will increase [20, 68]. This suggests that anthropomorphising RAPP would cause a similar effect. If this is the case, it could be used in the implementation of the other building blocks unto the new police platform as well.

1.3 Anthropomorphism

Anthropomorphism is a term that is used for the attribution of human characteristics to objects and non-human beings. It is also known as humanisation [1, 38]. Humanisation is: "The tendency to imbue the real or imagined behaviour of non-human agents with human-like characteristics, motivations, intentions, or emotions" (Rauschnabel & Ahuvia, 2014, p.375 [68]). In line with this is the Media-Equation. A theory that says that people tend to attribute human characteristics to computers and technology and treat them like they are real social beings [16]. Anthropomorphism can occur when a human recognises certain human traits in a system, for example a human face, a human-like shape or a voice. So, a system is human-like and because of that the user anthropomorphises the system.
Research has shown that anthropomorphising technology makes people express more positive thoughts towards the technology. The technology is seen as engaging and friendly. As a result, the willingness to use the technology increases [38].
Making RAPP more human-like could cause the police officer to anthropomorphise, which in turn might result in a more positive feeling towards RAPP.

2 Related work

Research shows that anthropomorphism has an influence on how humans interact with products and technology. Technologies which make use of anthropomorphism aim to form relationships with their users and gain trust. An illustration of an application that is designed to engage with its user is Talking Tom Cat, which repeats back spoken words, needs food and responds to touch [49]. Talking Tom Cat 2 even has clothes on and can look after his own pet [48]. Talking Tom Cat is not the only "virtual pet", a great number of such technologies exist [42, 64]. Robots are an additional example of engaging technologies, for instance health care robots for elderly [7, 71] or robots to aid children with autism [12]. Conversational Agents (CA's), like chatterbot, too, are interaction systems that make use of text-input to engage a user [13, 19]. Even the voices of navigation systems are adjusted to make utilisation more attractive [6].
A way to achieve the goal of engaging a user, is for the system to behave in such a way that people ascribe personality and emotions to the technology [6]. The human-like behaviour thus leads to anthropomorphism. So, adding human characteristics, humanising, can be used to anthropomorphise technologies. Examples of human characteristics could be imitation, intentions, movement, timing of the movement, facial features, sounds/voices, communication ability and human-like shape [1]. Par example: adding a face to a conversation with a chatterbot can cause a user to treat it more human-like (anthropomorphise).

2.1 Anthropomorphism and facial expression

Research by Koda & Maes [45] shows that a face makes people engage more and has the additional value of being more likable and comfortable. A poker game was designed to test agent-based interfaces. There were five conditions in which the user could play against either two or three computer faces. Depending on the condition this could either be a realistic male or female face, a caricature male or female face, a caricature dog face, a smiley face or an invisible man (white block). Facial expressions would change depending on the action the computer would take. The results indicate that facial expression is not seen as a distraction. On the contrary, it makes users pay attention and engage in the task, because they try to interpret the faces. However, a computer player with or without a face was rated equally intelligent as well as a human and a dog face. This shows that intelligence is not based on appearance, but on competence.

The study also investigated the difference between participants who are supportive of personification and those who are not. People who are supportive of personification preferred the dog face and attributed more intelligence to expressive faces, while those who are not supportive preferred the human face and attributed more intelligence to faces without expression. This highlights the urgency of accounting for what the user wants when designing a personified interface.

2.2 Anthropomorphism and speech interfaces

This is in concurrence with research by Yarosh et al. [82] which investigated the influence of personification and personalisation in speech interfaces. They defined personification as a system referring to itself and personalisation as the system referring to the user. According to the researchers, the use of personification and personalisation in speech interfaces is open for debate. Yet, most interfaces that make use of speech in commercial environments, like Google Home, Siri and Amazon Alexa, make use of both. Adults as well as children utilise these technologies, while they might do so differently. However, the functionality for both user groups is the same and not user focused. The aim of this research was to investigate children and adult preference for personalisation and personification in speech interfaces. For this study 87 children between the age of 5 and 12 asked questions to three voice assistants. One voice assistant talked about itself in first person (personification) and also used the child's first name (personalised), the second voice assistant was personified but not personalised and the last voice assistant was neither personified nor personalised. Findings show that children do not prefer interfaces that are adjusted to their age and name, although they do prefer the interface that mentions its own name. The research also investigated the way adults ask questions to the voice assistants. The descriptive results suggested that adults show a preference for the personalisation condition. However, no significant difference was found between the personalisation condition and the personification condition. The conclusion stipulates the urgency to take into account the users' needs when designing a speech interface that shows human characteristics.

2.3 Task difficulty and cognitive fluency

Research conducted by Rauschnabel & Ahuvia [68] explains that anthropomorphising an object depends on a persons' expertise. If a user has a lot of knowledge about the object, the need to anthropomorphise is lower. Anthropomorphism is described as a heuristic: when people do not have the knowledge to understand a product, they use the knowledge they have about human behaviour to make sense of it. Engaging in anthropomorphic thinking, results in an increase of cognitive fluency: the difficulty of a cognitive process. When cognitive fluency is high (low difficulty), it results in a good

feeling and an intuitive fit with the product. In contrast to when a product has a low cognitive fluency (high difficulty). This will lead to frustration.
The link between a good feeling and an intuitive fit, is further explored in research by Aggarwal & McGill [1], which hypothesizes that a match between features of a product and expectations of the user, result in a certain satisfaction. This will add to a possible positive evaluation. Yet, the research shows anthropomorphism does not always have a positive result when evaluating a product. It is explained that users have a certain schema, an idea of what the product will look like and how it is used. When evaluating a product, a comparison is made between the expected product and the actual product. If the actual product does not fit with the expected schema, the evaluation will be less positive. This means that anthropomorphising the product is easier when the expectation fits with the schema. The researchers concluded that for a positive evaluation of a product with human characteristics, it needs to be presented as being human. If the product is presented as a human, but does not have human characteristic, it will be evaluated less positive.
This positive evaluation is in line with research by Delbaer, McQuarrie & Phillips [20] which concludes that photo realistic images that display a product engaging in human behaviour, motivates people to anthropomorphise. The research used these images to personify brands in commercials, showing that it will lead to more attributions of brand personality and more emotional response. This in comparison to advertisements which did not use this personification.

2.4 Anthropomorphism and social presence

Research by Reeves & Nass [69] investigates the link between anthropomorphism and politeness. When a human asks the question: "How am I doing?", the reaction will almost always be that the person who asks the question is doing well. This is the case, because it is a polite answer. When someone else asks the question for another person: “How is he doing?”, the response can be negative as well as positive. This is because being polite does not matter anymore. The researchers hypothesised that when a human anthropomorphises the computer, the response to the question asked by a computer will be polite. Thus, when the human anthropomorphises the interface, and the interface askes: “How am I doing?”, the human will respond positively. While when someone else asks: “How is the interface doing?”, the human will respond honestly. The research showed that, as long as there are signs of social presence, the response to the question will be polite.

3 Research question

3.1 Research question

Current research is focused on deliberate making a system more human-like to trigger anthropomorphism. Typical human characteristics are added to highlight that a system is supposed to be more human, so obviously making it human. Robots and technology are given a name, human voice, clothes or body to trigger anthropomorphism, but there is no research that investigates non-obvious anthropomorphism: adding human-like adjustments that are not obviously human characteristics. As described, research by Aggarwal & McGill [1] shows that for a positive evaluation of a product with human characteristics, it needs to be presented as being human. However, so far no research has been found that investigates a positive evaluation of a product with non-obvious human characteristics, without it being presented as a human. The user must not expect the system to be human, but the non-obvious cues should lead the user to anthropomorphise.
Additionally, most research is done to contribute to the entertainment business, advise bureaus, marketing strategies and Conversational Agents like, speech interfaces or chatterbots but there is no research that investigates the influence of anthropomorphism on usability or everyday life working systems.
Such an everyday working system is RAPP. RAPP is built to support the user while working "on the street", in the car and other (rough) police work environments. It is a task-oriented system in which the police officer has to go over certain steps to correctly finish a task, currently shoplifting. It can be used on a computer as well as on a phone, but this research will focus on RAPP that is used on the phone. Anthropomorphism could be of additional value here, since RAPP is a new system and officers are not yet used to the new implementation and layout. The functionality is not completed, therefore it is still a work in progress. Police officers work with the system everyday, so usability is a hot topic. If an officer works with RAPP as if it were a human, this could improve the usability further. This research is based on the assumption that subtle changes to the RAPP interface will make the police officer anthropomorphise RAPP while using it. The research question is as follows:

Can non-obvious human-like changes in behaviour and communication to the interface of RAPP improve usability, credibility, trust and task efficiency?

Using a product while anthropomorphising is experienced as relatively pleasant. Therefore, the desire to use it again will increase [20, 68]. Using a system that is pleasant while working, may increase the overall happiness of the job. However, it starts with the improvement of task satisfaction in the form of usability. This leads to the following hypothesis:

H1: When a police officer anthropomorphises RAPP by means of non-obvious adjustments in behaviour and communication, this will lead to improved usability of RAPP.

According to McCroskey & Young [51], credibility is a characteristic which can be used to indicate the extent to which a person is seen as trustworthy, believable and competent. Anthropomorphic avatars are seen as more credible and attractive [78]. Which can in turn improve trust further. Research suggests that anthropomorphism can increase the amount of trust in a system [30, 80, 39]. Trust is the degree of confidence in technology which manifests itself in the willingness to rely on automated systems in uncertain situations [39]. Research by Hoff & Bashir [39] reviewed existing research on trust. They concluded that trust in a system, improves the connection between user and system. Improved trust and reliance, may lead to greater efficiency. In line with the above, the following hypotheses are stated:

H2: Anthropomorphising RAPP by means of non-obvious adjustments in behaviour and communication will improve credibility in RAPP

H3: Anthropomorphising RAPP by means of non-obvious adjustments in behaviour and communication will improve trust in RAPP

In line with the above research, Spatola et al. [75] investigated the influence of social and non-social robots presence on performance. During the experiment, participants were asked to either describe the robot design without any response of the robot: the non-social robot condition, or interact with the robot by means of speech: the social robot condition. The research found that the presence of a social robot reduces the amount of mistakes and increases performance compared to a non-social robot. It was suggested that anthropomorphism has an influence on the perceived presence of the robot, influencing the performance. Following from this research, the below hypothesis is proposed:

H4: Anthropomorphising RAPP by means of non-obvious adjustments in behaviour and communication will lead to improved task efficiency while using RAPP

Task efficiency is defined as: the capability of performing right tasks in the right way. Working with RAPP as if an officer is collaborating with his partner, can improve task efficiency and with that improved task satisfaction. This will improve the entire national police and hence, the safety of the whole country. This is hypothesized by the UX/UI team of the Dutch police [52]. The results of this research could have additional value for future research in interface design as well, more specifically to Conversational Agent (CA) interfaces and systems that are used in daily work environments. It would be an example of how non-deliberate human-like elements influence anthropomorphism, which could in turn improve a systems interface in daily work environments even more. Anthropomorphism can thus be a critical characteristic when designing an interface.

3.2 Prototype design

To test the hypotheses, three prototypes were built. The first prototype, RAPP, has no anthropomorphic adjustments and is used as base case. The two other prototypes do have anthropomorphic adjustments (see table 3.1). This first human-like prototype of RAPP will be called ARAPPV (verbal condition) from now on. The second will be called ARAPPN-V (non-verbal condition). This section will explain what RAPP currently looks like and how the anthropomorphic prototypes (ARAPPV & ARAPPN-V) are built.

Table 3.1. A schematic overview of the difference between RAPP, ARAPPV and ARAPPN-V in terms of anthropomorphic adjustments

Type of adjustments			
	None	**Verbal**	**Non-verbal**
RAPP	✓		
ARAPPV		✓	
ARAPPN-V			✓

3.2.1 Current design

The images 3.1 & 3.2 [1] provide an overview of the current RAPP design on a desktop. The application for the phone is visible in figure 3.3. The phone application of RAPP is used as framework for both ARAPPV and ARAPPN-V.
Currently, RAPP works in the following way: A shoplifting incident is registered on the police website by a victim, for example a supermarket. This appears in RAPP (3.1 & 3.3,1) and the officer on duty "picks it up". This means that it is going to be processed. The officer needs to fulfil a certain amount of tasks (acties), depending, among others, on the gravity of the crime, the age of the perpetrator, (monetary) value of the stolen

[1] This is not a real incident. It is created as a case study for this prototype.

goods, and if it is a first-time offense. All these steps need to be registered by the officer in RAPP to finish a shoplifting incident. Each step can result in a different output. RAPP is a smart system which provides recommendations for certain actions to help the user in the step-by-step process. However, because there is a lot of variety in actions to take, considering that there are many factors to take into account, it is sometimes hard to figure out what needs to be done in certain situations.
This is why ARAPPV and ARAPPN-V are created as support for the user and to guide the officer through ambiguous situations.
The following section will provide an overview of the literature taken into consideration when building the anthropomorphic prototypes and will explain exactly which adjustments are made to both ARAPPV and ARAPPN-V. Non-obvious adjustments are added to realise these prototypes. The applied features will be small, but it is expected that they will have a powerful influence.

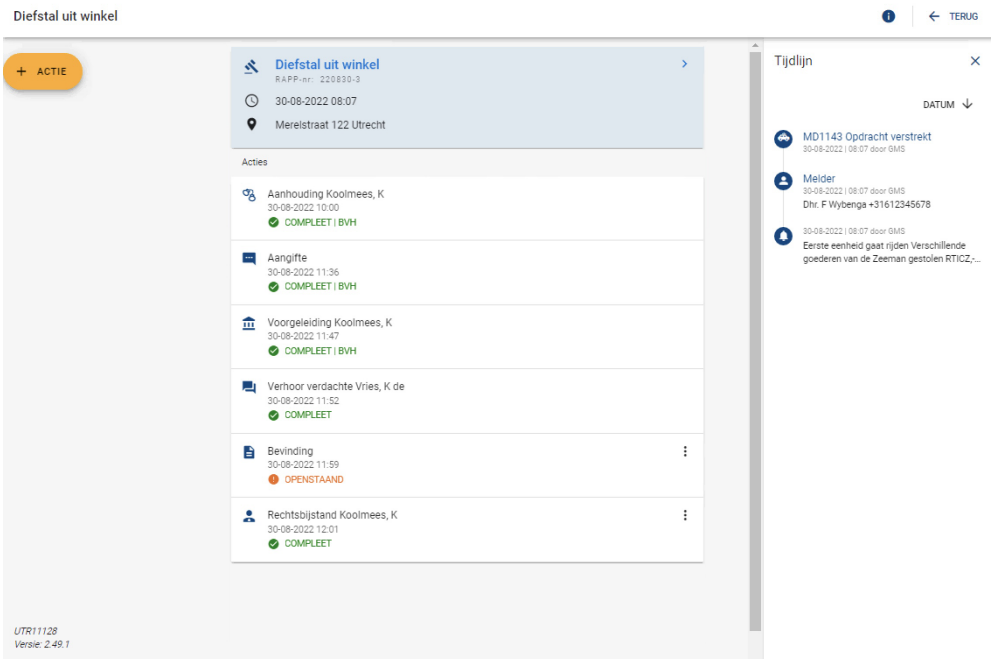


Figure 3.1. When the shoplifting incident is "picked up" by the officer, this screen is shown. This picture screen already contains additional actions, which must be added one by one using the "+ actie" button. All the actions that are signed off are marked green ("COMPLEET"). The orange one ("OPENSTAAND") is not yet completed. The timeline on the right, provides an overview of the time passed since picking up the incident. The system provides suggestions of actions that might be useful.

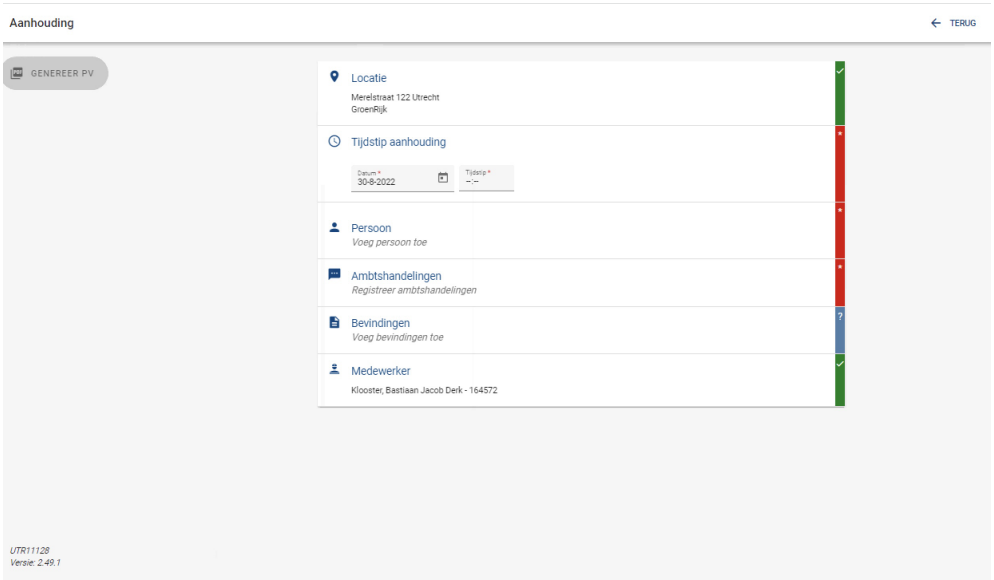


Figure 3.2. An overview of the screen when a new action (in this case aanhouding) is started. On the right the colour red shows that the task is not yet completed, it will turn orange when it is opened, but something is forgotten. When it turns green it is completed. The colour blue shows that the action is not obligatory but can provide additional information. When all the tasks are fulfilled (all green) the button "genereer PV" will turn yellow. Clicking on the button shows an overview of what is filled out. When this is checked the generated PV can be send off. After that the officer has to press "terug" to turn back to figure 3.1. A new action can be started if needed.

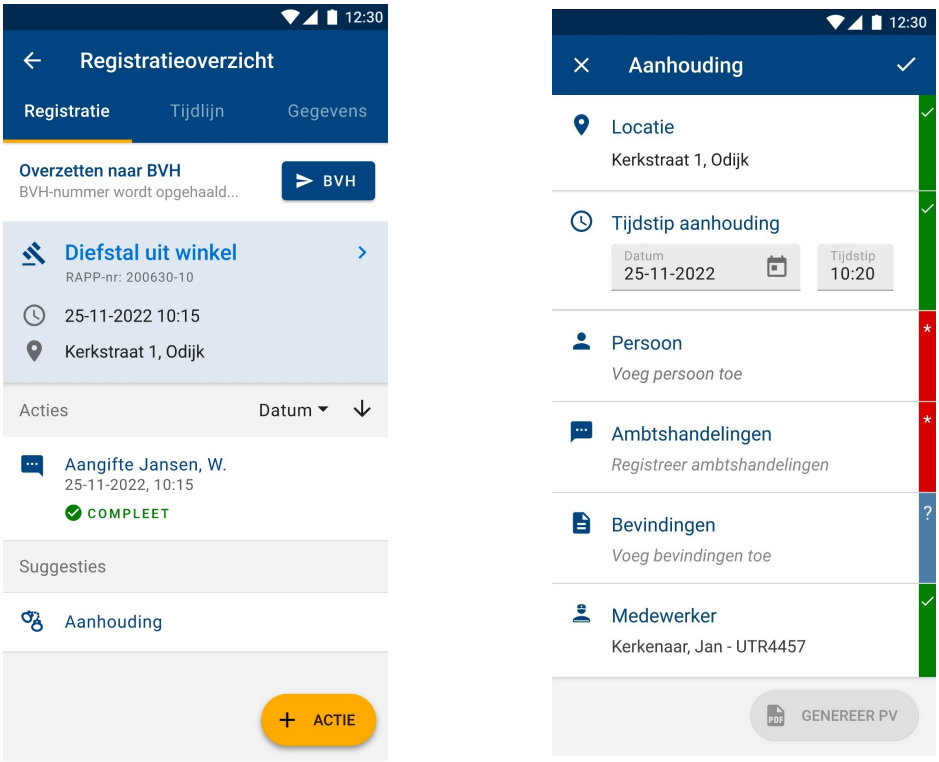


Figure 3.3. These images represent figure 3.1 and figure 3.2, respectively when used on a phone.

3.2.2 Theory of anthropomorphism

According to the theory of anthropomorphism [24], the likelihood of anthropomorphising is based on the following stimuli:

- Sociality motivation
- Effectance motivation
- Elicited agent knowledge

A distinction is made between the motivational determinant and the cognitive determinant. Sociality and effectance motivation are part of the motivational determinant. Sociality motivation touches on the urge and desire to form social connections with other humans. Anthropomorphism can ensure a perceived human-like connection with a nonhuman agent. When such a social connection is not available, "human" agents are built out of nonhuman agents with the help of anthropomorphism. This to fulfil the need for a social connection.
Effectance motivation entails the motivation to interact with non-human agents is an effective fashion. It supports the ability to make sense of complex situations in the present and anticipates how futures behaviour will be. The attribution of human characteristics and motivations to nonhuman agents, anthropomorphism, increases the capability of

understanding an agent's actions. It lowers the uncertainty and helps to predict future actions.
Elicited agent knowledge is grouped under the cognitive determinant and refers to the degree to which anthropocentric knowledge is accessible and available. So, technology has anthropomorphic characteristics and has knowledge about how humans behave. When a human interacts with an unknown, nonhuman, system, or any unknown object in general, and it seems similar to the human itself or other humans, the user is likely to activate knowledge about other humans to judge the object. It could provide a good foundation for inductive reasoning. This is mostly because general knowledge about humans and self-knowledge is obtained in early stages of life and is broader compared to the knowledge about nonhuman agents. The theory of anthropomorphism claims that the elicitation of agent knowledge is encouraged when the appearance and behaviour of a nonhuman object is perceived as human-like [24].
When applying this theory to RAPP, the effectance motivation becomes apparent. RAPP is a relative new application, therefore using it could cause uncertainty. When a person has no knowledge about a non-human agent (technology), it is likely that anthropomorphising occurs to make sense of it, to reduce uncertainty and increase confidence, thus the effectance motivation. Elicited agent knowledge strengthens this. If there is anthropocentric knowledge available and accessible, plus a person wants to understand and control the situation, anthropomorphising increases. What thus needs to be done to build ARAPPV and ARAPPN-V in a non-obvious way, is increase the elicited agent knowledge and therefore increase human-like behaviour in both prototypes, since the effectance motivation is already applicable.

3.2.3 Design alterations

According to research [24, 73], enhancing elicited agent knowledge can be achieved by human-like behaviour and appearance. The design of an interface should thus prompt anthropomorphism. To prompt anthropomorphism in ARAPPV and ARAPPN-V, elicited agent knowledge was enhanced by altering the human-like behaviour and appearance of RAPP.
Human-like behaviour and appearance can be enhanced in three different ways, design dimensions, divided in 2 different types of observabilities:

- Features that are observable:
  - Human identity cues
- Behaviour that is observable:
  - Verbal cues
  - Non-verbal cues

3.2.3.1 Human identity cues

Human identity cues are observable features because human-like characteristics are added to technology. Human characteristics are divided in human-like visual representations, such as the addition of real human faces and demographic information, like gender and age. These features determine how a person (literally) sees technology. Inducing human identity cues in technology could add to anthropomorphism.

3.2.3.2 Verbal cues

Verbal cues can influence observable behaviour in three different ways. By means of social dialog, e.g., small talk, emotional expressions, not by means of faces but textual, like saying "I am happy for you" and lastly, context-sensitive responses. This entails responses adjusted to the users' responses. Unlike human identity cues, verbal cues do not influence the observable features. A chatterbot, for example, can make use of verbal cues, but it does not necessarily need a body (human identity cues). According to Isbister & Nass [41], verbal cues include both the way in which the speaker refers to itself and others and choice of words and sentences.

3.2.3.3 Non-verbal cues

Non-verbal cues are part of the observable behaviour category as well. Emoticons are part of non-verbal cues, next to temporal cues and turn-taking gestures. Temporal cues refer to for instance delayed response cues in a conversation, waiting to answer a user. Turn-taking gestures are an indication that the technology is active. Blinking dots when a chatterbot is typing for example, provide information that the user is not wasting time waiting for an answer, but the system is actually working.

A schematic overview of human identity, verbal and non-verbal cues is provided in figure 3.4. These cues are however not used in designing ARAPPV and ARAPPN-V, since RAPP is not a chatterbot and because the cues have to be non-obvious. Nevertheless, the verbal and non-verbal adjustments added to create the prototypes are inspired by the cues shown in figure 3.4.

Anthropomorphic Cue		Example
Human Identity	Human-like visual representation	Images of humanoid robots, human-like avatars, real human faces
	Demographic information	Gender, age, name, ethnicity
Verbal	Social dialogue	Greeting rituals (“Hi, how are you?”), anecdotes, non-task related questions (“How is the weather in Berlin?”)
	Emotional expressions	Apologies (“I am sorry”), congratulations (“well done!”), concerns ("I am worried")
	Verbal style	Self-references ("I", "me"), active voice, variability of syntax and words
	Context-sensitive responses	Adjust responses to individual user input
Nonverbal	Emoticons	😊, 😉, 😐, 😒, 👍
	Temporal cues	Delayed responses to signal writing or thinking
	Turn-taking gestures	Blinking dots, “is typing” indicator

Figure 3.4. Identified Anthropomorphic cues. Adjusted from research by Seeger et al. [73]

Research shows that combining all the above dimensions does not result in an optimal design. It is likely that this is caused by the uncanny valley effect [73]. If technology is humanised almost as a real human, but it lacks a lifelike appearance, the focus shifts to the imperfect details, thereby undoing the anthropomorphism effect in total [56]. A good anthropomorphic design is not the one with the cues that are most human-like. It is the one that is most consistent in the overall human-like appearance [73].

3.2.4 ARAPPV design

Figure 1.1 displays the idea behind the creation of RAPP. A user values specific characteristics in a system. These user values are translated into software characteristics. If both user and software characteristics are in agreement, the system should function optimally. To realise this agreement, a list was created with key values for the system [52]:

1. I am transparent and consistent in how I work.
2. I am effective, proactive and well prepared. I have know-how, am experienced and contain all the tools needed.
3. I know how to simplify complexity to prevent mistakes, misunderstandings and confusion.
4. I am a source of information and provide relevant information by linking the information available.
5. I will support you in your work process in a goal-oriented way. I will help you with new elements.
6. You know who I am, what I say, what I know and what I can do.

One of the main frustrations when currently using RAPP is the lack of communication, while this is a important part of the key system values (1,3,4). The need for communication to make sense of the system, is in concurrence with the effectance motivation. When communication is increased, the system behaves more human-like, which is in agreement with elicited agent knowledge. To distinguish this research from other anthropomorphic research, it is important to stipulate that the anthropomorphic adjustment will be non-obvious. ARAPPV is built by means of increasing elicited agent knowledge in a non-obvious way. The following section will explain how this is done.

3.2.4.1 Human identity cues

As explained before, human identity cues are observable human-like characteristics. It is not focused on behaviour, but on making the technology look like a human [24]. This can only be done by adding human-like cues in an obvious way, therefore these cues will not be used when building ARAPPV and ARAPPN-V.

3.2.4.2 Verbal cue design

The focus of ARAPPV will be on anthropomorphism as a result of adding subtle verbal cues. To prevent the verbal cues from being obviously humanlike, ARAPPV will not refer to itself (self-reference) or to the user and the output will not be vocalised/spoken

(active voices). Purely textual input is used, as if a colleague would tell you what to do. For example, after all tasks are completed the data is generated and can be send off, the officer has to press "BVH" to do so. In ARAPPV, the system tells you: "All tasks are complete. Transfer to BVH is possible". This touches on the sociality motivation: to form a social connection as if it is a connection with a police partner. ARAPPV will also provide information which explains what to do in situations were a lot of mistakes are made. It will guide police officers through the work environment in a friendly, human-like and gentle way. This supports the effectance motivation: support the ability to make sense of complex situations.
These textual adjustments are in line with research by Duffy [22], which explains that communication is an anthropomorphic ability and it can contribute to acceptance of a social robot into the human social circle. The research specifies that the communication of the social robot does not have to be complex. It just needs to be sufficient enough to communicate with people.

Social dialogue
Since RAPP is a working system, social dialogue is not applicable. This is not relevant in the work environment and will not be of additional value to the workflow. Chattaraman et al [14], differentiate between social-oriented and task-oriented conversations. Social-oriented conversations are conversations in which the computer interacts in its most "natural" human-like way. Task-oriented conversations are conversations in which a user is supported in the task. RAPP is a task-oriented system. Therefore, only textual support is necessary which does not include small talk.

Emotional expression
As RAPP is task-oriented, a formal writing style is enforced [14]. Congratulations or concerns are informal and social-oriented. It is inconsistent with the way RAPP is designed and it will therefore be excluded from the ARAPPV design.

Context-sensitive responses
The emotional expression is in line with context-sensitive responses. The idea behind the creation of ARAPPV is to provide users with information relevant to the context. To guide the officer while using the application as if guided by a colleague. For example, the officer filled out the data to complete the arrest. A message shows: "All required fields are filled out. Press the arrow to continue". This message provides a context-sensitive response to guide the user through the current user interface. Context-sensitive responses are the main verbal design cues, all the verbal cues are context-sensitive responses.

Attention management

The fact that a system should provide context-sensitive responses, is strengthened by results found in the research by Goodrich & Olsen [32]. It explains that an interface should support attention management. When the same mistake is made repeatedly, an interface should provide the relevant information to solve the problem. Par example: the user wants to press the "Zet om naar Reprimande" button, but it does not work. Instead of a grey button and no text at all, ARAPPV must provide textual information that explains what is going wrong when the user is trying to do something that does not work. It should explain that the reason it does not work, is because not all the boxes are filled out.

Variability of words and syntax

The textual content must explain what is said in a strict way, as literal as possible (e.g., provide tips and advice). All sentences must be well structured with sophisticated words. The use of abbreviations is not allowed. Lastly, there must be variation in words when constructing the sentences. This is done by using different words (synonyms when there might be word repetition). According to Feine et al. [25] providing tips and advice and building complex sentences that are lexical diverse are identified, among others, as verbal anthropomorphic cues. A formal language will be used, in correspondence with the task-oriented design of ARAPPV [14]. Formality is also identified as a anthropomorphic cue, as long as it is in congruence with the overall text style [25]. Research by Lortie & Quitton [50] investigated human-like characteristics in dialogues and concluded that the number of posts as well as the number of words and the length of the words per post have an influence on perceived humaneness in dialogues. The more posts, the more words and the longer the words (more than 6 letters), the higher the perceived humanness.

3.2.4.3 Non-verbal cues

Meta research conducted by Seeger et al. [73] explains that non-verbal cues alone, do not provide enough ground for anthropomorphism. It should be combined with either human identity cues or verbal cues, not both because this might lead to the uncanny valley effect, as explained earlier [73].
Figure 3.4 is based on results found in CAs (Conversational Agents). RAPP is not a chatbot, therefore the results found for non-verbal cues are not generalisable to RAPP. Temporal cues and the turn-taking gestures are irrelevant in the scope of this prototype. Emoticons are a social-oriented way of communication. This does not fit the task-oriented design of RAPP.

Agent appearance

Research by Feine et al. [25], wields a different format of social cues in designing a CA, as is visible in figure 3.5. Social cues are used as a synonym for anthropomorphic cues.

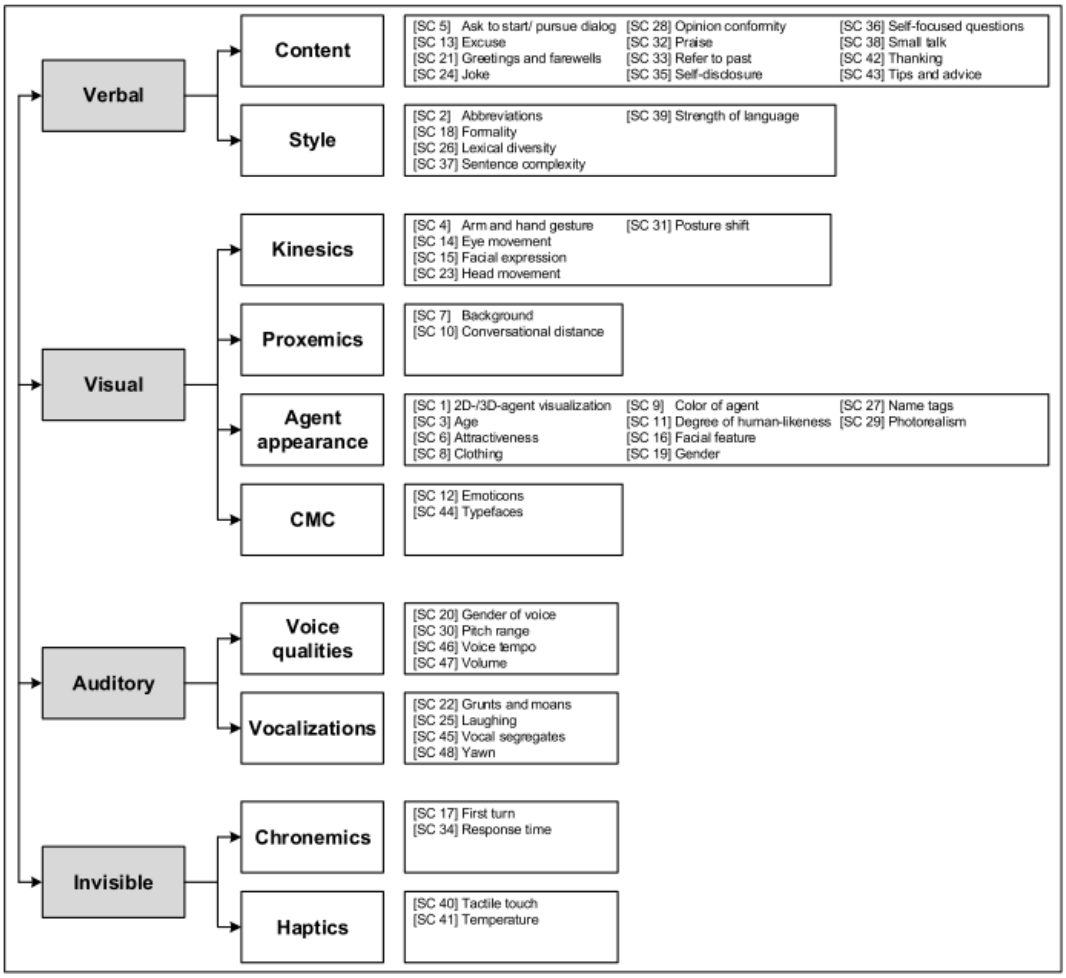


Figure 3.5. A taxonomy of anthropomorphic cues for conversational agents [25].

The research explains different degrees of agent appearance. An interface does not necessarily have to contain human identity cues to be social, as is applicable to this prototype design research. Thus, agent appearance does not necessarily have to be human-like. The degree of human-likeness can, for example, refer to the shape of the Conversational Agent (CA). This can range from artificial object to natural human, in the case of ARAPPV artificial. Attractiveness is used to explain how pleasant or beautiful the CA is perceived to be. Research by Altaboli et al. [2] makes use of the term visual aesthetics to describe the beauty or pleasing appearance of things. The level of symmetry was investigated by means of unity in size and number of objects. The results show that high level of symmetry, unity in size and number of objects, leads to increased visual aesthetics.

Feine et al. [25], also marks colour of the agent as anthropomorphic trigger. Emotional

design research investigates the influence of colour and shape on perceived pleasantness in interface design. Meta-analysis shows that the use of pleasant colours in interface design, improve learning and intrinsic motivation. It also slightly increases enjoyment and perceived difficulty [8]. Research by Shangguan et al. [74] shows that a visual emotional design, a visually attractive and colourful interface, has a positive effect on inducing and maintaining positive emotions.
The current RAPP design already considers colour. Take for example the colour markings as visible in figure 3.2:

- Green means complete. The colour green is associated with positive/pleasant [15, 54]
- Orange means incomplete but opened. Orange is associated with positive, however less positive then green. It has an overall lower ranking than both green and red [15].
- Red means incomplete and unopened. Red is associated with negative and taking action [15, 54].

The overall design of RAPP makes use of the colour blue. According to research blue stands for loyalty [54]. Additionally, blue is the colour of the Dutch National Police and in this case stands for unity. RAPP must be seen as part of the police force [52]. This is in correspondence with research by Fraune [29], which explains that computers belonging to the same team as the humans, were favoured over humans and computers who did not belong to the team.
In addition to the colours, shape is also considered. When looking at the corners of the buttons, they are not sharp, but slightly rounded. Research by Tien, Chiou & Lee [76] shows that round, face-like shapes induce positive emotions, lower perceived task difficulty and increase comprehension. This was confirmed in research by Gong, Shangguan, Zhai & Guo [31].
Research shows that non-verbal cues alone are not sufficient enough to anthropomorphise [73]. Communication research states that verbal and non-verbal human communication go hand in hand. Non-verbal communication must support verbal communication and they cannot be separated. A mismatch between verbal and non-verbal communication is considered disturbing and confusing [3].
Since RAPP already makes use of certain non-verbal cues, this is not an issue for the design of ARAPPV. In addition to the fact that human identity cues cannot be used, the above findings provide further supports for the use of verbal cues in the ARAPPV design.

3.2.4.4 ARAPPV implementation

Taking the above research into account, ARAPPV will provide as many supportive and explanatory messages as possible and needed. A formal writing style is used with complete, full sentences without abbreviations and without grammatical errors. Long words (6 or more letters) are incorporated when possible. However, it should not disrupt the efficiency and readability of the sentences.
Font and text size in the ARAPPV design will be the same as the current text and font used in RAPP to support consistency and retain symmetry.
A green police logo is added to the verbal cues to make ARAPPV part of the same team: the police force.
The supportive textual cues, as well as the police logo are placed inside a round cornered, green speech cloud, pointing toward the part of the interface it refers to. By adding the speech cloud, it looks like the system is speaking.
Lastly, an "ondersteuning?" on/off switch was added to the ARAPPV homepage, which could not be turned off. This was done to investigate whether participants would approve of such a switch or not.
Figure 3.6 displays the added verbal cues to ARAPPV. A complete list of adjustments added to create ARAPPV can be found in appendix C.

Figure 3.6. An enlargement of the added verbal cues to ARAPPV. Image 1 shows the verbal cue before "verstrekken gegevens" is started and image 2 shows the verbal cue after "verstrekken gegevens" is completed.

3.2.5 ARAPPN-V design

RAPP thus already makes use of certain non-verbal cues. However, research shows that this can be further elaborated. This is done for ARAPPN-V by adding attention cues. Attention cues are non-verbal cues which cannot be implemented without verbal cues. Therefore, ARAPPV is used as a basis for the construction of ARAPPN-V. The addition of attention cues (to ARAPPV to create ARAPPN-V) should engage the user to participate in joint attention.

3.2.5.1 Joint attention

Joint attention refers to the human ability to coordinate attention with a social partner [57]. It is a crucial element in language learning and other social abilities. Non-verbal communication provides information about a person's internal state [18]. This is also the case for joint attention which emerges in infancy. The infant starts to see another person as someone with intentions. They comprehend that when a person in attending to something, an understanding of the persons' intentions is needed [77]. Joint attention thus requires empathy; empathise with the other person, to understand intentions. Without joint attention, there is no social competence. Responding to joint attention means, following the gestures and/or gaze of the sender towards the attended, common point of reference. Hand gestures are a form of non-verbal communication [18] therefore, a (joint) attention cue is a non-verbal one.
Additionally, joint attention has a key influence in language learning. This is because, joint attention enables a link between a word and its actual meaning [11]. This way, joint attention supports verbal communication. Pointing gestures can for example be used to point to the attended stimulus. As said, hand gestures are a form of non-verbal communication.

3.2.5.2 Attention cues

ARAPPN-V contains attention cues to stimulate joint attention and with that attempts to evoke anthropomorphism. Adding attention cues to induce joint attention is done by means of visual hierarchy (VH). It refers to the hierarchy used by the human eye to identify what is observed. The component that is most apart from their direct environment, is the one that draws the most attention.
Visual hierarchy is based on the gestalt principle. This principle explains that single elements can be grouped as one based on the following principles [44, 66]:

- Proximity: elements are perceived as a group if they are close together
- Similarity: elements are perceived as a group if they possess similar elements, such as colour, shape, form and orientation
- Continuity: points seem connected to each other if perceived in a connected, smoothly curving or straight line
- Closure: open curvy elements can be perceived as closed
- Common fate: elements are grouped together if they move in the same direction at the same speed

These principles are of great importance in (interface) design [53, 66]. The link between visual hierarchy and the gestalt principle can be found in the arrangement of elements.

Grouping and representing elements, can make or break a design. This is in congruence with research by Altaboli et al. [2] which investigated the level of symmetry by means of unity in size and number of objects. The results show that high level of symmetry, unity in size and number of objects, leads to increased visual aesthetics, which is described as the beauty or pleasant appearance of something.
To draw attention, a stimulus must be set apart from its surroundings, because an element that is most distinct is noticed first. This can be done by creating a stimulus that is different from the rest, thereby breaking the gestalt principle. The visual field processes visual information in a certain order, showing that some elements are noticed earlier in comparison to others. Visual hierarchy design principles take this into account and push focus in a wanted direction to reach a certain goal [23]. Shifting the focus to a specific object is done by creating contrast. The object that has the highest contrast is processed first.
Contrast can be created by means of four different characteristics that are used by the human brain to differentiate between objects [23]:

- Colour. Contrast in colour shows that a light-coloured object is noticed earlier on when they are displayed on a dark background. It includes the saturation, hue, perceived texture and value of forms.
- Size. Contrast created by size shows that objects which are displayed bigger are noticed earlier when compared to smaller objects.
- Alignment. Alignment can create contrast by means of orientation, pattern, or direction of objects relative to each other. Objects placed at the top are noticed earlier in comparison to lower placed elements. Additionally, elements are noticed earlier when they are displayed in isolation.
- Character. Character contrast entails contrast by means of (pattern) shape. When a pattern is more complex, it draws more attention, whereas this is not the case for predictable, simple patterns.

This distinction is in line with the way the brain processes visual information. There are two types of cells that each process different retinal information (light that is observed with the eyes): Smaller P ganglion cells and larger M ganglion cells. The P cells process form, brightness and colour (colour, size, character), while the M cells process information about motion perception (alignment). This information is further processed by other brain areas by means of the so-called ventral stream and dorsal stream. The ventral stream is responsible for form and object recognition (colour, size, character). Object recognition entails identifying shape, colour and texture. The dorsal stream processes motion and (positional) relationships between objects (alignment) [67].

3.2.5.3 ARAPPN-V implementation

In conclusion, attention cues are added to ARAPPV by means of visual hierarchy to create ARAPPN-V. They are placed in contrast to their surroundings, by means of the above characteristics and should engage the user in joint attention, which could prompt anthropomorphism. The more noticeable the attention cues, the more likely the engagement in joint attention.
The attention cues are added to direct the user attention to what is verbally communicated. An example for the addition of attention cues are the notification colours in RAPP. As explained before, RAPP makes use of notifications by means of colour when an action is not complete (see figure 3.1). Zooming is on these notifications; these can be adjusted to attention cues, to support what is verbally communicated. When the text shows "OPENSTAAND", the orange box can blink in support of the verbal message.
Another example is the addition of an check mark, which has a contrasting colour and appears to be blinking. A white check mark is shown when an action is completed in RAPP. The user must press the check mark to return to the main menu. In ARAPPV a verbal message is presented to support this check mark. In ARAPPN-V this check mark is an attention cue with a different, contrasting background colour [2]. The check mark is placed at the top of the page at the right corner. According to research, this location is noticed early on [21, 23]. Additionally, the check mark is blinking/moving. This is in line with research, which explains that when a stimulus is moving in the attended direction and has an irrelevant colour, neural activity of the colour-selective region will increase quickly. This is however not the case when the stimulus is moving in an unattended direction [37, 67]. In turn, the moving check mark aims to direct the user's attention to the check mark (attended location) and with that prompts to stimulate joint attention and thereby anthropomorphism.
Figure 3.7 displays the added non-verbal cues to ARAPPN-V. A complete list of non-verbal adjustments to create ARAPPN-V is displayed in appendix C.

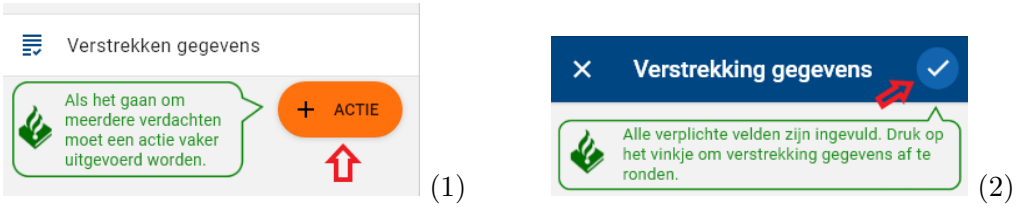


Figure 3.7. An enlargement of the added non-verbal cues to ARAPPN-V. Image 1 shows the non-verbal cue (different colour of "actie" button, see red arrow) before "verstrekken gegevens" is started and image 2 shows the non-verbal cue after "verstrekken gegevens" is completed (the check mark has a contrasting blue background colour, see red arrow).

[2] The following website is used to extract contrasting colours: https://color.adobe.com/nl/create/color-wheel

As of yet, no research is found that can support the link between joint attention and anthropomorphism. However, since joint attention requires empathy, it could be that when an interface shows attention cues to direct attention to a stimulus, the user might engage in this joint attention. This would mean that the user empathises with the interface, which could lead to anthropomorphism.
Additionally, since joint attention facilitates in language learning, it could also facilitate in learning how to use RAPP. If this is the case, it is in support with the effectance motivation mentioned earlier. This because the attention cues support the understanding of complex situations. Anthropomorphism might be a result.

To clarify the differences between all three prototypes, the table 3.2 below provides a summarised overview and figure 3.8 provides additional visualisation.

Table 3.2. A schematic overview of the difference between RAPP, ARAPPV and ARAPPN-V

Difference between RAPP, ARAPPV and ARAPPN-V			
	RAPP	**ARAPPV**	**ARAPPN-V**
Anthropomorphism	No anthropomorphic cues	Anthropomorphising by means of verbal cues	Anthropomorphising by means of non-verbal cues
Goal	Handle a shoplifting incident	Instruct the user as if communicating with a colleague	Support verbal cues as if interacting with a colleague
Design elements	No added design elements	Instruction messages as speech-clouds	Attention cues as blinking elements

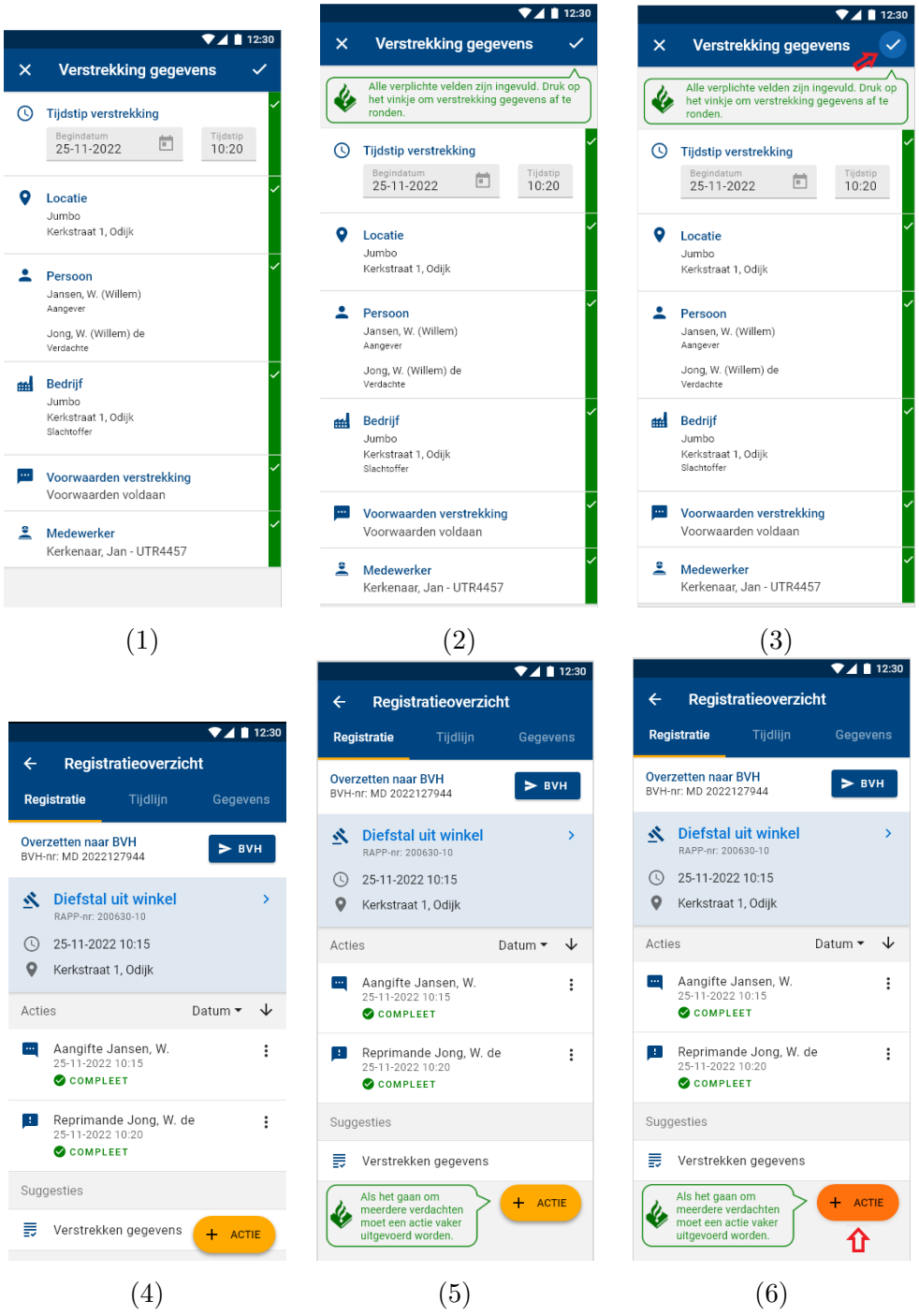


Figure 3.8. A visualisation of the differences between the prototypes in three different situations. The images 1 & 4 represent RAPP with no adjustments. The images 2 & 5 represent ARAPPV with verbal cues and the police logo. The images 3 & 6 represent ARAPPN-V with the non-verbal cues (see red arrow). The non-verbal condition switches between the images 2 & 3 and 5 & 6 to produce the blinking effect.

4 Method

This section will discuss the used method for this research. The subsections experiment design, procedure, materials, participants and analysis will be elaborated to clarify the steps which are required to answer the research question.

4.1 Experiment design

Over a course of six days, a total of 13 participants took part in this study. A within-subjects design was used to measure the differences in individual participant characteristics. The conditions are as follows:

1. RAPP as it is currently designed (control condition)
2. RAPP which is adjusted by means of text input to make it more "human"
3. RAPP which is adjusted by means of non-verbal input

For each condition, a prototype was built using Adobe XD. But only for the second (ARAPPV), and third condition (ARAPPN-V), the RAPP interface is humanised.
To account for priming and learning effect (internal validity), participants were presented with the conditions in different order (see appendix A).
Results were analysed both qualitatively and quantitatively.

4.2 Procedure

An overview of the step by step process is provided in figure 4.1.
At the start of the experiment, the participant was informed about the procedure of the experiment (A in 4.1). It was highlighted that it was allowed to stop with the experiment without giving a reason and consent was asked to use the data. The participant was not informed about the nature of the study to avoid interference with the results.
If there were no further questions a short questionnaire was presented with demographic questions and questions about technology in general. When this was completed, the case study was handed out. Next, the screen recording was started on MS Teams as well as on the phone and the phone with the corresponding condition (prototype) was handed to the participant (B in 4.1). Both the ARAPPV and ARAPPN-V prototypes contained

an "Ondersteuning?" on/off switch at the RAPP home page to control the support cues. It was explained to the participants that this button could not be manipulated in this study, since the support/ondersteuning cues were researched. The subject was asked to process a burglary (case study) while using the phone. During this task, the working process was monitored and the participant was asked to think aloud.
Processing the burglary case in this investigation was done in three steps. First, the participant had to start with "Aanhouding" to register the suspect. After that, the participant needed to check if the suspect was eligible for "Reprimande", which turned out to be the case. Lastly, the officer had to assign a prosecutor and transfer the data of the perpetrator to the shop (verstrekken gegevens). All the details necessary to come to these conclusions were provided in the case study.
When the burglary declaration process was completed for the first condition, a questionnaire was provided by the researcher to measure SUS scores, anthropomorphism, credibility and trust (C in 4.1).
After this, the participant had a two minute break (D in 4.1). This break was meant to take the mind of things and do something totally different. The participant was presented with a "Where is Waldo?" book or a rebus and was able to choose between the two. When the break was finished, the participant was asked to continue with the second prototype if able. When the participant was able to continue, the second condition started. The above process continued again.
When the questionnaire for the last condition finished, the researcher asked about the preferred condition and the need for an on/off switch. To conclude the experiment, the participant was questioned if there were any further comments or questions and after that was debriefed and thanked for participation (E in 4.1).
The whole experiment took about 30-50 minutes.

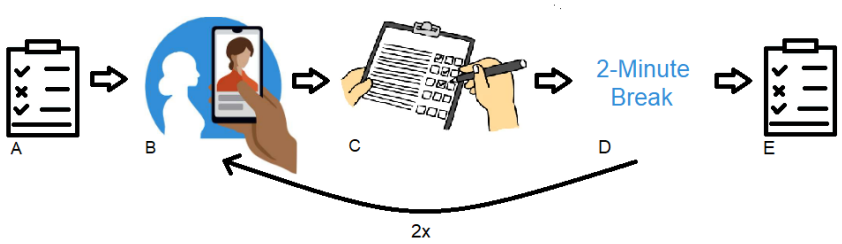


Figure 4.1. A schematic overview of the experiment process. A) briefing B) interaction process C) questionnaire D) two minute break E) debriefing. The steps B-D are the same in each condition and are repeated two times (a total of three).

4.2.1 Questionnaire

This research made use of two online questionnaires created by means of Qualtrics. The first questionnaire contained questions about feelings towards technology in general. The second one measured usability, anthropomorphism, credibility and trust. This

questionnaire was presented three time after each interaction with one of the three prototypes.
Before the experiment started, the participant was asked to fill out the questionnaire with demographics questions and questions about feeling towards technology in general. Demographic questions included age, years of service, RAPP usage and home-base of operation. These questions aloud to distinguish between experience with RAPP. More experienced police officers might rate RAPP differently, because of earlier frustrations. Longer in service means more experience with the system. This could mean that less practice is needed to learn the systems' use.
The questions about feelings towards technology in general were included to investigate whether the answers could be influence by the user's experience with technology. It could be that a participant who feels negative about technology might rate a question less positive and vice versa. The questions were based on research by Munnukka, Talvitie-Lamberg & Maity [58] and contained a 5-point Likert Scale.
The second questionnaire, which was conducted at the end of each interaction with RAPP, ARAPPV and ARAPPN-V, measured the users' experience while interacting with the prototypes. The questions were divided in two parts:
The first part of the questionnaire was focused on usability and task satisfaction. This is measured by means of so-called SUS scores (System Usability Scale). SUS was invented by John Brook and entails 10 statements which are evaluated by means of a 5-point Likert scale. A SUS score is an average score between 0 and 100, where 100 is the highest score. A score of 71 is generally seen as good. Usability should be measured by means effectiveness, efficiency and satisfaction [9, 10]. The statements in the SUS score questionnaire are focused on evaluating these values.
Since the questionnaire is conducted at the Dutch National Police, it is translated to Dutch. This will improve reliability and validity [5, 27]. The National Police had already created such a translated version of the SUS questions, which is used in this research.
The second part includes statements about anthropomorphism, credibility and trust. Research shows that statements about system knowledge/competence, sociability and naturalness [65, 72], can be used to measure anthropomorphism. For this research, the 5-point Likert Scale from research by Bartneck et al. [5] was combined with the statements of Munnukka, Talvitie-Lamberg & Maity [58]. Small adjustments were made, since this current research focuses on interface design of a work system. The statements about human-like body language and voice were, therefore, left out.
For the 5-point Likert Scale questionnaire about trust and credibility, research by West [81] was used. Credibility was measured by means usefulness (is the system telling the whole story), integrity, accuracy and honesty. For trust the participants was asked to rate if the system stated all the facts, was unbiased and respectful.
The complete questionnaire is displayed in appendix B.

4.2.2 Variables

This research measures the influence of anthropomorphism, applied by adding subtle features to an interface, on usability, credibility, trust and task efficiency. The independent variable is the manipulation of ARAPPV and ARAPPN-V by means of non-obvious, subtle, human-like adjustments.
The dependent variables are anthropomorphism, usability (SUS score), credibility, trust and task efficiency, all measured by means of 5-point Likert Scale questions except for task efficiency, which was measured by means of two qualitative interview questions at the end of the experiment: Which prototype did you prefer? & Why did you prefer that prototype?

4.3 Materials

The following materials were used:

- Laptop to record and monitor the research process (Asus VivoBook)
- MS Teams to monitor and record the research process
- Ipad to present the questionnaires (Ipad Air 2)
- Questionnaires as explained in section 4.2.1
- Phone to work with the prototypes (Samsung Galaxy S22 5G & Huawei P30)
- Prototype 1 with normal RAPP (RAPP, Normal), created by means of Adobe XD and presented on the phone's web-browser
- Prototype 2 with an adjusted RAPP (ARAPPV, Verbal), created by means of Adobe XD and presented on the phone's web-browser
- Prototype 3 with an adjusted RAPP (ARAPPN-V, Non-verbal), created by means of Adobe XD and presented on the phone's web-browser
- Where is Waldo? book and rebus to use during the break
- Software to do statistical tests (SPSS)

4.4 Participants

This experiment was conducted with 13 participants, which were quite difficult to recruit. One female and twelve male police officers between the ages of 26 and 54, stationed across the Netherlands were participating. All participants had experience with RAPP, as most of them use it in their daily work, which is "on the street". The majority of participant was very experienced in RAPP and 5 participants were less experienced. It was attempted to select the participants as random as possible by means of random sampling. An email was send to each district to ask for participation. This way, every member of the population had an equal chance of being selected. Two participant were selected from each district. They had to reply to the email to participate. The email did not contain information about the nature of the study to avoid bias.

4.5 Analysis

4.5.1 Data processing

Due to malfunctioning on the first test day, one participant did not participate in the non-verbal condition. This data was excluded from analysis. Therefore, a total of 12 participants participated in all three conditions (normal/RAPP, verbal/ARAPPV, non-verbal/ARAPPN-V). To convert the response for each question into one variable (anthropomorphism, trust, credibility), the median of the corresponding questions was extracted for each participant in every condition separately, resulting in three variables per participant per condition. The median was chosen instead of the mean, to take into account the fact that participant might interpret the values for a Likert Scale differently, e.g., participant one could rate a question with a 4, but for participant two (who feels exactly the same) it might be rated as a 3.
The data gathered to measure anthropomorphism, credibility and trust did not fulfil the parametric requirements, therefore a non-parametric test was chosen. Since the data was retrieved using a Likert Scale questionnaire, for a small sample size ($N = 13$) and by means of a within-subject design, a Friedman test was chosen to compare the three conditions. Data was processed by means of SPSS and an alpha level of 0.05 was used. The RAPP condition was used as a base case and ARAPPV and ARAPPN-V were compared to measure the difference in anthropomorphism and to extract the influence of anthropomorphism on credibility and trust.
To measure the statistical difference for usability, a repeated-measures ANOVA was used with an alpha level of 0.05. Again, the RAPP condition was used as a base case and ARAPPV and ARAPPN-V were compared to measure the difference in usability.

4.6 Pilot study

A pilot study was conducted to test if the experiment setup was sufficient to conduct the final experiment. The pilot study was used to adjust the experiment. For this experiment 2 participant were asked to participate. One participant had little experience with RAPP the second participant was involved in the creation of RAPP and used it while working. Both participants had experience with evaluation studies. Participant one conducted the experiment while participant two was observing.
After the pilot study the questionnaire was split in a demographics part and an evaluation part as is visible in appendix B. Additionally, a view small adjustments in the verbal cues were added to make sure that the text was accurate and in accordance with the actual RAPP.
Lastly, it was decided that a second phone was necessary to conduct the experiment, which made use of its one Wi-Fi-connection.

5 Results

The quantitative and qualitative results found in this research are presented in the following section.

5.1 Quantitative

5.1.1 System Usability Scale

To measure usability, the System Usability Scale (SUS) was used. The results are displayed in table 5.1 & figures 5.1 & 5.2. The highest SUS score was found for ARAPPV (verbal), a score of 82.5. The second highest for ARAPPN-V (non-verbal), namely 79.0, and the lowest score was found for RAPP (normal), a score of 74.4. An average SUS score of 71 in this field of research means a good interface [4].
A repeated-measures ANOVA determined that the differences between the mean SUS-scores were not significant ($F(1.4, 22) = 5.699, p = 0.17$).

Table 5.1. Overview of the average SUS scores in each condition.

Average System Usability Scale	
	Score
RAPP	74.423
ARAPPV	82.5
ARAPPN-V	78.958

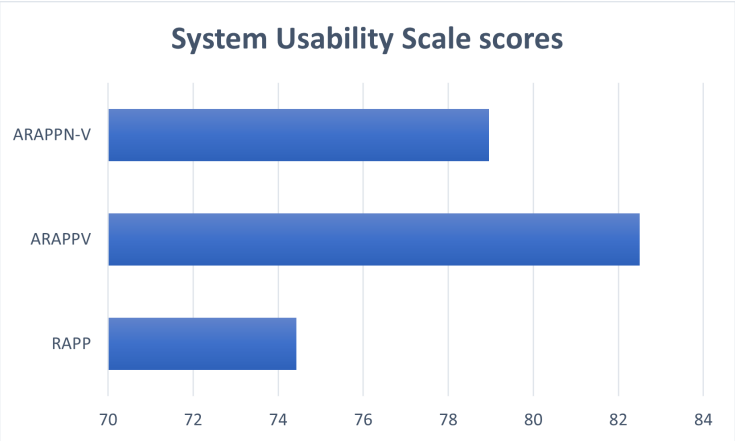


Figure 5.1. Bar chart representing average SUS scores in each condition

5.1.2 Inferential statistics

To measure the difference in anthropomorphism, credibility and trust a Friedman test was conducted. Figure 5.3 contains a detailed overview of the inferential statistical results.

5.1.2.1 Anthropomorphism

Table 5.3 displays the influence of anthropomorphism on RAPP. The results show that anthropomorphism does not significantly change for the ARAPPV (verbal) and ARAPPN-V (non-verbal) condition ($X^2(2) = 5.636, p = 0.060$).
Although p is not significant, the effect sizes for these comparisons could still be useful to see the magnitude of group differences. Therefore, a Post Hoc, Wilcoxon Signed-Rank Test was conducted, as recommended by Field [26]. Since p is almost significant, it could also establish if there was a condition in which a significant difference could be found. The results are shown in table 5.2.
Based on the Wilcoxon Signed-Rank test, it appears that anthropomorphism did significantly change from RAPP to ARAPPV ($p = 0.035, T = 32, r = 0.585$). However, it does not significantly change from RAPP to ARAPPN-V ($p = 0.527, T = 17.5, r = 0.175$), and from ARAPPN-V to ARAPPV ($p = 0.059, T = 24, r = 0.524$).
Strengthened by the relatively high effect size, it can be concluded that it seems like anthropomorphism increases for ARAPPV compared to RAPP. The difference between ARAPPV and ARAPPN-V is however not significant.

Table 5.2. Statistical significance of increase in anthropomorphism measured by a one-sample test between the difference conditions RAPP, ARAPPV and ARAPPN-V. Order of comparison was based on the mean ranks assigned by the Friedman test: RAPP - 1.71, ARAPPN-V - 1.83, ARAPPV - 2.46. No significant difference was found at an alpha level of 0.05.

Wilcoxon Signed-Rank Test		
Comparison	**p-value**	**Accept Hypothesis**
RAPP / ARAPPV	0.035	Yes
RAPP / ARAPPN-V	0.527	No
ARAPPN-V / ARAPPV	0.059	No

5.1.2.2 Credibility

In table 5.3 the influence of anthropomorphism on credibility is presented. The results show that credibility does not significantly change for ARAPPV (verbal) and ARAPPN-V (non-verbal) ($X^2(2) = 1.152, p = 0.562$). Therefore, no follow up test was conducted. It can be concluded that it seems like verbal and non-verbal cues do not have an influence on credibility in this case.

5.1.2.3 Trust

Table 5.3 contains the results found for the effect of anthropomorphism on trust. The results show that trust does not significantly change for ARAPPV and ARAPPN-V ($X^2(2) = 1.471, p = 0.479$). Again, no follow up test was conducted. It can be concluded that, in this case, it seems like verbal and non-verbal cues do not have an influence on trust.

Table 5.3. Statistical significance of increase in anthropomorphism and effect of anthropomorphism on credibility and trust measured by a two-way analysis over three difference conditions RAPP, ARAPPV and ARAPPN-V. No significant difference was found at an alpha level of 0.05.

Friedman's Two-Way Analysis of Variance		
	p-value	**Accept Hypothesis**
Anthropomorphism	0.060	No
Credibility	0.562	No
Trust	0.479	No

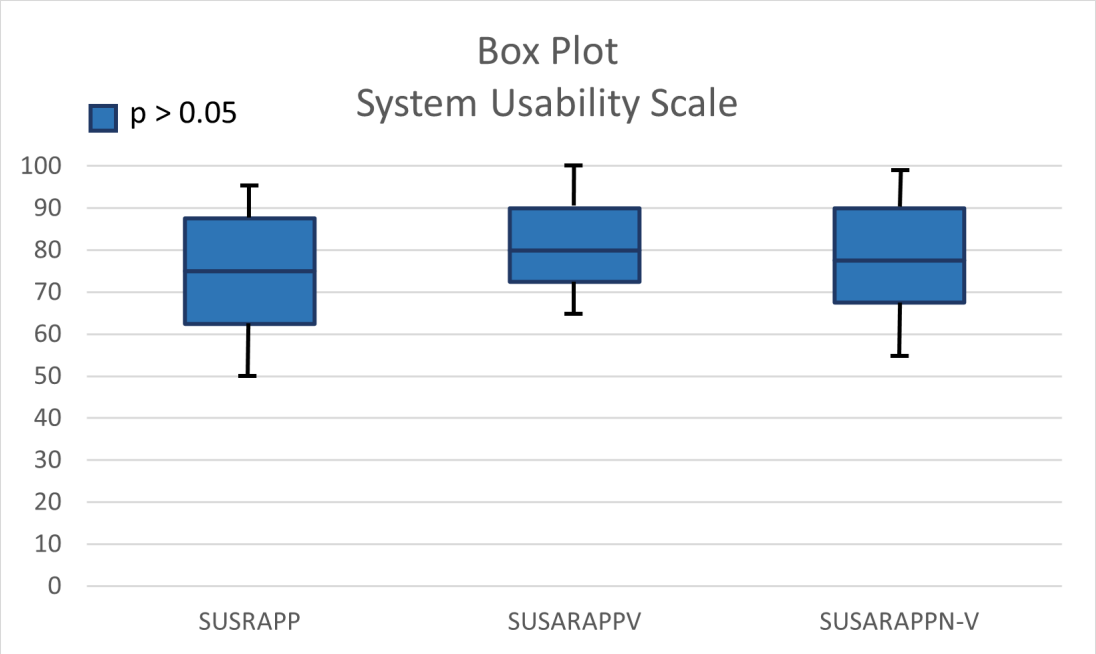


Figure 5.2. Box plot of the between-group difference in average System Usability Scale score. The p value was calculated using a repeated-measures ANOVA. No significant difference was found at an alpha level of 0.05.

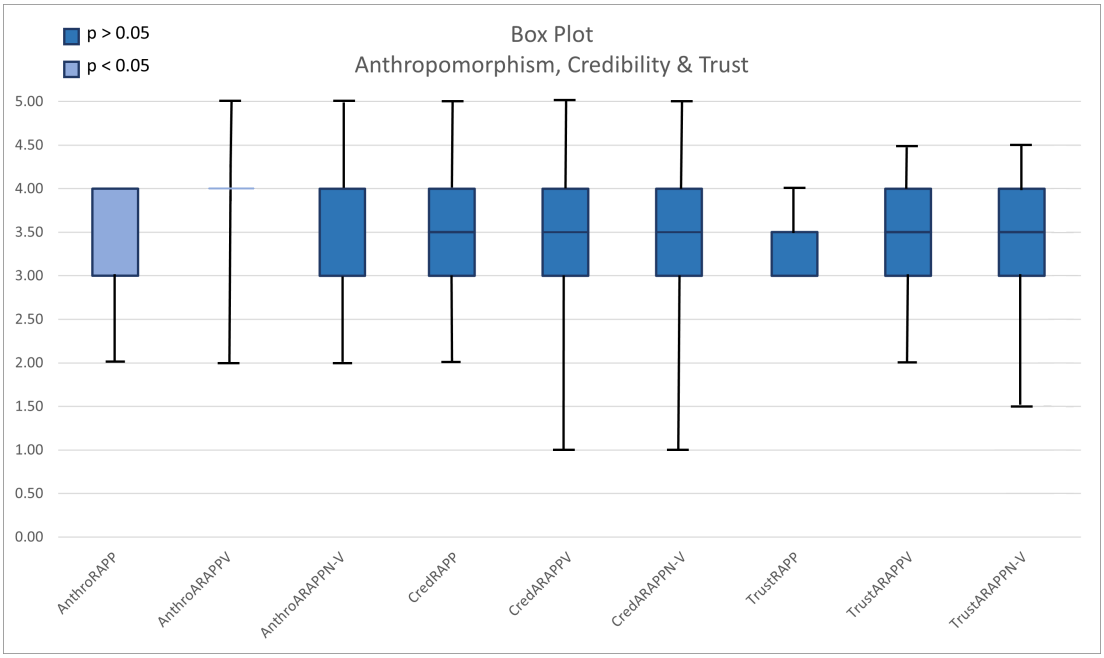


Figure 5.3. Box plot of the between-group difference in mean scores for anthropomorphism, credibility and trust, respectively. The p values were calculated using a Friedman's test. No significant difference was found at an alpha level of 0.05 for anthropomorphism, between the RAPP & ARAPPN-V and ARAPPN-V & ARAPPV condition and for credibility and trust between the RAPP, ARAPPV and ARAPPN-V conditions. There was a significant difference at an alpha level of 0.05 for anthropomorphism between the RAPP & ARAPPV condition.

5.1.3 Descriptive statistics

The boxplot in figure 5.3 displays the descriptive results found in this experiment, since no significant difference could be found for the inferential statistics.

5.1.3.1 Anthropomorphism

For anthropomorphism, the result show that the median in ARAPPV (verbal) is higher, 4 in comparison to 3 in RAPP (normal) and ARAPPN-V (non-verbal), as well as the maximum in both ARAPPV and ARAPPN-V, 5 compared to 4 in RAPP. The minimum did not change between the different conditions.

5.1.3.2 Credibility

For credibility the minimum decreased from 2 in RAPP, to 1 in both ARAPPV and ARAPPN-V. The maximum did not change between conditions. The median increased from 3.5 in RAPP, to 4 in both ARAPPV and ARAPPN-V.

5.1.3.3 Trust

For trust the minimum decreased, from 3 in RAPP to 2 in ARAPPV and 1.5 in ARAPPN-V. Contrasting, to the maximum which increase from 4 in RAPP, to 4.5 in both ARAPPV and ARAPPN-V one. The median in did not change between conditions.

5.2 Qualitative

5.2.1 Condition preference in support of task efficiency

Overall, the feelings towards RAPP (normal) in this experiment were neutral and the feelings toward ARAPPV (verbal) were positive. However, the feelings towards ARAPPN-V (non-verbal) were mostly negative.

Starting with ARAPPN-V, only half of the participants was able to notice the difference between ARAPPV and ARAPPN-V. Of the other half, 4 participants found the blinking elements to be annoying and/or too compelling. One participant explained that for himself, the blinking was unnecessary, however, he could imagine others might like it. A second participant preferred ARAPPN-V and said that he felt even more guided by the system in comparison to ARAPPV. He even got confused while using the normal RAPP.
Secondly, out of 13 participants, 7 preferred RAPP and said they did not need the instructions provided by ARAPPV. However, all 13 participants claimed that the instructions would be very useful for participants with little to no experience in RAPP.

One participant even mentioned that the given instructions were exactly where the problem was:
"Ik zie daar de tips in staan, waar ik voor wordt gebeld[1]."

Another participant mentioned that ARAPPV provided:
"Een steuntje in de rug[2]."

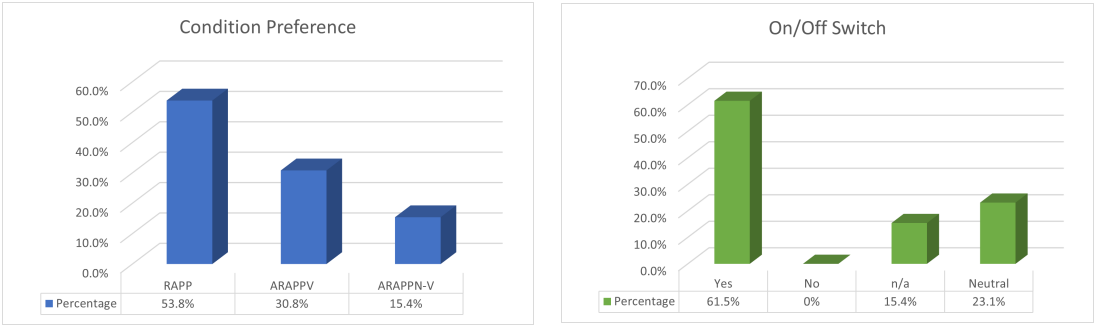
One participant found the content of the instructions to be too belittling/patronizing, although supportive and someone claimed that the text got in the way. Two other participants say there was too much text which made it hard to read. Still, after reading they concluded that the instructions were useful.
Contrasting, one of the participants said that he felt that more text could be used. Especially, since the instructions were mostly added at the beginning and the end.

[1] Translation: These instructions are the kind of questions officers call me for.
[2] Translation: A helping hand

Because of this, extra instructions in the middle were missed. Additionally, it was explained by a participant that the text was no cause of frustrations, since it did not get it the way. Another participant mentioned that the text was short, not belittling/patronizing, but precise, exactly as police officers like to see it.

All in all, the feelings towards ARAPPV are positive and it is the preferred condition for 4 participants. It would be helpful to support users with less experience is and it could improve task efficiency. For experienced users it would be nice if the (non-)verbal cues could be turned off by means of an on/off switch. To the question if this switch was necessary, 8 participants answered yes. Two participants were neutral: one participant highlighting that the text did not bother him, so it was not specifically needed. The other participant explained that it was not needed, but if the option was there he would choose it. Two participants did not answer the question since it was not asked.

An overview of the overall condition preference and the need for an on/off switch is represented in figure 5.4.



(A) (B)

Figure 5.4. Bar chart representing A) the distribution of the preferred condition. A total of 7 participants preferred RAPP, 4 participants voted for ARAPPV, of which one participant did not see the difference between ARAPPV and ARAPPN-V and 2 participants rated ARAPPN-V highest from which one participant did not see the difference between any of the conditions but choose ARAPPN-V because he was the fastest in this condition and B) the need for an on/off switch for the verbal and non-verbal cues.

Four other opinions were expressed by different participants. Although not all participants expressed these thoughts, these participants claimed to speak for other colleagues. Their opinions are stated here:

- Arrows and check marks are used interchangeably, which causes uncertainty. Intuitively, an arrow means go back, however, in this context it means continue. A check mark represents done, continue. The use of check marks exclusively, would be preferred.
- There is more background information necessary to support the officer during their work. RAPP should be a source of information, providing the information which the officer would otherwise google/look up in the police system. Par example: when the officer communicates the rights (cautie meedelen), it is not explained what happens when the suspect does talk. Secondly, the victim also has rights, but which rights? This is the kind of information RAPP should provide as well.
- RAPP is too much intertwined with BVH. BVH is a system that needs to be fed information. It feels like by using RAPP, we are still stuffing BVH with information.
- The key to understanding how to use RAPP is by using it. However, officers experience problems while using it and are therefore less eager to use it. These problems are caused by the fact that sometimes the servers/systems, on which RAPP is running, fail. To get officers to use RAPP more often, means restoring faith in the system. First make sure that the system does not fail, then see what happens.

6 Discussion

This research attempted to find a relationship between anthropomorphism, applied by adding subtle features to an interface, and usability, credibility, trust and task efficiency. This was done by means of adding verbal and non-verbal cues to the RAPP police system which is currently used. To the best of knowledge, this research explains the effect of non-obvious anthropomorphic trigger cues on an application that is used in a daily work environment.

6.1 Anthropomorphism

Results show that anthropomorphism significantly increases for ARAPPV (verbal condition) compared to RAPP (normal condition), however no significant difference was found between ARAPPN-V (non-verbal condition) & ARAPPV and RAPP & ARAPPN-V. Descriptive results however show that maximum rating indeed improves for ARAPPN-V compared to RAPP, however not for ARAPPN-V compared to the ARAPPV.
Since ARAPPV provides verbal support in complex situations, it increases elicited agent knowledge. Anthropomorphic knowledge is better accessible and available, hence more human-like. To make sense of a system such as RAPP, the effectance motivation can be used. Anthropomorphising RAPP reduces uncertainty and helps to predict future actions [68]. The verbal cues increase the likelihood to anthropomorphise because it is easier to attribute human characteristics to something that already feels more human [24]. As the qualitative result show, ARAPPV indeed provides more support for the police officer. The verbal cues (anthropomorphic knowledge) make ARAPPV the recommended version for police officers less experienced in RAPP.
The increase in anthropomorphism between ARAPPV (verbal) and RAPP (normal) can also be explained by the fact that communication is an anthropomorphic ability [22]. ARAPPV communicates tips and advice which are seen as verbal anthropomorphic cues [25]. This helps ARAPPV to be accepted in the human social circle [22].
The lack of improvement between ARAPPV and ARAPPN-V could be explained by a mismatch in the cognitive schema. It could be that participants expected a larger difference between ARAPPN-V and ARAPPV, and the lack of difference, a mismatch in schema, did not result in increased anthropomorphism [1]. This is strengthened by the fact that the quantitative results show that half of the participants did not see the difference between both interfaces.

The fact that half of the participants did not see the non-verbal cues and four of the participants that did see the differences said to be annoyed by them, shows that the attention cues did not result in joint attention. Joint attention means following gestures of the sender (ARAPPN-V) towards the attended [77]. The lack of joint attention, explains the lack of empathy, ergo the lack of anthropomorphic increase.
Another reason that could explain the difference between ARAPPN-V & RAPP and ARAPPN-V & ARAPPV is a disruption in the level of symmetry. When the level of symmetry is high something is described as beautiful and pleasant [2]. It could be that the non-verbal adjustments disrupted the symmetry resulting in a less pleasant appearance. Interface appearance does not necessarily has to contain human identity cues [25]. A less pleasant appearance of ARAPPN-V could thus negatively influence anthropomorphism. As the quantitative results pointed out, from the six participants that saw the changes in ARAPPN-V, four found them to be annoying, suggesting that it indeed negatively influences anthropomorphism. However, the other six participants did not see the difference, but rated ARAPPN-V less anthropomorphic than ARAPPV. This suggests that the lack of symmetry is not the only reason for these results.
The fact that the non-verbal cues did not stick out as much could be explained by the lack of contrast necessary to push focus in a wanted direction [23]. However, creating too much contrast would interfere with the pleasantness of the interface appearance even more, which could result in a more unpleasant experience. Another explanation is the fact that when using RAPP, participants need to finish a certain amount of steps to complete the process. The participants were too focused on completing the steps and checking the green/orange/red complete boxes which indicate that everything is filled out, that the workflow was in the way of noticing the non-verbal cues.

6.2 Usability

Figure 6.1 places the System Usability Scale score for the three conditions in perspective. RAPP, ARAPPV and ARAPPN-V are categorized as good systems. A large increase between the SUS scores of RAPP (lower bound of good) and ARAPPV (upper bound of good) & ARAPPN-V (middle bound of good) is visible in figure 5.2 [4]. Although this difference is not significant, it is still a large growth when it comes to SUS scores [47].
Therefore, this research found support for the following hypothesis:

H1: When a police officer anthropomorphises RAPP by means of non-obvious adjustments in behaviour and communication, this will lead to improved usability of RAPP.

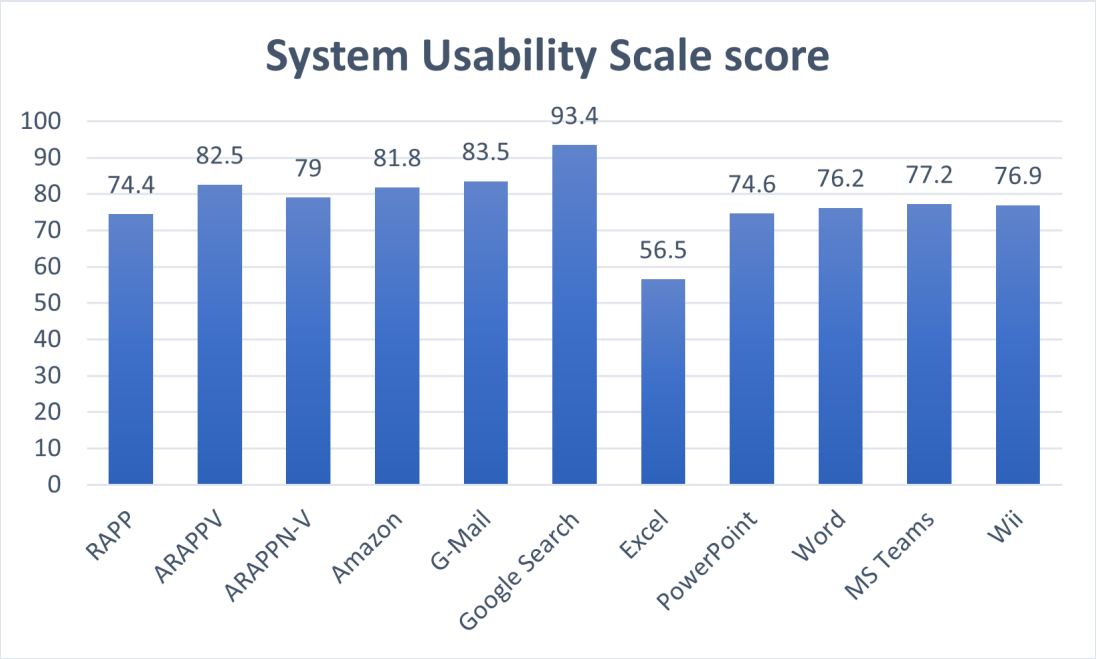


Figure 6.1. Graphical representation of average SUS scores for RAPP, ARAPPV and ARAPPN-V with respect to eight other average SUS scores extracted from research by Kortum & Bangor [46].

As the qualitative results point out, users experience the instruction messages as supportive. They guide the user through the system, which makes the system easier to use. This explains the increased SUS score for ARAPPV (verbal). It also partly explains the increased SUS score for ARAPPN-V (non-verbal). Since only half of the participants noticed any difference between the ARAPPV and ARAPPN-V, both conditions are the same for these participants. Taken together with the two participants of the other half, who said the non-verbal cues were supportive, this accounts for the slight increase in the ARAPPN-V SUS score.
The other four participants who found the non-verbal cues to be annoying, explain the fact that the SUS score for ARAPPV and ARAPPN-V are not the same, and the SUS score for ARAPPN-V is slightly less. The pleasant experience by means of improved usability, because of anthropomorphism, leads to an increased desire to use the system again [20, 68]. It seems likely that the addition of verbal cues to the interface of RAPP causes users to anthropomorphise, which improves usability.

6.3 Credibility

The result show that credibility does not significantly improve between the different conditions. The descriptive results show a decrease of 1 in the minimum, suggesting that credibility might even decrease. Nevertheless, a small increase of 0.5 in the median is also measured for ARAPPV (verbal) and ARAPPN-V (non-verbal) in comparison to

RAPP (normal), suggesting that although credibility is rated lower in some cases, it is rated higher more often. However, the difference is too small to safely say that it makes a difference. Therefore, this research did not provide support for the second hypothesis:

H2: Anthropomorphising RAPP by means of non-obvious adjustments in behaviour and communication will improve credibility in RAPP

Credibility can be measured by four aspects [28]:

- Device credibility: the physical appearance of the product
- Interface credibility: the interface design and interaction experience (usability)
- Functional credibility: what the product is able to do, which is also related to trust
- Information credibility: whether the output is believable or not

The small increase in credibility could thus be explained by an increase in interface credibility. Which is the result of anthropomorphic changes, as is visible in the increase in usability.
Since the phone (physical appearance) was the same in every condition, device credibility remains unchanged.
The verbal and non-verbal cues guide the user, while processing all the information and provide helpful instructions when necessary. This support could have caused an increase in informational credibility. As the qualitative results indicate, police officers show a positive attitude towards ARAPPV and the majority rates the instructions as supportive for less experienced users.
As RAPP is a new system, it could be expected that the need for information is high. Since ARAPPV and ARAPPN-V provide more information, it is expected that credibility should increase [36]. However, as the results show, this is not always the case. This could be explained by the fact that most participants in this study were familiar with RAPP, since they train students to use it. The supportive cues were unnecessary for these participants. The less experienced participants did show a preference for either ARAPPV or ARAPPN-V, showing the need for information. Nevertheless, the cues were of less importance to the majority of the participants explaining the findings for credibility.
Additionally, users who are familiar with the content will judge the system more harsh and are more likely to evaluate a system as less credible [40, 43], compared to a user who is less experienced and unfamiliar with the system, hence more likely to rate the system as more credible [36, 43]. This could explain the slight decrease in the credibility of ARAPPV and ARAPPN-V. Most participants were familiar with RAPP and therefore rated ARAPPV and ARAPPN-V more harsh.

Functional credibility could also explain the results found for ARAPPV and ARAPPN-V. People do not evaluate a face by means of appearance, instead they evaluate in by means of its competence or performance [45]. Police history shows that the servers do not always function optimally [17], especially BVH has its problems. RAPP is connected to BVH, and will eventually replace it. However, the distrust in BVH is still visible in the attitude towards RAPP, as is confirmed by the qualitative results. One participant explained that BVH and RAPP are to intertwined. Distrust in the system effects functional credibility. Luckily, it can be improved by making sure the servers do not fail. This is in line with another participant that mentioned to restore faith first by making sure the systems do not fail, then see what happens.
Although functional credibility might be low, it is important to note that a general credibility score of 4 out of 5 is already quite high and pure perfection is a hard thing to achieve.

6.4 Trust

The results show no significant difference to conclude that anthropomorphism by means of either verbal nor non-verbal cues increase trust. Although the descriptive results show that the maximum rating increased by 0.5, the median did not change over the different conditions and the minimum rating even dropped from 3 to 2 in ARAPPV and 1.5 in ARAPPN-V. This drop, however, does not mean that ARAPPV and ARAPPN-V were less trustworthy. ARAPPV and ARAPPN-V are simply not more trustworthy in comparison to RAPP. All in all, this research did not find enough evidence to support the third hypothesis:

H3: Anthropomorphising RAPP by means of non-obvious adjustments in behaviour and communication will improve trust in RAPP

Credibility is one of the predominant predictors for trust [59] and is an indicator for the degree to which a person is regarded as believable, competent and trustworthy [51]. The continuity in trust between the different conditions, can thus be explained by the small difference in credibility.
As the qualitative results indicate, faith in the system needs to be restored before the system can be trusted. Robot characteristics, especially system performance, are of great influence when it comes to trust [35]. The best way to gain trust is by providing users with the ability to give feedback regarding the reliability and situational factors that have an effect on the system's performance [39]. Fortunately, this is already happening more and more at the design department of the Dutch National Police.

6.5 Task efficiency

The qualitative results show that the verbal cues and to some extend non-verbal cues are indeed support for the user and help to guide the user through the system. Therefore, this research does find strong support for the fourth hypothesis in relation to the verbal cues, however no support regarding the non-verbal cues:

H4: Anthropomorphising RAPP by means of non-obvious adjustments in behaviour and communication will lead to higher task efficiency while using RAPP

Only half of the participants saw the difference between ARAPPV and ARAPPN-V. From the other half, four participants were annoyed by the blinking elements in ARAPPN-V, reducing task efficiency. The focus was directed towards the blinking elements, causing irritation. Attention should be on the task and not on the interface. Focusing on the interface reduces task efficiency [32]. However, two participants thought ARAPPN-V would be an efficient and helpful way to guide the user. Although this highlights the differences in user preferences, it is not enough evidence to support the hypothesis.
On the other hand, all participants show a positive attitude towards the verbal cues in the ARAPPV condition. All participants agreed that the support cues are helpful for less experienced users. The verbal cues provide support and explain what the police officer needs to do in a certain situation to avoid doing the wrong thing. The officers spend less time thinking about what should be done. The attention is drawn to the task that needs to be completed and not to the interface design. Therefore, ARAPPV supports the interaction with the presented information. This allows the user to move quicker since the officer does not need to look things up in the police system or call the help-desk for an explanation. Hence, ARAPPV increases free time, time to spend on other tasks, because it demands less attention [32].
Lastly, the anthropomorphic, verbal cues enhance social presence, which in turn improves performance [75]. The results suggest that ARAPPV is perceived as more human-like, anthropomorphic. As if collaborating with a police partner. This in turn improves task efficiency.

6.6 Limitations

This research poses several limitations starting with the sample size which was relatively small. This made it hard to draw significant conclusions.
Secondly, the same case study was used in every condition. A different case study for each condition could have resulted in different results. Participants were familiar with the case study after the first condition and knew what steps to take. This could be the reason that participants did not see the difference between ARAPPV and ARAPPN-V, because they were moving too fast.
Thirdly, most of the participants were experienced RAPP users, while a lot of police officers in the field do not use RAPP as much. Nevertheless, the result still show that ARAPPV could be of great help for less experienced users. Additionally, RAPP is going to replace the current system BVH. So, eventually, all police officers will be experienced RAPP users. This research looks ahead and provides advice to take into account when further improving RAPP.
Fourthly, the sampling method used required the participants to respond to participate in the study themselves. It could be that this affected the variety of the sample group and only people who were very willing to participate responded, for example, because they had the most to complain or like to be well informed. One thing of note is that the experienced participants knew a lot about the struggles of less experienced users, since they are the ones that are contacted about these struggles. So, even though they are more experienced and willing to participate, they can still make a good assessment for other users.
Lastly, on the first day of testing the police phone turned out to be insufficient. It was a secure police phone on which only one specific web-browser could be used. This browser was not compatible with Adobe XD and full screen-mode was not allowed. Therefore, participant 3 did not participate in the non-verbal condition. The construction of this prototype in this browser made refreshing the page too easy, causing the participant to be "kicked-out" of the prototype and needing to start all over. One of the other two participants that day mentioned this effected the response to the survey questions. When looking at the results for this user case in more detail, it did not account for the lower rating in trust and credibility. A normal, non-police phone was used as replacement for the remaining test days.

7 Conclusion

7.1 Future work

Future research should include eye tracking to investigate if the blinking elements truly go unnoticed and/or if they might influence participants unconsciously. In addition, EMG scans can be used to monitor if anthropomorphism has an influence on emotional change which may positively affect the user.
The current RAPP prototypes were built in a task-oriented way. Future research should explore if social dialog is more beneficial in building trust in a daily work system. This could be done by increasing expressiveness by means of conveying feelings and emotions in response to user actions. Making an interface more sociable, might also make it less professional which may not be desirable in a work environment.

7.2 Conclusion

This research attempted to find an answer to the following research question:

Can non-obvious human-like changes in behaviour and communication to the interface of RAPP improve usability, credibility, trust and task efficiency?

This was investigated by creating three different prototypes RAPP, ARAPPV and ARAPPN-V. RAPP was used as base case, ARAPPV contained verbal cues, whereas ARAPPN-V consisted of non-verbal cues. Results indicate that anthropomorphism indeed improves usability for ARAPPN-V and even more so for ARAPPV. Furthermore, the results show that task efficiency improved in the ARAPPV condition. Although, descriptive results might suggest that there is a small improvement for credibility and trust, there were no significant results found to support this.
Qualitative results indicate that to improve credibility and trust in RAPP, the stability of the system needs to be restored first.
Participants show support for ARAPPV (verbal condition) to guide less experienced RAPP users. Almost all participants mentioned that the implementation of an on/off switch for the verbal cues would be of additional value to avoid irritation for long term usage.

All in all, this research provides evidence that non-obvious anthropomorphism, by means of subtle, verbal cues, prompts anthropomorphism which in turn improves usability and task efficiency of RAPP. Descriptive statistics show that to some extend trust and credibility are improved.
These results add to Conversational Agent (CA) research, since verbal cues are used in a more formal capacity which, in this context, is more subtle, thus less obvious in comparison to existing research. Anthropomorphism by means of subtle, verbal cues can thus be a critical characteristic when designing a daily working interface.

7.3 Recommendation

As this thesis was written in order of the National Police, I was asked to provide recommendations, if possible, which might be taken into account when further implementing RAPP. Based on the results found in this research, I would suggest the following:

- The addition of verbal cues (as described in this research), since this could be a helpful tool to get to know RAPP and learn how to use it.
- Do not use the non-verbal cues, because for most of the participants this caused irritation. However, if there are better ways to adjust it, it could be of additional value.
- Only add the verbal cues if an on/off switch is implemented as well. This way, experienced RAPP users are not bothered by the cues and users can adjust the RAPP interface how they want. It is recommended to safe the user settings, so the users does not need to switch on/off every new login. RAPP should remember the last settings.
- Investigate if other verbal cues could be implemented which contain additional background information, e.g. what are the victim's rights.
- Adjust arrows to check marks only. This stimulates consistency and reduces confusion.

Bibliography

[1] Pankaj Aggarwal and Ann L. Mcgill. "Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products". In: *Journal of Consumer Research* 34.4 (2007), pp. 468–479.

[2] Ahamed Altaboli and Yingzi Lin. "Effects of unity of form and symmetry on visual aesthetics of website interface design". In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 56. 1. SAGE Publications Sage CA: Los Angeles, CA. 2012, pp. 728–732.

[3] Michael Argyle. *Social interaction*. Routledge, 2017.

[4] Aaron Bangor, Philip Kortum, and James Miller. "Determining what individual SUS scores mean: Adding an adjective rating scale". In: *Journal of usability studies* 4.3 (2009), pp. 114–123.

[5] Christoph Bartneck et al. "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots". In: *International journal of social robotics* 1.1 (2009), pp. 71–81.

[6] David Benyon and Oli Mival. "Landscaping personification technologies". In: (2008), pp. 3657–3662.

[7] Elizabeth Broadbent, Rebecca Stafford, and Bruce MacDonald. "Acceptance of healthcare robots for the older population: Review and future directions". In: *International journal of social robotics* 1.4 (2009), pp. 319–330.

[8] Cyril Brom, Tereza Starkova, and Sidney K D'Mello. "How effective is emotional design? A meta-analysis on facial anthropomorphisms and pleasant colors during multimedia learning". In: *Educational Research Review* 25 (2018), pp. 100–119.

[9] John Brooke et al. "SUS-A quick and dirty usability scale". In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7.

[10] John Brooke. "SUS: a retrospective". In: *Journal of usability studies* 8.2 (2013), pp. 29–40.

[11] Jerome Bruner. "Child's talk: Learning to use language". In: *Child Language Teaching and Therapy* 1.1 (1985), pp. 111–114.

[12] John-John Cabibihan et al. "Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism". In: *International journal of social robotics* 5.4 (2013), pp. 593–618.

[13] Chayan Chakrabarti and George F Luger. “Artificial conversations for customer service chatter bots: Architecture, algorithms, and evaluation metrics”. In: *Expert Systems with Applications* 42.20 (2015), pp. 6878–6897.

[14] Veena Chattaraman et al. “Should AI-Based, conversational digital assistants employ social-or task-oriented interaction style? A task-competency and reciprocity perspective for older adults”. In: *Computers in Human Behavior* 90 (2019), pp. 315–330.

[15] Yun Chen et al. “A method for exploring word-colour associations”. In: *Color Research & Application* 45.1 (2020), pp. 85–94.

[16] Sonia Chiasson and Carl Gutwin. “Testing the media equation with children”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2005, pp. 829–838.

[17] Landelijk CIO-office. “Aanvalsprogramma Informatievoorziening Politie 2011-2014”. In: november (2011).

[18] Andrew J Cowell and Kay M Stanney. “Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility”. In: *International journal of human-computer studies* 62.2 (2005), pp. 281–306.

[19] Antonella De Angeli, Graham I Johnson, and Lynne Coventry. “The unfriendly user: exploring social reactions to chatterbots”. In: *Proceedings of the international conference on affective human factors design, london*. 2001, pp. 467–474.

[20] Marjorie Delbaere, Edward F. McQuarrie, and Barbara J. Phillips. “Personification in advertising: Using a visual metaphor to trigger anthropomorphism”. In: *Journal of Advertising* 40.1 (2011), pp. 121–130.

[21] Soussan Djamasbi, Marisa Siegel, and Tom Tullis. “Visual hierarchy and viewing behavior: An eye tracking study”. In: *International conference on human-computer interaction*. Springer. 2011, pp. 331–340.

[22] Brian R Duffy. “Anthropomorphism and the social robot”. In: *Robotics and autonomous systems* 42.3-4 (2003), pp. 177–190.

[23] Doaa Farouk Badawy Eldesouky. “Visual hierarchy and mind motion in advertising design”. In: *Journal of Arts and Humanities* 2.2 (2013), pp. 148–162.

[24] Nicholas Epley, Adam Waytz, and John T Cacioppo. “On seeing human: a three-factor theory of anthropomorphism.” In: *Psychological review* 114.4 (2007), p. 864.

[25] Jasper Feine et al. “A taxonomy of social cues for conversational agents”. In: *International Journal of Human-Computer Studies* 132 (2019), pp. 138–161.

[26] Andy Field. *Discovering statistics using IBM SPSS statistics*. Sage Publications Ltd, 2018.

[27] Kraig Finstad. “The system usability scale and non-native English speakers”. In: *Journal of usability studies* 1.4 (2006), pp. 185–188.

[28] Brian J Fogg and Hsiang Tseng. “The elements of computer credibility”. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 1999, pp. 80–87.

[29] Marlena R Fraune. “Our robots, our team: Robot anthropomorphism moderates group effects in human–robot teams”. In: *Frontiers in psychology* 11 (2020), p. 1275.

[30] Li Gong. “How social is social responses to computers? The function of the degree of anthropomorphism in computer representations”. In: *Computers in Human Behavior* 24.4 (2008), pp. 1494–1509.

[31] Shaoying GONG et al. “The effects of emotional design on multimedia learning”. In: *Acta Psychologica Sinica* 49.6 (2017), p. 771.

[32] Michael A Goodrich and Dan R Olsen. “Seven principles of efficient human robot interaction”. In: *SMC’03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483)*. Vol. 4. IEEE. 2003, pp. 3942–3948.

[33] F.B.J. Grapperhaus. “Bijlage: Toelichting op de ICT-vernieuwing bij de politie”. In: (2018), pp. 1–12.

[34] F.B.J. Grapperhaus. “Tweede Kamer der Staten-Generaal”. In: *Tweede Kamer der Staten-Generaal* 32 851.71 (2021), pp. 1–8. URL: `http://arno.unimaas.nl/show.cgi?fid=14775`.

[35] Peter A Hancock et al. “A meta-analysis of factors affecting trust in human-robot interaction”. In: *Human factors* 53.5 (2011), pp. 517–527.

[36] Richard J Hanowski, Susan C Kantowitz, and Barry H Kantowitz. “Driver acceptance of unreliable route guidance information”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 38. 16. SAGE Publications Sage CA: Los Angeles, CA. 1994, pp. 1062–1066.

[37] Ben M Harvey and Serge O Dumoulin. “Visual motion transforms visual space representations similarly throughout the human visual hierarchy”. In: *Neuroimage* 127 (2016), pp. 173–185.

[38] Frank D. Hodge, Kim I. Mendoza, and Roshan K. Sinha. “The Effect of Humanizing Robo-Advisors on Investor Judgments*”. In: *Contemporary Accounting Research* 38.1 (2021), pp. 770–792.

[39] Kevin Anthony Hoff and Masooda Bashir. “Trust in automation: Integrating empirical evidence on factors that influence trust”. In: *Human Factors* 57.3 (2015), pp. 407–434.

[40] L Michael Honaker, Vicki Schwartz Hector, and Thomas H Harrell. "Perceived validity of computer-versus clinician-generated MMPI reports". In: *Computers in Human Behavior* 2.1 (1986), pp. 77–83.

[41] Katherine Isbister and Clifford Nass. "Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics". In: *International journal of human-computer studies* 53.2 (2000), pp. 251–267.

[42] *It's the New Tamagotchi Pix Party! your Tamagotchi character is ready for a celebration!* Nov. 1996. URL: `https://tamagotchi.com/`.

[43] Barry H Kantowitz, Richard J Hanowski, and Susan C Kantowitz. "Driver acceptance of unreliable traffic information in familiar and unfamiliar settings". In: *Human Factors* 39.2 (1997), pp. 164–176.

[44] Miles A Kimball. "Visual design principles: An empirical study of design lore". In: *Journal of Technical Writing and Communication* 43.1 (2013), pp. 3–41.

[45] Tomoko Koda and Pattie Maes. "Agents with faces: The effect of personification". In: *Robot and Human Communication - Proceedings of the IEEE International Workshop* May 2014 (1996), pp. 189–194.

[46] Philip T Kortum and Aaron Bangor. "Usability ratings for everyday products measured with the system usability scale". In: *International Journal of Human-Computer Interaction* 29.2 (2013), pp. 67–76.

[47] Patrick Lawson. *How to sus out usability scores*. Sept. 2022. URL: `https://www.thinkcompany.com/blog/how-to-sus-out-usability-scores/`.

[48] Outfit7 Limited. *My talking tom 2*. Sept. 2018. URL: `https://apps.apple.com/nl/app/my-talking-tom-2/id1337578317`.

[49] Outfit7 Limited. *Talking Tom Cat*. June 2010. URL: `https://apps.apple.com/nl/app/talking-tom-cat/id377194688`.

[50] Catherine L Lortie and Matthieu J Guitton. "Judgment of the humanness of an interlocutor is in the eye of the beholder". In: *PLoS One* 6.9 (2011), e25085.

[51] James C McCroskey and Thomas J Young. "Ethos and credibility: The construct and its measurement after three decades". In: *Communication Studies* 32.1 (1981), pp. 24–34.

[52] Edo-Jan Meijer. *UX & Politie Nederland User Experience Competence Center*. Dutch National Police, 2022.

[53] Zeynep Mennan. "From simple to complex configuration: Sustainability of gestalt principles of visual perception within the complexity paradigm". In: (2009).

[54] Saif Mohammad. "Colourful language: Measuring word-colour associations". In: *arXiv preprint arXiv:1309.5942* (2013).

[55] Peter Mom. *Politie-ict al veertig Jaar Een Hoofdpijndossier*. Nov. 2017. URL: https://www.nporadio1.nl/nieuws/onderzoek/4ed6fd58-4e12-4966-b033-cd35af94a44a/politie-ict-al-veertig-jaar-een-hoofdpijndossier.

[56] Masahiro Mori, Karl F MacDorman, and Norri Kageki. "The uncanny valley [from the field]". In: *IEEE Robotics & automation magazine* 19.2 (2012), pp. 98–100.

[57] Peter Mundy and Lisa Newell. "Attention, joint attention, and social cognition". In: *Current directions in psychological science* 16.5 (2007), pp. 269–274.

[58] Juha Munnukka, Karoliina Talvitie-Lamberg, and Devdeep Maity. "Anthropomorphism and social presence in Human–Virtual service assistant interactions: The role of dialog length and attitudes". In: *Computers in Human Behavior* (2022), p. 107343.

[59] Kristine L Nowak and Christian Rauh. "The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction". In: *Journal of Computer-Mediated Communication* 11.1 (2005), pp. 153–178.

[60] dr. O. Truijens ing. P.K. de Vries MBA drs. G.H. Hoijtink. "Programmaplan 2019 - 2024". In: (2018), pp. 1–79.

[61] Nationale Politie. *Geschiedenis Nederlandse Politie*. 2014. URL: https://www.politie.nl/informatie/geschiedenis-nederlandse-politie.html.

[62] Nationale Politie. *Politiewet 2012*. 2012. URL: https://wetten.overheid.nl/BWBR0031788/2019-02-01.

[63] Nationale Politie. *Waakzaam en diestbaar, PVR*. May 2022.

[64] Gary Polson. *Virtual Pets Blog The World of Virtual Pets*. Jan. 2017. URL: http://virtualpet.com/.

[65] Aaron Powers and Sara Kiesler. "The advisor robot: tracing people's mental model from a robot's physical attributes". In: *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 2006, pp. 218–225.

[66] Robert W Proctor and Trisha Van Zandt. *Human factors in simple and complex systems*. CRC press, 2018.

[67] Dale Purves et al. *Cognitive neuroscience*. Vol. 6. 4. Sunderland: Sinauer Associates, Inc, 2008.

[68] Philipp A. Rauschnabel and Aaron C. Ahuvia. "You're so lovable: Anthropomorphism and brand love". In: *Journal of Brand Management* 21.5 (2014), pp. 372–395.

[69] Byron Reeves and Clifford Nass. "The media equation: How people treat computers, television, and new media like real people". In: *Cambridge, UK* 10 (1996), p. 236605.

[70] Algemene Rekenkamer. “ICT politie 2010”. In: 29 350 Nr. 9 (2011).

[71] Hayley Robinson, Bruce MacDonald, and Elizabeth Broadbent. “The role of healthcare robots for older people at home: A review”. In: *International Journal of Social Robotics* 6.4 (2014), pp. 575–591.

[72] Young June Sah and Wei Peng. “Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website”. In: *Computers in Human Behavior* 45 (2015), pp. 392–401.

[73] Anna-Maria Seeger, Jella Pfeiffer, and Armin Heinzl. “Texting with humanlike conversational agents: designing for anthropomorphism”. In: *Journal of the Association for Information Systems* 22.4 (2021), p. 8.

[74] Chenyu Shangguan et al. “More attractive or more interactive? The effects of multi-leveled emotional design on middle school students’ multimedia learning”. In: *Frontiers in psychology* 10 (2020), p. 3065.

[75] Nicolas Spatola, Sophie Monceau, and Ludovic Ferrand. “Cognitive Impact of Social Robots: How Anthropomorphism Boosts Performances”. In: *IEEE Robotics and Automation Magazine* 27.3 (2020), pp. 73–83.

[76] Li-Chu Tien, Chei-Chang Chiou, and Yueh-Shian Lee. “Emotional design in multimedia learning: Effects of multidimensional concept maps and animation on affect and learning”. In: *EURASIA Journal of Mathematics, Science and Technology Education* 14.12 (2018), em1612.

[77] Michael Tomasello et al. “Joint attention as social cognition”. In: *Joint attention: Its origins and role in development* 103130 (1995), pp. 103–130.

[78] Lyn M. Van Swol and Emily Seinfeld. “Differences between minority, Majority, and unanimous group members in the communication of information”. In: *Human Communication Research* 32.2 (2006), pp. 178–197.

[79] Ministerie van Justitie en Veiligheid. *Hoofdlijnenplanning Programma Vernieuwend Registreren*. Dec. 2021. URL: `https://www.rijksoverheid.nl/documenten/rapporten/2020/12/17/tk-bijlage-hoofdlijnenplanning-programma-vernieuwend-registreren`.

[80] Ewart J de Visser et al. “The world is not enough: Trust in cognitive agents”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 56. 1. Sage Publications Sage CA: Los Angeles, CA. 2012, pp. 263–267.

[81] Mark Douglas West. “Validating a scale for the measurement of credibility: A covariance structure modeling approach”. In: *Journalism quarterly* 71.1 (1994), pp. 159–168.

[82] Svetlana Yarosh et al. “Children asking questions: speech interface reformulations and personification preferences”. In: *Proceedings of the 17th ACM conference on interaction design and children*. 2018, pp. 300–312.

List of Figures

List of Tables

8 Appendices

A Presentation order of prototypes

Order of presentation per condition			
Participant	RAPP	ARAPPV	ARAPPN-V
P1	1	2	3
P2	3	1	2
P3	1	2	x
P4	2	1	3
P5	3	2	1
P6	1	3	2
P7	2	3	1
P8	1	2	3
P9	2	1	3
P10	3	1	2
P11	3	2	1
P12	2	3	1
P13	1	3	2

Figure 8.1. An overview of the order in which the prototypes were presented to each participant. The colours represent the different possibilities in which the prototypes were presented. Each order/colour is represented two times, except for P3. P3 has a different colour, since participant 3 did not participate in the non-verbal condition.

B Questionnaire design

B.1 Demographics

At the start of the research, the participant was presented with a questionnaire containing statements focusing on demographics and feelings towards systems in general (see the questionnaire below). The questions were presented in Dutch, since all participants were native Dutch speakers.

B.2 User interaction

A second questionnaire was produced, which was conducted at the end of each interaction with RAPP, ARAPPV and ARAPPN-V. This questionnaire is used to measure the user experience while interacting with the prototypes. The questions are divided in two parts, which are visible below.
The first part of the questionnaire is focused on usability and task satisfaction (question 1). This is measured by means of so-called SUS scores (System Usability Scale) [9, 10].
The second and final part includes statements about anthropomorphism (questions 2a, 3 & 4), trust (question 5) and credibility (questions 2b, & 6).
At the end of the questionnaire the participants was thanked for participating.

Demografisch en Technologie

Start of Block: Briefing

Er volgen nu een aantal persoonlijke vragen en een vraag over hoe u in het algemeen naar technologie kijkt.

End of Block: Briefing

Start of Block: Open questions

Question 1
Wat is uw leeftijd?

_____

Question 2
Hoeveel jaar bent u in dienst bij de politie?

_____

Question 3
Hoeveel jaar daarvan hebt u BVH en nu RAPP gebruikt?

_____

End of Block: Open question

Start of Block: Likert Scale

Question 4

Nu volgen een aantal standpunten over hoe u naar technologie in het algemeen kijkt.
Geef bij de onderstaande uitspraken aan in hoeverre dit op u van toepassing is.

Het gebruik van nieuwe technologie voelt voor mij:

	1	2	3	4		
Ongunstig	○	○	○	○	○	Gunstig
Slecht	○	○	○	○	○	Goed
Nadelig	○	○	○	○	○	Voordelig
Negatief	○	○	○	○	○	Positief

End of Block: Likert Scale

Start of Block: Debrief

Dit is het einde van de vragenlijst. Bedankt voor het invullen!

End of Block: Debrief

RAPP evaluation

Start of Block: Briefing

U heeft zojuist gewerkt met een variant van RAPP. Er volgt nu een korte vragenlijst. Deze vragenlijst heeft alleen betrekking op de variant waarmee u zojuist heeft gewerkt en niet op eerdere ervaringen.
Door middel van deze vragenlijst, wordt onderzocht hoe u het design en het gebruik van RAPP heeft ervaren.

End of Block: Briefing

Start of Block: Usability SUS

Question 1

Er volgen nu een aantal vragen over uw ervaring ten aanzien van het design van RAPP.

Geef bij de onderstaande uitspraken aan in hoeverre u het er mee eens bent.
De schaal is verdeeld van 1 tot en met 5.
1: Heel erg mee oneens
2: Beetje mee oneens
3: Neutraal
4: Beetje mee eens
5: Heel erg mee eens

Met het systeem wordt RAPP bedoeld waarmee u zojuist heeft gewerkt.

	1	2	3	4	5
1. Ik denk dat ik dit systeem vaak zou willen gebruiken	○	○	○	○	○
2. Ik vind het systeem onnodig complex	○	○	○	○	○
3. Ik vond het systeem makkelijk te gebruiken	○	○	○	○	○
4. Ik denk dat ik hulp nodig heb van een technisch persoon om dit systeem te kunnen gebruiken	○	○	○	○	○
5. Ik vond dat de verschillende functies in dit systeem goed geïntegreerd zijn	○	○	○	○	○
6. Ik vond dat er te veel tegenstrijdigheid in dit systeem zit	○	○	○	○	○

7. Ik kan me voorstellen dat de meeste mensen dit systeem snel leren gebruiken	○	○	○	○	○
8. Ik vond het systeem erg lastig te gebruiken	○	○	○	○	○
9. Ik voelde me erg zelfverzekerd tijdens het gebruik van dit systeem	○	○	○	○	○
10. Ik moest een hoop dingen leren voordat ik aan de slag kon met het systeem	○	○	○	○	○

End of Block: Useabilty SUS

Start of Block: Anthropomorphism

Question 2a

Er volgen nu een aantal vragen over het functioneren van RAPP waarmee u zojuist heeft gewerkt.
Geef hieronder aan in hoeverre u de onderstaande begrippen van toepassing vindt.

Het functioneren van het systeem voelt...

	1	2	3	4	5	
Nep	○	○	○	○	○	Natuurlijk
Machine-achtig	○	○	○	○	○	Menselijk
Zonder bewustzijn	○	○	○	○	○	Met bewustzijn
Kunstmatig	○	○	○	○	○	Echt
Gevoelloos	○	○	○	○	○	Sympathiek

Question 2a
Geef bij de onderstaande uitspraken aan in hoeverre u het er mee eens bent.

De schaal is verdeeld van 1 tot en met 5.
1: Heel erg mee oneens
2: Beetje mee oneens
3: Neutraal
4: Beetje mee eens
5: Heel erg mee eens

Met het systeem wordt RAPP bedoeld waarmee u zojuist heeft gewerkt.

	1	2	3	4	5
1. Het contact met dit systeem voelde menselijk aan	○	○	○	○	○
2. Het contact met het systeem voelde persoonlijk aan	○	○	○	○	○
3. Ik had het gevoel dat ik met een politie-collega communiceerde	○	○	○	○	○

End of Block: Anthropomorphism

Start of Block: Credibility

Question 2b

4. Ik had het idee dat het systeem op mij was aangepast	○	○	○	○	○	

	1	2	3	4	5	Niet van toepassing
5. Ik heb iets gehad aan de instructies van het systeem	○	○	○	○	○	○

End of Block: Credibility

Start of Block: Anthropomorphism

Question 3

Geef hieronder aan in hoeverre u de onderstaande begrippen van toepassing vindt op het systeem waarmee u zojuist heeft gewerkt

	1	2	3	4	5	
Niet leuk	○	○	○	○	○	Leuk
Onvriendelijk	○	○	○	○	○	Vriendelijk
Onaardig	○	○	○	○	○	Aardig
Onaangenaam	○	○	○	○	○	Aangenaam
Vreselijk	○	○	○	○	○	Fijn
Niet sociaal	○	○	○	○	○	Sociaal

Question 4

Geef hieronder aan in hoeverre u de onderstaande begrippen van toepassing vindt op het systeem waarmee u zojuist heeft gewerkt

	1	2	3	4	5	
Incompetent	○	○	○	○	○	Competent
Onwetend	○	○	○	○	○	Goed geïnformeerd
Onverantwoordelijk	○	○	○	○	○	Verantwoordelijk
Onintelligent	○	○	○	○	○	Intelligent
Dom	○	○	○	○	○	Verstandig

End of Block: Anthropomorphism

Start of Block: Trust

Question 5

Geef hieronder aan in hoeverre u de onderstaande begrippen van toepassing vindt.

Dit systeem is:

	1	2	3	4	5	
Belerend	○	○	○	○	○	Ongedwongen/vrij
Druk	○	○	○	○	○	Rustig
Vol	○	○	○	○	○	Leeg
Lang van stof	○	○	○	○	○	Kort van stof
Afhankelijk	○	○	○	○	○	Zelflerend
Onduidelijk	○	○	○	○	○	Duidelijk

End of Block: Trust

Start of Block: Credibility

Question 6
Geef bij de onderstaande uitspraken aan in hoeverre u deze van toepassing vindt op het systeem.

De schaal is verdeeld van 1 tot 5.
1: Heel erg mee oneens
2: Beetje mee oneens
3: Neutraal
4: Beetje mee eens
5: Heel erg mee eens

Met het systeem wordt RAPP bedoelt waarmee u zojuist heeft gewerkt.

	1	2	3	4	5
Dit systeem voelt als een echte expert	○	○	○	○	○
Dit systeem begrijpt wat ik nodig heb als het een probleem betreft	○	○	○	○	○
Dit systeem zet mij op nummer 1	○	○	○	○	○
De hulp die dit systeem biedt werd voor mij als passend ervaren	○	○	○	○	○
Dit systeem voelt eerlijk	○	○	○	○	○
Dit systeem voelt integer	○	○	○	○	○
Ik kan op dit systeem vertrouwen	○	○	○	○	○

End of Block: Credibility

Start of Block: Debrief

Dit is het einde van de vragenlijst. Bedankt voor het invullen!

End of Block: Debrief

C Prototype adjustments

To create the prototypes, the following adjustments are made to RAPP.

C.1 List of adjustments:

ARAPPV:

- Page showing Persoon, added: "voeg een andere rol toe" in the colour red. Red since the previously chosen role is erased or adjusted because it was wrong. It is written at the bottom of the page so the words are not written on top of each other. The same font and size as the rest of the text on the page is used to add to the similarity.
- Page when role is assigned: "De rol is toegewezen. Druk op de pijl linksboven om de gestarte actie af te ronden." This is in black, as it is an instruction and it is displayed at the bottom of the page. Same font and size as the rest of the text to add to the similarity.
- Page after hOvJ is assigned: "De hOvJ is toegewezen. Druk op de pijl linksboven om de gestarte actie af te ronden." This is in black because it is an instruction. It is written at the bottom of the page. Same font and size as the rest of the text to add to the similarity.
- After the box at Verstrekken gegevens is checked, a message in black shows: "Er is aan de voorwaarde voldaan. Druk op de pijl linksboven om de gestarte actie af te ronden.". Again, it is an instruction, displayed at the bottom of the page. The same font and text size as the rest of the text is used to add to the similarity.
- Reprimande voorwaarden: An instruction message is displayed in black: "Om in aanmerking te komen voor een reprimande moet de verdachte aan alle voorwaarden voldoen.". It is shown at the top of the page, so it is noticed early on. Text size and font are the same as the rest of text on the page.
- Page Reprimande when all the boxes are checked the message: "De reprimande is compleet. Druk op het vinkje rechtsboven om de reprimande af te ronden." is displayed. This message appears at the top of the page. It could not be displayed at the bottom of the page. The user then must scroll down to read the message otherwise, it is not visible. The text is written in black in the same font, colour and size as the rest of the text.
- At the bottom of the page Verstrekken gegevens the message appears when all boxes are checked: "verstrekken gegevens is compleet. Druk op het vinkje rechtsboven om verstrekken gegevens af te ronden." This is in the same font, colour (black) and text size as the rest of the text.

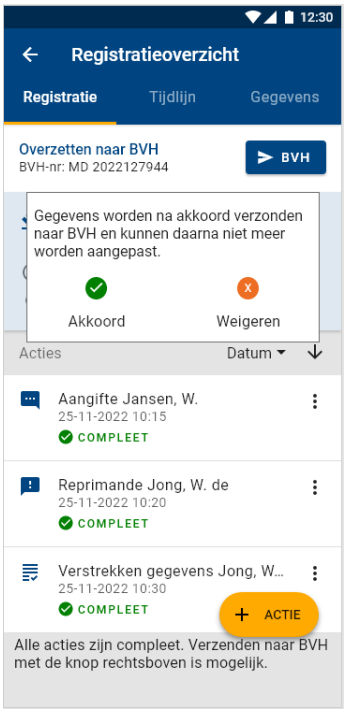


Figure 8.2. The message displayed when the user wants to send the data to BVH

- When all tasks are complete a message at Registratieoverzicht shows: "Alle acties zijn compleet. Verzenden naar BVH met de knop rechtsboven is mogelijk." This is placed at the bottom of the page to not overlap other text. It is in the same font, colour (black) and text size as the rest of the text.

- Figure 8.2 shows the notification appearing when the user is about to send the data to BVH. It is a warning to ensure the user knows that the decision is a final one and adjustments are not possible afterwards. The message is placed as an overlay on the previous page to ensure that the user knows that it is still possible to go back. The message is shown in the same font as the text on the page beneath. The Akkoord button is green (same green as the COMPLEET text at the page beneath) to ensure that the user feels comfortable pressing it. Weigeren is orange (the same colour as the text shown when a task is incomplete), because it is an important decision the user must feel free to press the Weigeren button to double check. If it is red, the user might be hesitant to press the button.

- Page showing Persoon the following message is added in black with the same font and size as the rest of the text: "De persoonsgegevens zijn compleet. Druk op het vinkje rechtsboven om de gestarte actie af te ronden." This is placed at the top of the page, as this location is noticed first. The prototype does not enable further registration of data. This message makes sure the user knows that the data is complete.

ARAPPN-V:

- Mijn werk shows a shoplifting incident with high priority. To highlight this priority, the bell is moving (shifts from left to right) as attention cue for the user to pick up that specific incident. The time between the shifts is 0.3 seconds to make the bell moves as real as possible.

- At Aanhouding the red box at Persoon appears, to indicate it needs to be filled out. To make this box appear to be blinking, another type of red is added (from #D90000 to #FF1919, complementary colour). This is done to create a slight contrast and make it more noticeable. This is also done at Verstrekken gegevens (Voorwaarden verstrekking) and Reprimande (hOvJ). To make the blinking as clear as possible, the time between colours is set at 0.3 seconds.

- Just as done with the red boxes that indicate something needs to be filled out, the orange boxes to indicate that something is incomplete also change colour. For example at Verstrekken gegevens. The colour is changed slightly to create contrast (from #ED6D22 to #FF843D, complementary colour). It appears as if the box is blinking. This is also, the case for the orange box at Aanhouding.

- As Aanhouding is closed without being completely filled out, the screen turns to Registratieoverzicht and it shows OPENSTAAND at Aanhouding. This is adjusted to make OPENSTAAND appear to be blinking. The same adjustment is added to the page when the Reprimande is closed without being completely filled out and again, when Verstrekken gegevens is not completely filled out (adjusted colour: from #FFAA00 to #FF710D, double complementary colour, increased text size: from 12 to 14 and blinking time 0.3 seconds).

- If the message at Persoon is displayed (after the registration of the passport) the check mark (✓) is increased in size (font size 40) to make it look like it is blinking. The colour is not changed since the difference between the background (blue) and the colour of the arrow (white) is already contrasting. The time between the blinking arrows is 0.3 seconds to make it as noticeable as possible. This adjustment is also added to the check mark when the message "De reprimande is compleet..." is shown and again at Verstrekken gegevens when it is completed.

- Whenever a role is assigned at Persoon, the message shows "de rol is toegewezen...", here the arrow at the top right is increased in size (font 40) to support the message. Just like the check mark mentioned before, the arrow appears to be blinking to draw attention towards it (blinking time 0.3 seconds). The colour is not changed since the differentiation between the background (blue) and the colour of the arrow (white) is already contrasting. The same adjustment is included at Voorwaarden verstrekking when the checkbox is checked and the message to press

the arrow is shown and once more when the message "de hOvJ is toegewezen..." appears.

- When all the data is filled out and all the tasks at Registratieoverzicht are completed, the data needs to be sent to BVH. A message appears to explain this. As support for this message the colour of the button is changed so it appears to be blinking (from #004682 to #1578CF, complementary colour, blinking time 0.3 seconds).

Adjustments to ARAPPV based on feedback:

- In the messages the reference to the location is removed. Providing the location in a message means that the location is not clear, if the location is not clear, the design is wrong. In this case, the location is clear, therefore the location reference is unnecessary.

- The instructional texts are moved to the top of the page. This way the instruction is provided at ones. This is in congruence with research [21, 23], that explains that information presented at the top of the page is processed at an earlier stage.

- The text colour is changed to green (#008000), the same as the boxes. This is done to associate it with progress because green is associated with progress. Additionally, the text pops-out more. The pop-out is meant to clarify that the text is different from the other text as it is an instruction. Also, black text was not "personal" enough and the green colour adds to a less distant feeling, because of colour associations.

- ARAPPV needs to add to the feeling of communicating with another person. However, that feeling is not there. It was suggested to add a police avatar, but since this research does not investigate obvious cues, the police logo is used as a marking. It is added next to the instruction messages. Since it is a police logo it should add to unity. It should evoke the feeling of communicating with a partner and hence create a more personal environment. Normally, this logo is blue however, to correspond with the presented message it is changed to either green or red. Green refers to progress, the officer is making progress in the step-by-step progress. Red means the officer is not making progress, but needs to pay extra attention since something is incomplete (taking a step back).

- As speech interfaces (chatterbots) present information in a speech cloud to anthropomorphise, ARAPPV makes use of such clouds as well, again to add to the personification. This is done by grouping the presented information (text and logo) at the top of the page and present it as distinctive, instructional information. Every instruction is presented in the same way to make it more recognizable

and distinctive. The speech cloud is placed in such a way that it looks like the system is talking (from the top left, see red circle in figure 8.3).

- To make the instruction appear more spacious and less full, the text size is adjusted (from 16 to 14). Otherwise, the text would be to close together inside the speech cloud, making it appear more artificial and stressful.

- Instead of using "gegevens zijn compleet", "alle verplichte velden zijn ingevuld" is used. This because the system does not know if it is correct. It just knows that the required fields are filled out.

- It was suggested to add more verbal cues. Therefore, a verbal cue is added to Persoon at Aanhouding. A common mistake is adding the wrong role. The addition of a verbal cue can clarify the ambiguity in this situation. A green speech cloud (instruction message) is placed at the top of the Persoon box. The message would have been presented in a hover state, meaning that it appears when the officer hovers over the box. This would be advantageous since hovering enables a message to emerge only when needed and not in a constant capacity. However, since the application is used on a phone, hovering is not possible. Instead, the Persoon page slides to the page with the message, when Persoon is pressed, which then automatically slides to the Persoon toevoegen page after 1.5 seconds (< 1.5 seconds is too slow to notice and register and > 1.5 seconds is too long and intervenes with the workflow). This way, the message does not appear constantly but only when needed. To improve readability, the font size is set to 14, the image and text already in the box are decreased in size and moved down. It now appears as if the system responds to the user by speaking with an instruction. The message shows: "Het gaat hier om de verdachte". The message appears with a delay of 0.2 seconds (> 0.2 seconds is too long, it looks like the system needs to "think" too long before responding, while < 0.2 seconds appears too fast, unnatural). The slide delay adds to the personification of the interface, as if your colleague reminds you before moving to the next page.

- A textual instruction is added at the top of the page Voorwaarden verstrekking. It explains what voorwaarden verstrekking entails, since police officers sometimes have difficulty understanding this. It is place at the top to be noticed first. The text is the same font, size and colour as the other instructional speech clouds (font size 14 (roboto regular), colour code: #008000)

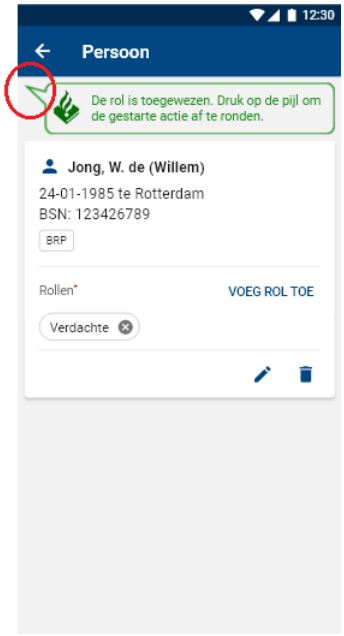


Figure 8.3. An example of a speech cloud: The green text and logo are displayed as an instruction at the top of the page. The red circle marks the direction of the speech cloud. It is placed in such a way that it looks like the system is speaking.

Adjustments to ARAPPN-V based on feedback:

- It was said that the movement of the arrows (blinking effect) might cause irritation. Therefore, instead of increasing size, the arrows change colour (blinking). The colour white causes a contrasting effect on the blue background, therefore earlier on the choice was made not to change the colour of the white arrows. However, if a colour similar to white is used the contrast does not change. A shadow colour (#BFBFBF) was selected[1] which is darker than white. The change between the two colours causes a slight difference is visibility, but not enough to irritate. To contribute to this, the timing is changed to 0.7 seconds instead of 0.3. Additionally, the transition is changed to "dissolve". It now looks like the arrow is slightly "pulsing". The change of colour is in correspondence with the literature [23]. This adjustment was also added to the check marks.

- To reduce the irritation effect, the blinking of the boxes to indicate what needs to be filled out is reduced to 0.5 seconds. This way, the change of colour is still visible, but not in a stressful, demanding way. Changing it to a transition longer than 0.5 seconds is a too slow for this notification, since the time spend on the page is not that long.

- The colour transition of the word "OPENSTAAND" is adjusted from 0.3 to 0.4 (0.3 was too fast, while 0.5 or higher was too slow for the time spend on the page) and the transition is changed to "dissolve". This was done to create a smoother transition (pulse instead of blink). Additionally, the change of the font size is removed: the font size no longer changes. This creates a less stressful design.

[1] Colour selected from `https://color.adobe.com/nl/create/color-wheel`

- The colour change of the BVH button is adapted in such a way that the change from the dark blue to lighter blue is faster (0.3 sec) then the change back (1.5 sec). The addition of the one colour dissolving in the other makes it look like a reminder of the possibility without it being annoying and too much.

Adjustments to ARAPPV based on further feedback:

- A on/off switch is placed at the bottom om the RAPP demo page. The text next to the switch shows "Ondersteuning?". This was added to enable the user to turn off (and on) the verbal and non-verbal support when using the application. This because after extensive use the support might no longer be needed and could only cause frustrating. However, for this research the button was always on and could not be turned off.

- The instruction provided at Aanhouding when persoon is pressed is moved to the top of the page with Persoon toevoegen. This was done, because at Aanhouding the layout is distorted and also the cue is given to early. At Aanhouding the officer still knows that it is about the suspect, but at Persoon toevoegen is could have slipped away (in the actual RAPP the Persoon toevoegen page is the same for multiple actions, however in this prototype the page is only implemented one time).

- The instruction: "De rol is toegewezen..." when the role at Persoon is assigned is adjusted to: "De gegevens zijn compleet. Druk op de pijl om de gestarte actie af te ronden.". This was done because "De rol is toegewezen" is self-explanatory. However, the message referring to pressing the arrow is still needed for consistency in the overall prototype. Providing the message: "Druk op de pijl om de gestarte actie af te ronden" alone is not consistent with the other messages since it will only provide an instruction and not a conformation.

- The message: "Alle verplichte velden zijn ingevuld. Druk op het vinkje om de gestarte actie af te ronden." at Persoon is adjusted, since it does not say that the data needs to be checked. It says that all obligatory fields are filled out, but it could be that the data is wrong. Therefore, the message is adjusted to: "Controleer of alle gegevens correct zijn ingevuld. Druk vervolgens op het vinkje om de gestarte actie af te ronden."

- The message at Voorwaarden verstrekking: "Dit heeft betrekking op het verstrekken van de gegevens van de verdachte aan de winkelier ten behoeve van schade verhaal." is moved to the page Verstrekken gegevens since the message is not only applicable to Voorwaarden verstrekken, but too the entire Verstrekken gegevens.

- At Reprimande voorwaarden an instruction message is added at the bottom of the page: "Zijn niet alle opties van toepassing, druk dan op het kruisje om verder te gaan met de aanhouding." This was done to aid the user in what to do when the Reprimande is not applicable. This prevents mistakes and improves workflow. The message is the same (colour, font, text size, layout) as the other green instruction messages.

Adjustments to ARAPPN-V based on further feedback:

- To indicate what needs to be done at GMS Melding, the button that shows "AANVULLEN" blinks (change colour to #FF710D, double complementary colour). The transition is set at 0.7 seconds in dissolve. This way it does not stress out the user and it is not annoying. It is a short "pulse" as a reminder for the user. This adjustment is also added to the "+ actie" at the Registratieoverzicht page.

- The colour change to the arrows and check marks makes is look like the they can only be pressed half of the time (when it blinks) because the other half of the time the colour is grey. Therefore, a circle was placed underneath the arrow/check mark that changes colour instead of the arrow/check mark itself (stays white). It now looks like a touch indicator that is presented after a user touches the screen. Only, the touch indicator is presented before the screen is actually touched. The same colour as the adjusted BVH button is used (#1578CF) and the shadow around the arrow was set to max to blend the circle in but to still make it distinguishable from the background.

Adjustments to ARAPPV based on one final review:

- At persoon, where a role must be added. Initially, a message in orange displayed: "Voeg minimaal 1 rol toe". Since all messages, either green or red are in a speech cloud marked with a police logo, it seemed odd that it was not the case here. Therefore, a orange speech cloud and orange police logo (same orange as the text already there) were added to the message. The cloud points towards VOEG ROL TOE.

- At the Registratieoverzicht screen where the button "+ actie" is placed and the participant has to start with "aanhouding", a new message is placed: "Ook voor een reprimande moet er gestart worden met een aanhouding." This was done because on this page there were no instructions yet and for some police officers it is unclear that they have to start with "aanhouding" when they already know or think there is a high change that it is going to be a reprimande. The message is the same (colour, font, text size, layout) as the other green instruction messages and the cloud is pointing towards Aanhouding.

- At the page with "Persoon toevoegen", where the officer can choose between different options to register a person, a message is placed beneath the option zoek op nummer saying: "Bijvoorbeeld een BSN-nummer". This was added to clarify what type of number it can be, since this is unclear sometimes. The cloud points towards zoek op nummer and has the same layout etc. as the other instruction messages.

- At the page where the instruction message initially said: "Controleer of alle gegevens correct zijn ingevuld. Druk vervolgens op het vinkje om de gestarte actie af te ronden." (at persoon) the message is adjusted to: "Controleer of alle gegevens compleet zijn. Scroll hiervoor de hele pagina door. Druk daarna op het vinkje om verder te gaan." This was added because the user needs to go over the entire page to see if something is missing. At the end of the page there is another option to add something which is often forgotten.

- At Reprimande voorwaarden: The instruction message shows "Om in aanmerking te komen voor een reprimande moet de verdachte aan alle voorwaarden voldoen." "alle" is displayed in bold font. This was done to highlight the significance.

- When all boxes are green at the Reprimande page the message now shows: "Alle benodigde handelingen zijn verricht. Druk op het vinkje om de reprimande af te ronden" instead of "De reprimande is compleet. Druk op het vinkje om de reprimande af te ronden.". The previous sentence was not correct, since the reprimande is only complete when the hOvJ has approved it, which you do not know at this point. It also contributes to word variability.

- After the aanhouding/reprimande is complete, the user is back at the Registratieoverzicht screen with the "+ actie" button. Here a new message is added in the same layout as the other messages and pointing towards the button: "Als het gaat om meerdere verdachten moet een actie vaker uitgevoerd worden." This was done, because a lot of police officers do not know this, but it is really important.

- The message saying that it is possible to send to BVH is adjusted to: "Alle acties zijn uitgevoerd. Verzenden naar BVH is mogelijk.". This was done, because the system does not know if all actions are complete, it only knows that they are all filled out, thus executed (uitgevoerd).