AN EXPLORATORY STUDY ON THE RELATIONSHIP BETWEEN (UN)TRUST AND MEMORY

Niki Volonasi, s1775014

Human Factors & Engineering Psychology

University of Twente

September 2019 – January 2021

First Supervisor: Mr Simone Borsci

Second Supervisor: Dr Rob van der Lubbe

Table of Contents

Acknowledgements

Looking back on this thesis journey, there were a lot of things I learned along the way and many people I will like to thank. Firstly, I will like to thank my first supervisor, Simone Borsci. After working together with Mr Borsci for my Bachelors Thesis, I knew that there were many things that we could further explore on the concept of (un)trust. I am therefore very grateful for giving me the opportunity to work with him again and explore this concept even further. I would also like to thank my second supervisor Dr. Rob van der Lubbe who gladly undertook this assignment even though this was not initially planned. Their extensive feedback and guidance throughout this journey were crucial to the success of this thesis.

I would also like to thank Mr Martin Schmettow who was keen on guiding me through this project, especially with his knowledge in Bayesian Analysis even though due to his personal circumstances the collaboration ended prematurely. I wish him all the best in his recovery.

Like every big journey, there were some rollercoaster on the way. Due to Covid-19 several changes and adaptations had to take place. For example, the eye-tracking analysis unfortunately had to be cancelled, something that both me and Mr Borsci were excited in exploring. Motivation was also a challenge for time to time, and for this I would like to thank again my supervisors for their flexibility in arranging meetings when I stabled into uncertainty.

Of course, I would not be able to carry on this journey without my “support group” of family and friends. I would like to thank my father Tolis, my mother Augoustina, my sister Alkisti and of course my dog Toulouse for their belief in me, and their daily dose of support. I would also like to thank El Niño and especially Michael Angelo Groeneveld, for providing me with the flexibility to work on my thesis even thought that meant that some days I had to work less. Lastly, I would like to thank my friends Marina, Rania, Lida, Constantinos, Aurora, Barbara, Jitske, John, Nienke, Lasse and many more for not only their help in proof-reading this paper but for always being there for me. Thank you everyone for believing in me, for supporting me and for keep pushing me in times when I barely believed in myself.

Abstract

Trust is an essential determinant of human-to-human interactions. Nowadays, human-to-technology interactions have emerged, which can lead to positive or negative outcomes. It is thus important for individuals to be able to correctly recognize (un)trust systems. The current study, aimed at exploring the recognition of (un)trust and whether recognition differences exist between (un)trustworthy stimuli. Furthermore, studies have also shown that individual differences could affect decision making. Thus, the current study explored whether individual differences (Faith in General Technology, Trust Stance and Personality) affected the recognition of (un)trustworthy stimuli. A recognition test was conducted according to the Signal Detection Theory. Participants (N=84) were blind to the scope of the experiment and to the (un)trustworthiness of the stimuli. The results, analysed using the Bayesian Approach, showed that people were equally able to remember both trustworthy and untrustworthy stimuli. However, a difference was identified which suggests that untrustworthy stimuli were better recognized than trustworthy when compared to situations in which stimuli were neutral. The analysis of the effect of individual differences on (un)trust recognition showed no evident effects. It can therefore be argued that the participants' trustworthiness judgments were not fully activated through this implicit exposure to the stimuli. The ability to recognize (un)trustworthy systems, protects the individual from potential threats and thus, future research should explore the factors that can affect the recognition of (un)trust, both between humans and between humans and technology.

Samenvatting

Vertrouwen is een essentiële determinant van menselijke interacties. Tegenwoordig zijn er opkomende interacties tussen mensen en technologie, die tot positieve of negatieve resultaten kunnen leiden. Het is dus belangrijk dat mensen de vertrouwbaarheid van technologische systemen correct kunnen herkennen. Het huidige onderzoek is gericht op de herkenning van (on)vertrouwbaarheid. Het is ook onderzocht, of er verschillen bestaan tussen de herkenning van (on)betrouwbare stimuli. Bovendien, heeft het huidige onderzoek aangetoond dat individuele verschillen, die bestaan tussen mensen, de besluitvorming kunnen beïnvloeden. Daarom is in de huidige studie onderzocht of individuele verschillen (Faith in General Technology, Trust Stance and Personality) de herkenning van (on)betrouwbare stimuli kunnen beïnvloeden. Er is een herkenningstest uitgevoerd volgens de signaaldetectietheorie (SDT). De deelnemers (N = 84) waren niet bewust van de reikwijdte van het experiment en van de (on)betrouwbaarheid van de stimuli. De resultaten, die zijn met behulp van de Bayesiaanse aanpak geanalyseerd, toonden aan dat mensen konden op dezelfde manier betrouwbare en onbetrouwbare stimuli herinneren. Er is een verschil gevonden, die suggereert dat onbetrouwbare stimuli beter zijn herkend dan betrouwbare stimuli, in vergelijking met situaties waarin stimuli neutraal waren. De analyse van het effect van individuele verschillen op de herkenning van (on) vertrouwen toonde geen duidelijke effecten. Daarom kan worden gesteld dat het betrouwbaarheidsoordeel van de deelnemers niet volledig geactiveerd is door deze impliciete blootstelling aan de stimuli. Het vermogen om (on) betrouwbare systemen te herkennen, beschermt de mensen tegen mogelijke bedreigingen en daarom moet toekomstig onderzoek de factoren onderzoeken, die van invloed kunnen zijn op de herkenning van (on) vertrouwen, zowel tussen mensen, als ook tussen mensen en technologie.

Introduction

Every day, humans interact with many individuals, that can differ from them in traits such as age, culture, race, personality. Despite these differences, humans can still decide, even in milliseconds, whether they want to interact with another person or not (Rule et al., 2012; Todovor 2008; Todovor & Duchaine, 2008; Lindgaard et al., 2006; Yu et al., 2014; Olivola et al., 2014, Willis & Todorov, 2006). Being able to make the correct decision gives a competitive advantage to the individual, by avoiding threatful situations. Studies have shown that when an interaction results in negative consequences, people will store these negative consequences to memory, and use them in future interactions with similar people (Muir, 1987; Wout & Sanfey, 2008; Bar, 2007). Trust plays a crucial role in aiding people to decide with whom to interact. Overall, trust is the belief that one party has towards another party's actions (Montague, 2010).

Nowadays, a new interaction has emerged, the one between humans and technology (devices, interactive systems, robots). People are now faced with the challenge to not only decide if they should interact with another human, but also with machines. Technology has entered our lives in several ways. Devices have entered our household, in the form of smartphones, smart fridges, smart light bulbs and have even replaced several human-to-human interactions through shopping, talking and paying online. Technology is not only used for entertainment but also to cover basic needs. It is therefore crucial for humans to be able to detect which products they can trust.

Due to the breakthroughs in Machine Learning and Artificial Intelligence (AI), more technologies are developed today, which do not only resemble humans in appearance, but also in function (Russel &Norvig, 2009). Humans are now faced with a challenging decision: whether the other party is authentic or not. The new technical possibilities that Machine Learning and Artificial Intelligence have brought to the field of media manipulation, has made it possible to fool mass audiences, by using realistic, artificially created media. Artificial Intelligence examines and learns a subject's visual and aural characteristics to map these characteristics into another subject (Russel &Norvig, 2009). An example of a manipulated medium that can have negative consequences in the masses is "deep fake". "Deep fake" uses Machine Learning and Artificial Intelligence to make images of fake events and people (Kietzmann et al., 2020). There are several examples of deep fake such as Barack Obama with Jordan's Peele voice (Suwajanakorn, Seitz & Kemelmacher-Shlizerman, 2017), or Mark Zuckerberg "admitting" that their platforms own their users (Chiu, 2019). These new technologies, and the threats they entail, show the importance of the correct recognition of the system's trustworthiness, which will in turn determine the user's interaction with the system. Due to the possibilities enabled by modern technology, how can it be assured that a system is authentic even before interacting with it?

As far as devices and products are concerned, a study conducted by Volonasi and Borsci (2019) showed that people were able to correctly rank groups of devices in terms of their untrustworthiness only based on appearance. These results led to the assumption that appearance cues must exist, which differentiate the trustworthy from the untrustworthy devices. The present work extends the study conducted by Volonasi and Borsci (2019) by exploring the mechanism(s) of recognition of trustworthy and untrustworthy visual stimuli. The first goal of the study is to examine the memorability of the stimuli and especially whether trustworthy and untrustworthy stimuli lead to differences in recognition. Secondly, the study explores the effect of further individual characteristics (i.e. Personality, Faith in General Technology and Trust Stance) in the recognition of (un)trust.

The next section briefly describes the theoretical background of the concepts of human-to-human and human-to-technology trust. The study then focuses on the effects of memory, previous knowledge and personality differences on decision making. An exploratory experimental study is conducted to investigate the difference in the recognition of trustworthy and untrustworthy visual stimuli (faces, scenes and devices), considering their memorability and the effect of individual characteristics (i.e. Personality, Faith in General Technology and Trust Stance).

Human-to-human Trust

In our everyday life, trust is a crucial determinant of our social and technological interactive exchanges. As a concept, trust can be seen as the overall belief in another person's or thing's actions. Specifically, trust is a relationship between two parties, the one who trusts (trustor), and the one that is being trusted (trustee). This relationship is affected by feelings of vulnerability, risk and expectations that the other person will compromise (Montague, 2010).

Definition of Trust

There have been several attempts in defining human-to-human trust. One definition was given by Barber, who emphasised the multidimensionality of the term, by introducing three different types of expectations that lead to trust: (1) general expectation on natural and moral social order (predictability); (2) expectation on the other person or system and its competent role (dependability); (3) the expectations to the others' future obligations and responsibilities (faith) (Muir, 1987; Madhavan & Wiegmann, 2007). When a person makes an initial judgment on whether to trust somebody, they consider how predictable the other party's future actions will be (Muir, 1987). Then, trust is based on behavioural evidence shown by the other person, such as on "trial-situations" of controlled risk, in which the other party had the opportunity to be unreliable, but was not (Muir, 1987). Following the judgment of dependability, trust between humans is based on faith that the other person will continue to be dependable (Muir, 1987).

The overall dynamic, multi-dimensional nature of trust, makes it an ever-changing process, affected by experience and shaped based on the stability of performance from the trustee (Muir, 1987; Madhavan & Wiegmann, 2007). Trust is therefore based on both confidence and expectations and also on an overall sense of vulnerability and uncertainty of the other party's actions (Chiou et al., 2020; Lee & See, 2004).

In social interactions, trustworthiness is a key factor in deciding whether to approach or avoid another party. Judgments are made very quickly in even 100ms, and these judgments have been found to stay persistent even when there are no time constraints (Rule et al., 2012; Todovor 2008; Todovor & Duchaine, 2008; Lindgaard et al., 2006; Yu et al., 2014; Olivola et al., 2014, Willis & Todorov, 2006). Even before any social interaction, when only brief encounters have taken place, some faces are better remembered than others (Rule et al., 2012). This shows that although interactions are crucial in the development of a trusting relationship (dependability and faith), first impressions judgments and especially previous experiences with familiar people also influence the decision-making process of the trustor towards another party (predictability).

Trust, Distrust, Untrust, Mistrust

Latest research has started to emphasize the negative aspects of trust as well, namely distrust (Marsh & 2005). While *trust* deals with the confidence and faith towards the future actions of another party, *distrust* is described as an overall feeling of doubt, characterised by suspicion and lack of confidence in someone's future actions. Distrust and trust are both based on previous experiences and interaction with the other party (Marsh & 2005). Among researchers, there is a debate on how trust and distrust are measured. On one hand, researchers argue that the concepts of trust and distrust are two opposites of the same spectrum and are distinct, while on the other hand, another school of thought supports that trust and distrust exist simultaneously (Marsh & 2005; Lewis & Weigert 1984; Lewicki et al. 1998; McKnight & Choudhury 2006; Dimoka, 2010).

Apart from trust and distrust, the concept of *untrust* was developed, to express the situation in which a trustor shows little confidence towards the other party (Marsh & 2005). Untrust is not the opposite of trust, but rather a form of positive trust that is not enough to lead to cooperation. Lastly, *mistrust* has to do with misplacing trust, in situations that for example trust was betrayed (Marsh & 2005). Therefore, in cases of mistrust, a trustor trusted another party that ended up betraying them. Interaction and experience are also required for mistrust since a situation in which the trustee betrays the trustor needs to occur for the second to understand that trust was misplaced.

Human-to-Technology Trust

Technological systems nowadays are used in many cases as a replacement of human-to-human interactions such as shopping online (e-commerce), e-banking, as well as social media and chatting platforms that allow individuals to keep in contact without having a face-to-face interaction. This type of human-to-technology interaction has created a debate among researchers. On the one side of the debate, are those believing that trust can and is being formed between humans and technology, and this is especially apparent on the way people accept, choose and interact with machines (Wang & Benbasat, 2005; Vance et al., 2008; Thatcher et al., 2011). On the other side of the debate, however, are those that do not believe that trust can be formed between humans and machines (Luhmann, 1979; Friedman et al., 2000). These researchers argue that interactions between humans and technology do not only lack the emotional bond created in human-to-human interactions (Luhmann, 1979) but are also not based in a trusted, reciprocal relationship. However, the highly observed use of technology shows that in practice people are, to some extent, able to recognize (un)trust systems and thus avoid negative consequences (Xu et al., 2014). This shows that a certain kind of trust must exist between humans and technology.

Three elements of trust in human-technology interactions

Similarly to human-to-human trust relation, the trust relationship between a human and a machine will also be based on the predictability of the machine's actions (Muir, 1987). To make these predictions, humans will use their previous knowledge with similar systems, the machine's properties, and its environment. This process implies that memory can affect decision making through the past experiences and knowledge that each individual has. As the relationship progresses, trust is based on dependability rather than expectation. Dependability is highly influenced by interactions and experience with the machine, which are built over time (Muir, 1987). Thus, it could be argued that Barber's "three elements of trust" concept (predictability, dependability and faith) could also be applied to human-to-technology interactions.

Barden's "three elements of trust" concept could be further implemented on the different interaction stages in which judgments are formed; (a) before interacting with the system (pre-use trust), (b) while interacting with the system and (c) following interaction (post-use trust). Predictability is related to the pre-use interaction, while dependability and faith are linked more to the stages of interaction and post-interaction. In all the different interaction stages, previous experiences with similar systems are necessary (Borsci et al., 2018; Salanitri et al., 2015; McKnight et al., 2002; McKnight et al. 2011).

Untrustworthy technology

Untrustworthy technology has not dominated the market because people can fairly assess the trustworthiness of a product, even before interacting with it (pre-use trust). The above observation implies the existence of different cues and characteristics which can be linked to trustworthiness. Being able to correctly detect an untrustworthy machine means that the deciding party was able to avoid potentially harmful consequences of trusting an unfaithful system (Huvila. 2017). Before the interaction, a moment of decision making takes place, in which a trustor assesses the trustworthiness of the other party. Thus, this first assessment is a crucial step in a person's future interactions and actions.

This research will focus on the first trust formation (pre-use trust), since it is the point of decision making that establishes interaction. Memory plays an important role during the first trust formation, since the first trust formation is primarily affected by previous knowledge and experiences. Several studies have been conducted on the role of memory in this first assessment towards cheating and uncooperative individuals, which will be reviewed in detail below (Verplaetse et al., 2007; Bell & Buchner, 2012; Oda, 1997; Mealy et al., 1996).

Memory

Memory is the process in which information is encoded, stored and retrieved. The purpose of memory is to store and retrieve information that is useful for future actions (Sherwood, 2016). Sensory inputs pick up information from the outside world and this information is stored in memory through the encoding process (Goldstein, 2015). Following the encoding process, the recently acquired information will be transformed into long-term memory, in a process called memory consolidation (Urcelay & Miller, 2008). Memory consolidation is "time-dependent" and deals with the strengthening of neuron connections, which in turn influence how efficiently information is stored (Urcelay & Miller, 2008).

Emotions can influence the encoding and the consolidation processes, due to their influence on the attendance and perception towards a stimulus (Roesler & MaGaugh, 2019; Phelps, 2004). Highly emotional events lead to a higher degree of arousal and consequently the release of more stress hormones, which are all responses to the events that are being stored (Wout & Sanfey, 2008).

When a stimulus is perceived, brain areas such as the amygdala receive the signal unconsciously and quickly and produce a rapid response to the environment. Highly emotional events are perceived and encoded in memory quicker (Phelps, 2004). Humans are therefore able to unconsciously code, learn and respond to an emotional stimulus. To prove the involvement of amygdala, neurological studies that tested patients with bilateral amygdala damage on their recognition of faces found that these individuals were not able to discriminate between untrustworthy and trustworthy faces (Adolphs, 1998). The study of Winston et al., (2002) also showed that when faces of trustworthy individuals were evaluated, the amygdala and insula were activated. They also found that when the untrustworthiness of the faces was increasing, these brain areas also showed higher engagement (Winston et al., 2002).

Positive vs Negative stimuli detection

Stimuli with positive or negative emotional value differ in their attentional orienting effects as shown in the “face in the crowd effect” (Ohman et al., 2001). The “face in the crowd” effect suggests that a threatening face found between neural faces is detected faster and more accurately than a friendly face. Thus, there is an automatic orienting effect towards negative information, which makes it more recognizable and as a consequence better remembered (Ohman et al., 2001). When a threatening event is experienced, it is more likely to be remembered, due to the intense responses that are associated with it (higher arousal and hormones) (Wout & Sanfey, 2008).

A mechanism that can explain the face in the crowd effect is the cheater-detection mechanism. Specifically, several studies suggest that humans are equipped with a brain mechanism integrated into a cheater-cognitive module that is used in social interactions (Verplaetse et al., 2007). Through this mechanism, people keep track and assess whether the other individual followed or violated social contracts in the past (Cosmides, 1989; Cosmides and Tooby, 1992; Verplaetse et al., 2007; Yamagishi et al., 2003; Buchner et al., 2009; Mehl & Buchner, 2008). This "track-record" is stored in memory, and it aids individuals in their future interactions (Bell & Buchner, 2012). Studies on the cheater-detection-mechanism have shown that faces of uncooperative individuals were better remembered than those of cooperative ones, even after short exposure times (Bell & Buchner, 2012; Oda, 1997; Mealy et al., 1996). The cheater-detection mechanism suggests that people code negative, emotional, information of uncooperative humans and systems to remember them in future interactions (Weymar et al., 2019).

The two mechanisms discussed; “face in the crowd” effect and the cheater detection mechanism, could also be linked with (un)trust. Since trusting an untrustworthy party could potentially lead to harm, it could be assumed that untrustworthy stimuli would elicit more negative emotions and would thus be more recognizable than trustworthy stimuli.

Association, analogies and predictions

Existing knowledge together with incoming sensory information is combined for predictions to occur, which then affect people’s actions and plans. Bar explains this idea by showing that actions are guided by three main factors; *associations, analogies* and *predictions* (Bar, 2007).

Firstly, *associations* are formed by combining a lifetime of repeating patterns and similar situations and storing these in memory. When faced with a new input, an *analogy* is created, which associates this new input with already stored representations in memory. Based on the associated representations that these analogies activate, *predictions* are formed. Actions, plans and thoughts are therefore guided by predictions that are made up from received sensory input and stored representations (Bar, 2007).

It is evident that stored memory guides, to a great extent, thoughts, actions, and emotions of individuals (Bar, 2007; Sherwood, 2016; Tuch et al., 2012). The threatening nature of untrustworthy stimuli and the negative emotions that past experiences with similar stimuli elicit (Rule et al., 2012), suggest that untrustworthy stimuli may be remembered better than trustworthy stimuli. Thus, it is shown that stored memory can further influence the judgment of trustworthiness and untrustworthiness.

Big 5 Personality Traits

Although a negative emotional event is more likely to be recognized and remembered than a positive one, there are further individual differences that can also influence the way that emotional stimuli are perceived and encoded. Different personality traits have been found to affect the way that people judge events and experience emotional stimuli (Zelenski, 2007). Based on the Big 5 model, there are five personality traits;

1. *Extraversion* (or extroversion) refers to highly sociable and outgoing individuals
2. *Agreeableness* is linked with altruistic and sympathetic behaviours (Alarcon et al., 2018)
3. *Conscientiousness* refers to highly competent individuals that prefer planning and carefully thought decisions
4. *Neuroticism* characterises individuals that experience more stress and emotional arousal
5. *Intellect Imagination* (or *openness*) characterises individuals that are open to new ideas and experiences (Freitag & Bauer, 2016)

Different emotional responses towards stimuli have been linked to different personalities. For example, extraversion has been found to result in more positive and intense emotions, while neurotic individuals experience more intense and negative emotions (Zelenski, 2007). Since different personality traits result in the experiences of different emotional responses, it can be said that emotions mediate the effect of personality on decision making and judgments (Zelenski, 2007).

Effect of personality on memory

There are cognitive differences between personality traits. Appraisal and memory are cognitive functions which are influenced by personality traits and at the same time closely linked to emotions (Zelenski, 2007). Personalities not only affect people's expression of highly affective events but also milder ones. For example, extraverts will all-in-all judge a mild affective situation as more positive compared to more introverted individuals.

Personality and Trust

As far as personalities and trust predispositions are concerned, agreeableness is the trait most closely linked to trusting behaviours (Alarcon et al., 2018). People scoring high in this trait, are more caring and place more importance in their relationships, and are found to score higher on trust (Freitag & Bauer, 2016). Extraverted individuals, due to their highly social nature are presumed to score higher in trust than introverts. Introverts are presumed to be selective, and less comfortable when surrounded by people, and therefore are expected to spend more time analyzing the person and/or product they are interacting with (Freitag & Bauer, 2016).

On the other hand, people high in conscientiousness make careful decisions, and generally do not trust easily, therefore these individuals score lower in trust. Lastly, people scoring high in openness tend to score higher in trust as well as risk-taking behaviours, due to their overall open-minded nature (Freitag & Bauer, 2016).

Goal of Current Experiment

Despite the latest integration of technology in human life, untrustworthy devices and systems have not dominated the market. This suggests that people are intuitively able to recognize trustworthy from untrustworthy technology. The aim of the current research is to explore the factors that can affect peoples' recognition of (un)trustworthiness, in an attempt to understand how these judgments are formed.

First research question

Negative emotions and threatening experiences are found to be better remembered and recognized than neutral or positive stimuli. The hypothesis of the current paper is that an untrustworthy stimulus, due to its threatening nature, is more recognized than a trustworthy stimulus. Furthermore, the face-in-the-crowd effect and the cheater-detection mechanism have shown that people can detect and recognize better non-cooperative individuals. This work explores whether a similar mechanism is active when people are dealing with stimuli that convey features of (un)trust.

The first exploratory question that this study aims to answer is (1) *do trustworthy and untrustworthy stimuli show differences in their memorability?* To answer this question the following hypothesis is formed:

After a certain time from the exposure to trustworthy and untrustworthy stimuli, there is a difference in the accuracy that such stimuli are recognized.

Second research question

Individual differences exist that affect how an emotional stimulus is perceived, encoded and thus remembered. Different personality traits lead to different emotional responses and also past experiences can affect future judgments. This research explores the effect of the following individual characteristics into the recognition of trustworthy and untrustworthy stimuli:

(a) Faith in General Technology, which deals with users' idea that general technology is consistent, reliable and provides the required features (Mcknight et al., 2011);
(b) Trust Stance, which deals with users' assumption that interaction with technology will bring them a positive outcome (Mcknight et al., 2011)
(c) The Big 5 Personality Traits, namely: extraversion, agreeableness, conscientiousness, emotional stability and intellect imagination.

The second research question of the current work, is related to the effect of individual characteristics on the recognition of trustworthy and untrustworthy stimuli; *(2) Do the individual characteristics as described above (Faith in General Technology, Trust Stance and Personality) affect the recognition of untrustworthy and trustworthy photos?*

Specifically, the questions can be summarized as follows:

- Faith in General Technology: *The recognition of trustworthy versus untrustworthy stimuli is affected by people's overall Faith in General Technology*
- Trust Stance: *The recognition of trustworthy versus untrustworthy stimuli is affected by people's overall Trust Stance*
- Personality: *The recognition of trustworthy versus untrustworthy stimuli is affected by people's personality types.*

Novelty of work

Previous studies on the recognition of cooperative and uncooperative individuals showed that uncooperative visual stimuli were better remembered. However, the visual stimuli used in these studies were complemented by captions providing supplementary information such as the socioeconomic status of the stimulus (Oda, 1997; Mealey et al., 1996). It can be argued that the judgements made in these studies could be biased due to the supplementary information provided. In an attempt to reflect more the way that first impressions are formed in real life, where no extra information is available to the decision-maker, this research employs a visual stimulus consisting of a standardised dataset of photos of faces, scenes and products without extra descriptions, and without being asked to remember the stimuli in any way.

Signal Detection Theory and Bayes Analysis

To test people's recognition of old and new stimuli after a certain time from the initial exposition, Signal Detection Theory (SDT; Macmillan & Creelman, 2004) is used. SDT focuses on the process of decision making which is based on uncertainty. The SDT approach is chosen to implicitly examine the participants' recognition and the effect(s) of the stimulus' appearance cues.

In combination with Signal Detection theory, a Bayesian approach is used to analyze the results. One of the reasons why the Bayesian Regression analysis is chosen over the Frequentist Analysis in this study is that it considers prior knowledge. Specifically, Bayesian

Statistics takes into account; (1) priors, which are based on previous knowledge and (2) likelihood, which is the data being explored. These two together (prior and likelihood), lead to the posterior, which is the inference that we are interested in. The priors are a great tool through which existing knowledge can be incorporated in the data analysis.

Method

Design

A within-subjects design was used in this study, in which subjects were presented with two different types of stimuli: i) trustworthy photos, and ii) untrustworthy photos. The dependent variable was each participant's sensitivity in recognizing the stimuli, measured by calculating the dprime.

Participants

A total number of 85 participants (Male: 42, Female: 43) were recruited by convenient and snowball sampling. Inclusion criteria for participants were to be above 18 years old and have intermediate language proficiency in English. One participant that was below 18 had to be removed resulting in a total of 84 participants. All the participants signed the informed consent before participation, agreeing to take part in the study.

Materials & Apparatus

Stimuli

The experiment was carried out with a digital cross-platform created using the open-source software package 'PsychoPy3'[1] a set of questionnaires, and additional materials. The study was carried out in two stages; the Pre-Trial stage and the Trial stage, which will be described in the following section. The Pre-Trial stage contained a total of 40 images of Flags. The Trial stage contained a total of 80 images of Faces, Scenes, and Products. The images containing faces were taken from the Chicago Face Database (Chicago Face Database, 2018), comprising a set of 40 images of people's faces that were already categorized as either trustworthy or untrustworthy (20 trustworthy and 20 untrustworthy images). The images containing scenes were taken from the Socio-Moral database (SMID; Crone et al., 2018), and were depicting different scenes (20 fair/trustworthy and 20 unfair/untrustworthy images). Lastly, 40 images of products were used (20 trustworthy / 20 untrustworthy) **(**CPSC, n.d.), which contain products that had been on the market in 2017 and other products that had not been on the market but had been characterized as problematic and dangerous.

[1] Provided by Dr. Simone Borsci.

Questionnaires

To tackle the second exploratory question and test the individual characteristics of (1) Faith in General Technology, (2) Trust Stance and (3) Personality, a set of questionnaires was prepared:

(a) A *Demographics* questionnaire, including questions on the age, gender and nationality of the participants (Appendix A)
(b) An *'Attitudes Towards Technology'* questionnaire containing 12 items in a 5-Likert Scale (Appendix B)
(c) The *IPIP* questionnaire assessing the Big 5 Personality Traits through 50 items (10 items per personality trait), in a 5-Likert scale from Very Inaccurate to Very Accurate (Appendix C)
(d) Two parts of the *'McKnight'* questionnaire (McKnight et al., 2011) for Trust in Technology: the 4-item Faith in General Technology scale, and the 3-item Trusting Stance/ General Technology Scale (Appendix D).

Additional materials were used, such as an information sheet containing important information about the study, and the informed consent of the participants (Appendix E).

Hardware & Software

An Apple iMac desktop was used for the testing (27-inch, Late 2012). For the remote testing, an Apple MacBook Pro laptop was used (13-in, Mid 2012), together with the Google Hangouts application for screen sharing.

For the execution of the experiment, the open-source software package PsychoPy3 was used, while for the statistical analysis the RStudio (Version 1.2.5042) program was used.

Video

In order to prevent the registration of the stimuli in the long term memory, a TEDx video (titled "The Power of Vulnerability") was used as a distraction between the Pre-Trial and the Trial stage. The video was retrieved from the TEDx website[2], with the possibility of subtitles based on the participants' preferences.

[2] Accessed February 25, 2020:
https://www.ted.com/talks/brene_brown_the_power_of_vulnerability?language=en)

Ethical Approval

The study got Ethical Approval from the Ethical Committee of the University of Twente (Project ID 1582296162).

Procedure

Firstly, the participants were presented with the information sheet and given the informed consent to sign. Then the PsychoPy platform was launched and the experiment started.

The participants were first presented with the "Pre-Trial" stage, which was programmed to include two rounds that presented neutral stimuli (flags). The aim of this stage was to familiarise the participants with the platform. Furthermore, the answers from this stage were used for subsequent statistical calculations. The first round included the presentation of a set of 20 flag images in random order, the one after the other, each presented for 3 seconds. The second round included a set of 20 flag images, out of which 10 had already been presented in the previous round, and 10 had not been shown before. After each image in the second round, the platform was programmed to include the question 'Have you seen this image before' with a binary answer set with the options 'Yes' and 'No'.

Following the "Pre-Trial" stage, the "Trial" stage started. Two rounds were also included in this stage. For each participant, in the first round the platform presented a total of 60 images of Faces, Scenes, and Products randomly selected from the aforementioned repository. The images were presented in a random order to the participants. Following the presentation of these images, a 30-minute break took place, in which the participants were asked to watch the TEDx video (titled "The Power of Vulnerability"). After the break, the second round started, in which, another set of 60 images was presented to the participants, which contained 30 pictures that were presented in the previous round (old) and 30 new photos. After each image, the platform was programmed to include the same question as before, 'Have you seen this image before' with a binary answer set with the options 'Yes' and 'No'.

After completing the experiment in PsychoPy, the participants were asked to fill in the *Demographics* questionnaire, the *'Attitudes Towards Technology'* questionnaire, the *IPIP* questionnaire, and the two parts of the *'McKnight'* questionnaire. Finally, participants were debriefed regarding the true aim of the experiment.

It should be noted that due to the new regulations for COVID-19, some experiments had to be conducted remotely, with the use of Google Hangouts. When this method was applied, the researcher shared their computer screen with the participant, through which the PsychoPy experiment was presented. The participants then conducted the experiment by saying to the researcher which keys to press, until the experiment was completed. The questionnaires were shared to the participants through a link, which still allowed the participants to answer them on their own.

Data analysis

Prior to the analysis, the data was prepared. Specifically, all raw data from the PsychoPy Platform and from the Questionnaires were merged into one excel document. Each column included 84 rows indicating each participant.

By using the Signal Detection Theory, the Hit, Miss, False Alarm and Correct Rejection were calculated for each participant (Macmillan & Creelman, 2004). The *Hit* shows the correct recognition of an Old stimulus, while *Miss* depicts the inability to recognize an Old stimulus. *False alarms*, describes the faulty recognition of a New stimulus as being an Old item. Lastly, the *Correct Rejection* is when the participant correctly answered No to a New item (Macmillan & Creelman, 2004). Once this table is made for each participant separately, the response rates were normalised. The hit rate (H) shows the proportion of Old trials in which the participant correctly responded positively, while the False-Alarm Rate (F) shows the proportion of New items in which the participant faulty responded positively. The H and F of each participant indicate their performance and were used to calculate each observer's sensitivity as described below (Macmillan & Creelman, 2004).

In Signal Detection Theory, the sensitivity is measured with the *d'prime*. For the *d'prime* to be calculated the Hit and False-Alarm rates are used in terms of *z* scores, converting them into a standard deviation unit, using equation (1) (Macmillan & Creelman, 2004):

$$d' = z(H) - z(F) \qquad (1)$$

A *d'prime* of zero $d'=0$, shows a participant that cannot discriminate the stimuli at all, and thus $H=F$. This means that the participant had an equal rate of saying yes both for Old and for New stimuli.

Firstly, the *d'prime* was calculated for the answers of the Pre-trial stage of the flags, *d'* Pre-Trial-Flags (DprimePre). The *d'prime* of the untrustworthy and trustworthy stimuli was also calculated, for both the stimuli that includes all the image types and for each image type separately;

1. *d'* Trust (DprimeTrust)
2. *d'* Untrust (DprimeUntrust)
3. *d'* Face-Trust (DpimeFT)
4. *d'* Face-Untrust (DpimeFU)
5. *d'* Scene-Trust (DpimeST)
6. *d'* Scene-Untrust (DpimeSU)
7. *d'* Device-Trust (DpimeDT)
8. *d'* Device-Untust (DpimeDU).

The finalized excel file was imported into RStudio (Version 1.2.5042). The data from all the participants was taken into account since none of them had a sensitivity score below zero for the Flag stimuli.

Using this data file, the Bayesian Analysis was conducted. The main R package used was "brms". Since Bayesian Statistics are a combination of a prior distribution and the obtained likelihood (data), the priors had to be determined. Because this study aims at exploring the differences in recognition of Trustworthy and Untrustworthy stimuli, the priors selected were derived from the participants performance on the Pre-Trial stage. Specifically, the mean and standard deviation of the Pre-Trial *d'prime* (DprimePre) were used as previous knowledge for the memory models. The code for setting up the priors can be found in Appendix F.

After setting the priors, the models had to be specified, by setting the dependent variable (d'prime), the variables of interest (separated by "~"), and the independent variables. The posterior of the model was then derived. The output of the model showed the posterior parameter for the Intercept and the other variables tested in each model as well as the 95% Credibility Interval for each variable. If the Credibility Interval contained a zero value, then no effect was concluded. The standard deviation of the model (Sigma), which indicates how strong the model is in terms of fitting, and therefore how satisfactory the model is, was also calculated. An example of the models used to measure the effect of General Trustworthy stimuli and General Untrustworthy stimuli on DprimePRE (memory ability) can be found in Appendix G.

The Bayes Factor was also calculated using the "brms" package in R for all the models. In Bayesian Statistics the Bayes Factor is used to give support for a model over another model

(Verhagen & Wagenmakers, 2016). The Bayes Factor is usually used for the comparison of a null and an alternative hypothesis. In the current study, to calculate the Bayes Factor the null hypothesis models were used, over the models testing the differences between the Dprime of Untrustworthy stimuli and the Dprime of Trustworthy Stimuli. An example of the models used to measure the effect of General Trustworthy stimuli and General Untrustworthy stimuli on DprimePRE (memory ability) can be found in Appendix H. For these models, the neutral items were used, without the priors. The approach of Kass & Raftery (1995) was used for the interpretation of the Bayes Factors, intended as an index that shows how much evidence the alternative hypothesis has over the null hypothesis (Table 1).

Table 1

Bayes Factor model by Kass & Raftery (1995). *The Bayes Factor is used to quantify support from one model (null model) over another (alternative model). When two models are compared the output of the Bayes Factor shows whether there is negligible (1-3), positive (3-20), strong (20-150) or very strong evidence (>150) in favour of one model over the other.*

Bayes factor	**Interpretation**
1 - 3	Negligible evidence
3 - 20	Positive evidence
20 - 150	Strong evidence
>150	Very strong evidence

Results

Memory

To explore the memorability of untrustworthy and trustworthy stimuli, a Bayesian regression was conducted between the *d' prime* Pre-Trial (memory index) and the different general stimuli variables: i) *d' prime Trustworthy Photos* (DprimeTrust), ii) *d' prime Untrustworthy Photos* (DprimeUntrust).

Table 2, shows that when compared with Dprime values from situations in which stimuli are neutral (DpimePre from the flag trial), the Untrustworthy stimuli are better recognized (0.25, [0.06 – 0.44]). No conclusion can be drawn for the recognition of Trustworthy stimuli (0.15, [-0.04 – 0.34]), since the Credibility Interval contains a zero.

Table 2

Posterior Results from Bayesian Regression Analysis between DprimePre [0.97] (d' prime for Memory Ability), DprimeTrust [0.25] (d' prime for Trustworthy Stimuli) and DpimeUntrust [0.15] (d' prime for Untrustworthy Stimuli).

	Estimate	Est. Error	Q2.5	Q 97.5
b_Intercept	0.97	0.13	0.71	1.22
b_DprimeUntrust	0.25	0.09	0.06	0.44
b_DprimeTrust	0.15	0.09	-0.04	0.34
Sigma	0.37	0.03	0.32	0.43

Note. If Credibility Interval [Q 2.5- Q 97.5] includes a zero (0), no effect is concluded.

The estimated Bayes Factor calculated between the Untrustworthy General stimuli and the Null hypothesis was found to be 83.53 in favour of the Untrustworthy stimuli. The Bayes Factor for the Trustworthy stimuli was calculated even though the Credibility Interval contained zero, and was found to be 12.31 in favour of the Trustworthy Stimuli, showing small positive evidence against the Null hypothesis.

To directly estimate whether there is a difference in the recognition of untrustworthy and trustworthy stimuli, a Bayesian model was run between *d'prime* Untrust and *d'prime* Trust. Table 3 shows that the posterior mean of DprimeUntrust (0.59) is higher than the DprimeTrust (0.52). The estimated Bayes Factor calculated between the model and the Null hypothesis was found to be 14.35 against the Null Hypothesis which shows a positive effect in favour of the model, suggesting, in line with the results in Table 2, that there is a difference between the recognition of trustworthy and untrustworthy stimuli.

Table 3

Posterior Results from Bayesian Regression Analysis between DprimeUntrust [0.59] and DprimeTrust [0.52]. DprimeTrust being the Intercept.

	Estimate	Est. Error	Q2.5	Q 97.5
b_Intercept	0.52	0.13	0.25	0.77
DprimeUntrust	0.59	0.09	0.41	0.77
Sigma	0.42	0.03	0.36	0.49

Note. If Credibility Interval [Q 2.5- Q 97.5] includes a zero (0), no effect is concluded.

To examine whether there is a similar mechanism to the cheater-detection assumption when dealing with Trustworthiness and Untrustworthiness, the photos of Faces were also explored on their own. Table 4 shows that the estimate for the recognition of Trustworthy Photos of Faces (0.15, [-0.02 – 0.32]) was higher than the estimate of the recognition of Untrustworthy Photos of Faces (0.06, [-0.12 – 0.24]). The estimated Bayes Factor calculated between the Untrustworthy Face stimuli model and the Null hypothesis was found to be 0.3886 in favour of the Untrustworthy model. The Bayes Factor for the Trustworthy Face stimuli model was found to be 1.010 in favour of the Trustworthy Stimuli.

Table 4

Posterior Results from Bayesian Regression Analysis between DprimePre [1.33] (*d' prime* for Memory Ability), DprimeFU [0.06] (*d' prime* for Untrustworthy Faces) and DprimeFT [0.15] (*d' prime* for Trustworthy Faces).

	Estimate	Est. Error	Q2.5	Q 97.5
b_Intercept	1.33	0.09	1.16	1.49
b_DprimeFU	0.06	0.09	-0.12	0.24
b_DprimeFT	0.15	0.09	-0.02	0.32
Sigma	0.39	0.03	0.34	0.46

Note: If Credibility Interval [Q 2.5- Q 97.5] includes a zero (0), no effect is concluded.

For Scenes the difference between Trustworthy and Untrustworthy stimuli was found to be minimal. The estimates were negligible, with only a 0.01 difference between the Untrustworthy and the Trustworthy photos of Scenes. The posterior mean of Untrustworthy photos over memory ability was 0.19, while the posterior mean of Trustworthy photos was found to be 0.21. Thus, photos of Scenes did not result in any recognition differences between the Trustworthy and Untrustworthy stimuli. The Bayes Factor of the model exploring the Trustworthy photos of Scenes over the Null Hypothesis was estimated to be 8.16. The Bayes Factor of the Untrustworthy model of Scenes over the Null Hypothesis was 12.84. Both in Bayes Factors were in favour of the models showing positive evidence against the null hypothesis and thus a partial effect for the Trustworthy and Untrustworthy photos of Scenes on memory.

For Devices, a small effect was found on the Untrustworthy photos of Device over Memory ability (0.20, [0.04 – 0.35]). The Bayes Factor for the model exploring the Untrustworthy photos of Devices over the Null Hypothesis was estimated to be 2.25 in favour of the model, showing negligible evidence against the Null hypothesis and thus no effect. No effect was reported for the Trustworthy stimuli (0.12, [-0.04, 0.29]) since the Credibility Interval contained a zero value. The low Bayes Factor of the Trustworthy model, also shows the small effect of the trustworthy photos of devices, with the value being 0.57, suggesting no evidence of an effect. The full code and output of the models exploring memory ability for (i) all the photos, (ii) photos of faces, (iii) photos of scenes and (iv) photos of devices can be found in Appendix H.

Individual Factors

The second aim of this research was to investigate other individual factors that can mediate the recognition of Trustworthy and Untrustworthy stimuli such as Faith in Technology, Trust Stance and Personality traits.

Faith in Technology and Trust Stance

Since both Faith in General Technology and Trust Stance deal specifically with attitudes towards devices and technological products only the Devices were explored for these two variables. Neither Faith in Technology (Table 5) nor Trust Stance (Table 6) showed any effect for the different stimuli, with all the Credible Intervals containing zero. For Faith in Technology, the Bayes Factors over the Null Hypothesis were found to be 2.50 in favour of the

Untrustworthy model of Devices, and 2.30 for the Trustworthy model of Devices. For Trust Stance, the Bayes Factor of the Trustworthy model of Devices was found to be 3.47 and for the Untrustworthy 2.56. All of these Bayes Factors, show negligible evidence in favour of the alternative hypothesis, and thus no effects can be concluded between Faith in Technology and Trust Stance over recognition of Trustworthy and Untrustworthy photos.

Table 5

Posterior Results from Bayesian Regression Analysis between Faith in Tech [29.73] (Intercept), DprimeDU [-0.49] (*d' prime* for Untrustworthy Stimuli of Devices) and DprimeDT [-0.21] (*d' prime* for Trustworthy Stimuli of Devices).

	Estimate	Est. Error	Q2.5	Q 97.5
b_Intercept	29.73	1.21	27.35	32.02
b_DprimeDU	-0.49	0.84	-2.09	1.18
b_DprimeDT	-0.21	0.85	-1.89	1.53
Sigma	3.85	0.30	3.309	4.49

Note: If Credibility Interval [Q 2.5- Q 97.5] includes a zero (0), no effect is concluded.

Table 6

Posterior Results from Bayesian Regression Analysis between Trust Stance [18.95] (Intercept), DprimeDU [0.30] (*d' prime* for Untrustworthy Stimuli of Devices) and DprimeDT [-0.85] (*d' prime* for Trustworthy Stimuli of Devices).

	Estimate	Est. Error	Q2.5	Q 97.5
b_Intercept	18.95	1.61	15.79	22.03
b_DprimeDU	0.30	1.11	-1.95	2.48
b_DprimeDT	-0.85	1.15	-3.19	1.40
Sigma	5.02	0.39	4.32	5.87

Note: If Credibility Interval [Q 2.5- Q 97.5] includes a zero (0), no effect is concluded

Personality

The five personality traits of the *IPIP* questionnaire assessing the Big 5 Personality Traits (extraversion, agreeableness, conscientiousness, emotional stability and intellect imagination) through 50 items (10 items per personality trait), were also explored. The results on the different personality traits, showed only minimal effects. For each model, the different personality types were taken as Intercept.

The table below (Table 7) shows the output results for each model; Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellect Imagination.

Table 7

Posterior Results from Bayesian Regression Analysis between Extraversion, Agreeableness, Conscientiousness, Emotional Stability, Intellect Imagination (Intercepts), DprimeUntrust (*d'prime* for Untrustworthy Stimuli) and DprimeTrust (*d'prime* for Trustworthy Stimuli)

Personality		Estimate	Est. Error	Q2.5	Q 97.5
Extraversion					
	b_Intercept	73.25	3.77	65.82	80.65
	b_DprimeUnrust	-2.59	2.98	-8.46	3.27
	b_DprimeTrust	-0.76	2.94	-6.31	5.08
	Sigma	10.86	0.88	9.32	12.76
Agreeableness					
	b_Intercept	81.05	4.06	73.05	88.98
	b_DprimeUntrust	-4.13	3.17	-10.21	2.27
	b_DprimeTrust	1.43	3.07	-4.64	7.41
	Sigma	11.72	0.93	10.07	13.70
Conscientiousness					
	b_Intercept	72.34	4.62	62.95	81.23
	b_DprimeUntrust	-1.20	3.67	-8.33	5.83
	b_DprimeTrust	-1.64	3.63	-8.52	5.61
	Sigma	13.34	1.07	11.46	15.57
Emotional Stability					
	b_Intercept	61.22	5.06	50.95	71.10
	b_DprimeUntrust	-5.39	4.02	-13.23	2.37
	b_DprimeTrust	7.27	3.93	-0.55	14.99
	Sigma	14.81	1.18	12.63	17.32
Intellect Imagination					
	b_Intercept	76.43	3.59	69.22	83.67
	b_DprimeUntrust	-2.92	2.78	-8.43	2.37
	b_DprimeTrust	2.05	2.74	-3.35	7.68
	Sigma	10.46	0.84	8.95	12.18

Note. If Credibility Interval [Q 2.5- Q 97.5] includes a zero (0), no effect is concluded.

The Estimates from all the models' output, exploring the relationship of Trustworthy and Untrustworthy photos over each specific Intercept (each personality trait) had small variations, that are not enough to defy a change in personality, thus the results were negligible. All the Credibility Intervals also included a zero, which shows no effect of the model. All the Bayes Factors the Personality Trait models, together with their code and output can be found in Appendix J.

Discussion

This paper examined two exploratory questions regarding the recognition of trustworthy and untrustworthy visual stimuli: whether there is a difference in the memorability of trustworthy and untrustworthy stimuli and secondly whether individual characteristics (Faith in General Technology, Trust Stance, Personality) can explain the recognition of trustworthy and untrustworthy stimuli. The focus of the current research was on the first trust formation (pre-use trust), since it is the stage of decision making in which interaction is established (Muir, 1987; Borsci et al., 2018; Salanitri et al., 2015; McKnight et al., 2002; McKnight et al. 2011). To explore the two research questions, a Signal Detection experiment was conducted, in which visual stimuli were presented to the participants. The presented stimuli were not accompanied with any supplementary captions, in an attempt to reflect the way that first impressions are formed in everyday life situations.

The exploration of the first research question showed that untrustworthy stimuli is better remembered in the experiment compared to the distribution in the pre-test, in which all the stimuli were neutral. The Bayes Factor (83.13), provides strong evidence towards the hypothesis that the recognition of untrustworthy stimuli can be explained by memory ability. When the difference between the recognition of trustworthy and untrustworthy stimuli was explored, the results showed that untrustworthy stimuli were better remembered than trustworthy stimuli. The Bayes Factor (14.38) shows positive evidence towards the hypothesis that a difference exists between the recognition of trustworthy and untrustworthy stimuli. These results are in line with the cheater-detection mechanism (Verplaetse et al., 2007) and the face-in-the-crowd effect (Ohman et al., 2001), which shows that a negative stimulus is all-in-all more detectable and recognizable than a neutral or positive stimulus.

When the different image types were explored (face, scenes and devices), the results did not show any evident effects. Despite the positive evidence that trustworthy and untrustworthy stimuli lead to differences in recognition, the detailed exploration of each specific image type showed negligible effects. These results are not in line with the results of Volonasi & Borsci (2019), Oda (1997) and Mealey et al. (1996). In their study, Volonasi & Borsci (2019) found that participants were able to detect cheater devices and rank them as the least trustworthy only based on aesthetics. Moreover, the studies of both Oda (1997) and Mealey et al. (1996), also showed that faces of cheaters were better remembered than faces of cooperative individuals.

However, in their studies, Oda and Mealey et al. used supplementary fictitious information to describe how cooperative the person in the image is.

A reason why the current research did not show similar results to Volonasi & Borsci (2019), Oda (1997) and Mealey et al. (1996) is because in the aforementioned studies the participants were either explicitly asked to rank the stimulus based on its trustworthiness (Volonasi & Borsci, 2019), or were presented with supplementary information about the stimulus (Oda, 1997; Mealey et al. 1996). As it was shown from Muir (1987), when people make trustworthiness judgments towards another party or stimuli, they consider how predictable the other party's future actions will be. However, in the current experiment, trustworthiness judgments were implicitly examined. Thus, participants did not have to think about the stimulus predictability in future actions, since they did not have to interact with it, or were not asked to explicitly think about its trustworthiness. It can therefore be argued that the participants' trustworthiness judgments were not fully activated through this implicit exposure to the stimuli.

The study of Suzuki and Suga (2010) provides supportive evidence to understand why the participant's trustworthiness judgments were not triggered in the current study. Suzuki and Suga showed that faces of trustworthy individuals were remembered more than faces of untrustworthy individuals, but only in the cases that these individuals behaved in an incongruent manner during an interaction. In a similar approach, Buchner et al. (2009) showed that the context in which encounters took place (i.e. source memory) was more valuable than recognition memory (Buchner et al., 2009). In the current study, the participants did not interact with the presented stimuli, and thus the potential behaviour of the stimulus could not be sufficiently assessed.

The need of interaction as shown in the studies of Suzuki and Suga (2010) and Buchner et al., (2009) demonstrates that context plays an important role in the recognition of (un)trust stimuli. The importance of context in the formation of (un)trust judgments, shows that appearance alone is probably not enough to assess trustworthiness and untrustworthiness. The importance of context was underestimated during the design of the current study, which solely focused on the effect(s) of appearance cues on recognition.

To answer the first exploratory question on whether there is a difference in the recognition of (un)trustworthy stimuli, the results showed positive evidence of an ability to recognize untrustworthy stimuli over trustworthy. However, no evidence was found for each image type separately, and thus no conclusions can be made on whether trustworthy and untrustworthy faces, scenes and devices lead to differences in recognition.

The second exploratory question aimed at finding individual factors that can affect the recognition of trustworthy and untrustworthy stimuli. The explored factors were i) Faith in General Technology, ii) Trust Stance and iii) The 5 Personality Traits (extraversion, agreeableness, conscientiousness, emotional stability and intellect imagination). Faith in General Technology and Trust Stance were only tested for the photos of devices, since these two variables are only related to the attitudes of people towards technology. The results of these tests showed no clear effect. Thus, the recognition of (un)trustworthy stimuli was not affected by either Faith in Technology nor Trust Stance, and thus the hypothesis was rejected. Familiarity can explain the no effects between untrustworthy and trustworthy photos of Devices. Specifically, as Luhmann (1979) states, familiarity "is the precondition for trust as well as distrust" (Luhmann, 1979). The Devices used in this study ranged from children toys, monitors, and cribs, to medical chairs and hearing aids. The majority of the participants of this study were university students. Thus, it can be assumed that this specific target audience was not familiar with many of the presented products, and therefore there was no distinction between the products.

As far as personality is concerned, all the credibility intervals from the results contained a zero, and thus no clear effects can be concluded on the recognition of (un)trust and personality. The differences between the recognition of trustworthy and untrustworthy stimuli against the different personality traits were not sufficient to suggest a difference of personality. Although different personality traits result in the experience of different emotional responses (Zelenski, 2007), due to the implicit exposure of the stimuli and the lack of interaction or conscious thinking towards it, it can be argued that there were no experiences created for personality differences to express themselves.

Overall the results of the current study suggest partial evidence towards the hypothesis that untrustworthy stimuli are better remembered and that there is a difference in the recognition of (un)trustworthy stimuli. However, this difference seems to be not attributable to the specific type of stimuli (faces, scenes and devices). At the same time, the exploration of individual factors (Faith in General Technology, Trust Stance and Personality) did not show any differences for (un)trust recognition. One of the reasons why the current study did not show significant effects when the image types were explored separately, might be that the differences between the trustworthiness and untrustworthiness of the stimuli used were not distinct enough, indicating again the importance of existing contextual factors. Moreover, this study was based solely on the effects of appearance on the judgment. However, even when people encounter others briefly, first impressions are formed with the belief of potential future interactions. Encountering images in an implicit recognition test, in which the participants did not have to interact with the target for example through a Trust Game, or did not have to rank the stimuli on factors like trustworthiness (something that can stimulate conscious thinking), lead to the assumption that the participants probably did not even assess the trustworthiness of the stimulus.

During first impression judgments, humans assess whether the environment is safe enough for future cooperation. Thus, they have a future goal in mind that they have to explicitly decide upon; to interact with the other person and/or product or not. Trust, as Button (2006) stated, is a highly "occasioned matter" (Clarke et al., 2006). This means that people do not constantly ask themselves if they should trust another party or not. On the other hand, people in their everyday life make observations of their surroundings, and question trust only in specific occasions or under particular circumstances which result in unexpected events. By viewing trust as an occasioned matter, the importance of context, occasions and experiences becomes evident. People observe and learn from their context and experiences and they use these to guide their (un)trust judgments.

Limitations of Current study

Due to the new Covid-19 regulations, the biggest part of the study had to be conducted remotely. In order for the experiment to be conducted, the PsychoPy program had to be installed. To prevent the need for the participants to download the program and run the experiment on their own, without the researcher's guidance, the researchers decided to conduct the virtual experiments through their own computers by sharing their screen to the participants. However, sharing the screen led to some limitation. Firstly, bad Wi-Fi connections and in general sharing connections can lead to lower image quality, screen freezings and delays. Since the photos are presented for only 3 seconds each, if a loading delay is experienced, it will result in an even lower exposure time for those images. Therefore, although the experiment was programmed to present each image for 3 seconds, delays in loading time could affect the exposure time of some images resulting in an uneven exposure time of the stimuli.

Moreover, during the second part of the study, in which the participants had to indicate if they had seen the pictures before or not, the participants could not do that on their own but rather had to share their answer to the researcher who then clicked the corresponding keys (Y for yes and N for no). The fact that the participants did not have the room to conduct the experiment on their own, could have lowered their involvement in the experiment. Moreover, the need of the participants to share their answers to the researchers, could have also affected their feeling of privacy.

By conducting the experiment remotely, the researchers had also no control over the participants' focus during the study. This means that the researchers could not control whether the participants were actually looking at the screen when the photos were presented or whether they watched the video during the break. At the same time, the researchers could not control the use of other electronic devices during the experiment such as mobile phones. All-in-all the current study was a lab experiment, and although it aimed to resemble the way that first impressions are formed, the highly-control environment of the study does not entirely represent the way that encounters and first impressions are formed in the real world. Moreover, since the necessary remote execution of the experiment can affect the participants' involvement and focus on the study, their (un)trust formation towards the stimuli can be questioned.

A further limitation of the study is that the presentation of trustworthy and untrustworthy stimuli was not equal. Specifically, in both rounds of the Trial stage, the participants were presented with 60 photos in total. After analysing the results, it became apparent that the distribution between trustworthy and untrustworthy stimuli was not equal. Thus, there was an imbalance in the total number of trustworthy and untrustworthy photos that each participant got.

Future Research

This study found only partial evidence that there is a difference in the recognition of untrustworthy and trustworthy stimuli, with no effects being found for the specific image types (faces, scenes and products) and the individual characteristics (Faith in General Technology, Trust Stance and Personality). With the latest increase in Machine Learning and Artificial Intelligence, technology is embedded more and more into people's everyday life, by also making it more challenging for humans to discern between a machine and a person. Since trust is essential for human-to-human interactions in avoiding threatening encounters, trust formations towards technology should also be explored, to understand how users can detect cheating systems.

This research aimed to explore whether there is a difference in the recognition of trustworthy and untrustworthy stimuli, that can be explained through the stimulus memorability, by using an implicit recognition experiment. However, judgments of (un)trust are more complex, affected by not only implicit factors, but also explicit, conscious thinking, appearance, visual complexity to name a few.

Studies have shown that facial characteristics and expressions can be used to infer several personality traits, such as trustworthiness and cooperation, all crucial during social interaction (Farmer et al., 2013; Todorov, 2008). Attractiveness is one of the facial characteristics available even prior interaction and it is considered as a contributing factor in detecting cheating and cooperative individuals (Yamagishi et al., 2003; Verplaetse et al., 2007; Wout & Sanfey, 2008). The study of Shepherd and Ellis (1973) showed that higher and lower attractive faces were more recognized than those of moderate attractiveness (Yamagishi et al., 2003; Shepherd & Ellis, 1973). The eye region is elemental to the face's recognisability, and it is another characteristic used when judgments are made (Perlman et al., 2009). Humans, especially adults, have been shown to fixate mostly on the eye-region, explained by the fact that the eyes are a great tool in communicating emotions (ex. fear) which can lead to increased emotional arousal. At the same time, by fixating on the eyes during social interactions, people express their attentiveness to the other person (Perlman et al., 2009). Although this research initially aimed to test the fixation of individuals when presented with the stimuli (during both encoding and recall), due to the Covid-19 regulations of social distancing, the eye-tracking analysis had to be cancelled. However, based on the evidence that the eye region is a crucial contributor to our assessment, further research using eye-tracking analysis should be conducted, in an attempt to explore whether there are fixation differences between trustworthy and untrustworthy stimuli.

Another factor that can affect the encoding and recognition processes is also mood (Zelenski, 2007). Research in mood-congruent memory has shown that people's judgments are affected by their mood. The study of Jonshon and Tversky (1983) showed that the participants who were in a negative mood rated unrelated negative events as more likely to happen to them. Even though the current research explored whether there are differences in the recognition of (un)trust due to personality characteristics, the participant's mood and emotions were not assessed. Studies have shown that emotions are not only affected by personality differences but they also influence the assessment process (Roesler & MaGaugh, 2019; Phelps, 2004; Zelenski, 2007). Thus in future studies, the emotional responses and the mood of the participants should be explored in more detail, in order to investigate whether mood variations affect the recognition processes differently.

In the current research, the participants were not given any supplementary descriptions of the stimulus and were not asked to rank the stimulus in any way. Since the participants did not expect to interact with the stimulus or were not placed in a position to decide whether to cooperate with it or not, it is questionable to what extent trustworthiness was considered. Future research can thus explore the differences in stimulus recognition when participants are asked to consciously assess the stimulus on its trustworthiness. For example, by asking participants to explicitly assess the stimuli on its trustworthiness and thus making them more involved personally. By creating this personal involvement, the individual differences on (un)trust judgments can also be better assessed.

Conclusion

To conclude, the study showed that although untrustworthy stimulus is better remembered, no conclusions can be made for the trustworthy stimuli. However, when the difference between trustworthy and untrustworthy stimuli was explored, the untrustworthy stimuli had a higher score compared to the trustworthy stimuli which in combination with the high Bayes Factor, provide positive evidence in favour of the existence of a difference between trustworthy and untrustworthy stimuli. Moreover, no effects were found for the explored individual factors (Personality, Faith in General Technology and Trust Stance) and the recognition of trustworthy and untrustworthy stimuli. Thus, the current study found only partial evidence supporting the existence of a cheater detection mechanism when dealing with (un)trust stimuli.

Although the research reflected more the way that brief encounters and first impressions are formed in everyday life, it can be concluded that this passive viewing of the stimulus was not enough to activate participants' assessments of trustworthiness and untrustworthiness. Assessments of (un)trustworthiness are highly subjective and can be affected by many factors, such as memory, emotions, context, personality, mood, appearance etc. With the rapid increase in technology usage, and the latest advantages in Robotics, Artificial Intelligence and Machine Learning, not only more products will be developed, but they will also replace several of the now human activities. Thus, humans in the future will not only need to assess other humans in their trustworthiness and untrustworthiness but also machines. Future research should explore further the encoding, assessment and recognition of (un)trust both towards humans and towards machines, in order to find differences and similarities between them, and potentially create a human persona that will describe individuals who are good at recognizing (un)trustworthy stimuli.

References

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature, 393*(6684), 470-474. doi:10.1038/30982

Alarcon, G. M., Lyons, J. B., Christensen, J. C., Bowers, M. A., Klosterman, S. L., & Capiola, A. (2018). The role of propensity to trust and the five factor model across the trust process. *Journal of Research in Personality, 75*, 69-82. doi:https://doi.org/10.1016/j.jrp.2018.05.006

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11*(7), 280-289. doi:https://doi.org/10.1016/j.tics.2007.05.005

Bell, R., & Buchner, A. (2010). Valence modulates source memory for faces. *Memory and Cognition, 38*(1), 29-41. doi:10.3758/MC.38.1.29

Bell, R., & Buchner, A. (2012). How Adaptive Is Memory for Cheaters? *Current Directions in Psychological Science, 21*(6), 403-408. doi:10.1177/0963721412458525

Borsci, S., Buckle, P., Walne, S., & Salanitri, D. (2018). Trust and Human Factors in the Design of Healthcare Technology. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018): Volume VII: Ergonomics in Design, Design for All, Activity Theories for Work Analysis and Design, Affective Design* (Vol. 824, pp. 207-215). (Advances in Intelligent Systems and Computing; Vol. 824). Springer.

Buchner, A., Bell, R., Mehl, B., & Musch, J. (2009). No enhanced recognition memory, but better source memory for faces of cheaters. *Evolution and Human Behavior, 30*(3), 212-224. doi:https://doi.org/10.1016/j.evolhumbehav.2009.01.004

Bürkner, C. P., & Gabry, J (2020). Bayesian Regression Models using "Stan". (2.13.5). Retreived from (Sept., 29, 2020): https://cran.r-project.org/web/packages/brms/brms.pdf

Chiou, E. K., Schroeder, N. L., & Craig, S. D. (2020). How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers & Education, 146*, 103756. doi:https://doi.org/10.1016/j.compedu.2019.103756

Chiu, A (2019, June 12). Facebook wouldn't delete an altered video of Nancy Pelosi. What about one of Mark Zuckerberg? [Web log post]. Retreived December 15, 2020, from https://www.washingtonpost.com/nation/2019/06/12/mark-zuckerberg-deepfake-facebook-instagram-nancy-pelosi/?noredirect=on

Christine Roy, M., Dewit, O., & Aubert Benoit, A. (2001). The impact of interface usability on trust in Web retailers. *Internet Research, 11*(5), 388-398. doi:10.1108/10662240110410165

Clarke K, Hughes J, Martin D, Rouncefield M, Sommerville I, Voß A, Procter R, Slack R, Hartswood M (2006) Its about time: temporal features of dependability. In: Clarke K, Hardstone G, Rouncefield M, Sommerville I (eds) Trust in technology: a socio-technical perspective, Comput Supported Coop Work, vol 36, Springer, chap 4, pp 105–121. doi: 10.1007/1-4020-4258-2_5

Crone, D. L., Bode, S., Murawski, C., & Laham, S. M. (2018). The Socio-Moral Image Database (SMID): A novel stimulus set for the study of social, moral and affective processes. *PLOS ONE, 13*(1), e0190954 e0190954. doi:10.1371/journal.pone.0190954

Dimoka, A. (2010). What Does the Brain Tell Us About Trust and Distrust? Evidence from a Functional Neuroimaging Study. *MIS Quarterly, 34*, 373-396. doi:10.2307/20721433

Freitag, M., & Bauer, P. C. (2016). Personality traits and the propensity to trust friends and strangers. *The Social Science Journal, 53*(4), 467-476. doi:https://doi.org/10.1016/j.soscij.2015.12.002

Friedman, B., Peter H. Khan, J., & Howe, D. C. (2000). Trust online. *Commun. ACM, 43*(12), 34-40. doi:10.1145/355112.355120

Goldstein, E. B. (2015). *Cognitive psychology: Connecting mind, research and everyday experience*. New york: Cengage learning.

Greenberg, S., Boring, S., Vermeulen, J., & Dostal, J. (2014). *Dark patterns in proxemic interactions: a critical perspective*. Paper presented at the Proceedings of the 2014 conference on Designing interactive systems, Vancouver, BC, Canada. https://doi.org/10.1145/2598510.2598541

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitiveaspects of physiognomy.Journal of Personality and Social Psychology,78, 837-852.

Huvila, I. (2017). Distrust, mistrust, untrust and information practices. *Information Research, 22*.

Johnson, E. J., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology, 45*(1), 20-31. doi:10.1037/0022-3514.45.1.20

Kass, R., & Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association, 90*(430), 773-795. doi:10.2307/2291091

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons, 63*(2), 135-146. doi:https://doi.org/10.1016/j.bushor.2019.11.006

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors, 46*(1), 50-80. doi:10.1518/hfes.46.1.50_30392

Lewicki, R., McAllister, D., & Bies, R. (1998). Trust and Distrust: New Relationships and Realities. *The Academy of Management Review, 23*(3), 438-458. Retrieved September 29, 2020, from http://www.jstor.org/stable/259288

Lewis, J. D., & Weigert, A. (1985). Trust as a Social Reality. *Social Forces, 63*(4), 967-985. doi:10.1093/sf/63.4.967

Liggett, J. C. (1974).The human face. New York: Stein & Day; 1974.
Mealy, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. Ethology and Sociobiology,17, 119-128.

Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology, 25*(2), 115-126. doi:10.1080/01449290500330448

Luhmann, N., Davis, H., Raffan, J., Rooney, K., King, M., & Morgner, C. (2017). *Trust and Power*: Wiley.

Ly, A., Verhagen, J., & Wagenmakers, E.-J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology, 72*, 19-32. doi:https://doi.org/10.1016/j.jmp.2015.06.004

Macmillan, N. A., Creelman, C. D. (2004). Detection theory: A user guide. Psychology press.

Macrae, C. N., & Bodenhausen, G. V. (2000) Social cognition: Thinking categorically about others. In*: Vol. 51. Annual Review of Psychology* (pp. 93-120).

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human automation trust: an integrative review. *Theoretical Issues in Ergonomics Science, 8*(4), 277-301. doi:10.1080/14639220500337708

Marsh, S., & Dibben, M. R. (2005). *Trust, Untrust, Distrust and Mistrust – An Exploration of the Dark(er) Side.* Paper presented at the Trust Management, Berlin, Heidelberg.

Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology, 17*(2), 119-128. doi:https://doi.org/10.1016/0162-3095(95)00131-X

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2), 12-32. https://doi.org/10.1145/1985347.1985353

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research, 13*(3), 334-359. doi:10.1287/isre.13.3.334.81

Montague, E. (2010). Validation of a trust in medical technology instrument. *Applied Ergonomics, 41*(6), 812 821. doi:https://doi.org/10.1016/j.apergo.2010.01.009

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies, 27*(5), 527-539. doi:https://doi.org/10.1016/S0020-7373(87)80013-5

Oda, R. (1997). Biased face recognition in the Prisoner's Dilemma Game. *Evolution and Human Behavior, 18*(5), 309-315. doi:https://doi.org/10.1016/S1090-5138(97)00014-7

Ohman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: a threat advantage with schematic stimuli. *Journal of personality and social psychology, 80*(3), 381–396. https://doi.org/10.1037/0022-3514.80.3.381

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences, 18*(11), 566-570. doi:https://doi.org/10.1016/j.tics.2014.09.007

Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology, 14*(2), 198-202. doi:https://doi.org/10.1016/j.conb.2004.03.015

Roesler, R., & McGaugh, J. L. (2019). Memory Consolidation. In *Reference Module in Neuroscience and Biobehavioral Psychology*: Elsevier.

Rule, N. O., Slepian, M. L., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition, 125*(2), 207-218. doi:https://doi.org/10.1016/j.cognition.2012.06.017

Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*: Prentice Hall Press.

Salanitri, D., Hare, C., Borsci, S., Lawson, G., Sharples, S., & Waterfield, B. (2015). *Relationship Between Trust and Usability in Virtual Environments: An Ongoing Study*, Cham.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science, 300*(5626), 1755-1758. doi:10.1126/science.1082976

Sherwood, L. (2016). *Human physiology: From cells to systems*.

Suzuki, A., & Suga, S. (2010). Enhanced memory for the wolf in sheep's clothing:: Facial trustworthiness modulates face-trait associative memory. *Cognition, 117*(2), 224-229. doi:https://doi.org/10.1016/j.cognition.2010.08.004

Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: learning lip sync from audio. *ACM Trans. Graph., 36*(4), Article 95. doi:10.1145/3072959.3073640

Todorov, A. (2008). *Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors*.

Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology, 25*(3), 395-410. doi:10.1080/02643290802044996

Thatcher, J. B., McKnight, D. H., Baker, E. W., Arsal, R. E., & Roberts, N. H. (2011). The role of trust in postadoption IT exploration: An empirical examination of knowledge management systems. *IEEE Transactions on Engineering Management, 58*(1), 56-70

Tuch, A. N., Presslaber, E. E., Stöcklin, M., Opwis, K., & Bargas-Avila, J. A. (2012). The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. *International Journal of Human-Computer Studies, 70*(11), 794-811. doi:https://doi.org/10.1016/j.ijhcs.2012.06.003

Urcelay, G. P., & Miller, R. R. (2008). 1.05 - Retrieval from Memory. In J. H. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference* (pp. 53-73). Oxford: Academic Press.

Vance, A., Elie-Dit-Cosaque, C., & Straub, D. W. (2008). Examining trust in information technology artifacts: The effects of system quality and culture. *Journal of Management Information Systems, 24*(4), 73-100.

van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108*(3), 796-803. doi:https://doi.org/10.1016/j.cognition.2008.07.002

Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: The sequel. A kernel of truth in predictive cheating detection. *Evolution and Human Behavior, 28*(4), 260-271. doi:10.1016/j.evolhumbehav.2007.04.006

Volonasi, N. (2019). Re-design of an online survey to assess trust before the use of technology. In.

Waldman, A. E. (2020). Cognitive biases, dark patterns, and the 'privacy paradox'. *Current Opinion in*

Psychology, 31, 105-109. doi:https://doi.org/10.1016/j.copsyc.2019.08.025

Wang, W., & Benbasat, I. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems, 6*(3), 72-101

Willis, J., & Todorov, A. (2006). First Impressions:Making Up Your Mind After a 100-Ms Exposure to a Face. *Psychological Science, 17*(7), 592-598. doi:10.1111/j.14679280.2006.01750.x

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5*(3), 277 283. doi:10.1038/nn816

Xu, J., Le, K., Deitermann, A., & Montague, E. (2014). How different types of users develop trust in technology: A qualitative analysis of the antecedents of active and passive user trust in a shared technology. *Applied Ergonomics, 45*(6), 1495-1503. doi:https://doi.org/10.1016/j.apergo.2014.04.012

Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You can judge a book by its cover: Evidence that cheaters may look different from cooperators. *Evolution and Human Behavior, 24*(4), 290-301. doi:https://doi.org/10.1016/S1090-5138(03)00035-7

Yu, M., Saleem, M., & Gonzalez, C. (2014). Developing trust: First impressions and experience. *Journal of Economic Psychology, 43*, 16-29. doi:https://doi.org/10.1016/j.joep.2014.04.004

Zelenski, J. (2007). The role of personality in emotion, judgment, and decision making. In (pp. 117-132)

Appendix A - Demographic Characteristics

1. What is your age?

   ____

2. What is your sex?
   a. Male
   b. Female
   c. Other
3. What is your nationality? (e.g. Dutch, English, German etc.)

   ____

4. What is the highest level of school you have completed or the highest degree you have received?
   a. Less than high school degree
   b. High school graduate (high school diploma or equivalent including GED)
   c. Some college but no degree
   d. Associate degree in college (2-year)
   e. Bachelor's degree in college (4-year)
   f. Master's degree
   g. Doctoral degree
   h. Professional degree (JD, MD)
5. What is your occupation/ What do you study?

   ____

Appendix B - Attitude towards Technology/ Dependency

1. I feel it is important to be able to find any information whenever I want online
2. I feel it is important to be able to access the Internet any time I want
3. I think it is important to keep up with the latest trends in technology
4. I get anxious when I don’t have my cell phone
5. I get anxious when I don’t have the Internet available to me
6. I am dependent on my technology
7. Technology will provide solutions to many of our problems
8. With technology anything is possible
9. I feel that I get more accomplished because of technology
10. New technology makes people waste too much time
11. New technology makes life more complicated
12. New technology makes people more isolated

Appendix C - IPIP Big 5 Personality Traits Questionnaire

Instructions: Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence.

Indicate for each statement whether it is 1. Very Inaccurate, 2. Moderately Inaccurate, 3. Neither Accurate Nor Inaccurate, 4. Moderately Accurate, or 5. Very Accurate as a description of you.

Very Inaccurate	Moderately Inaccurate	Neither Accurate Nor Inaccurate	Moderately Accurate	Very Accurate

1. Am the life of the party
2. Feel little concern for others
3. Am always prepared
4. Get stressed out easily
5. Have a rich vocabulary
6. Don't talk a lot
7. Am interested in people
8. Leave my belongings around
9. Am relaxed most of the time
10. Have difficulty understanding abstract ideas
11. Feel comfortable around people
12. Insult people
13. Pay attention to details
14. Worry about things
15. Have a vivid imagination
16. Keep in the background
17. Sympathize with others' feelings
18. Make a mess of things
19. Seldom feel blue
20. Am not interested in abstract ideas
21. Start conversations
22. Am not interested in other people's problems
23. Get chores done right away
24. Am easily disturbed.
25. Have excellent ideas.

26. Have little to say
27. Have a soft heart
28. Often forget to put things back in their proper place
29. Get upset easily
30. Do not have a good imagination
31. Talk to a lot of different people at parties
32. Am not really interested in others
33. Like order
34. Change my mood a lot
35. Am quick to understand things
36. Don’t like to draw attention to myself
37. Take time out for others
38. Shirk my duties
39. Have frequent mood swings
40. Use difficult words
41. Don't mind being the center of attention
42. Feel others' emotions
43. Follow a schedule
44. Get irritated easily
45. Spend time reflecting on things
46. Am quiet around strangers
47. Make people feel at ease
48. Am exacting in my work
49. Often feel blue
50. Am full of ideas

Appendix D - Trust in Technology

Faith in General Technology (Adapted from McKnight et al. 2002):

1. I believe that most technologies are effective at what they are designed to do.
2. A large majority of technologies are excellent.
3. Most technologies have the features needed for their domain.
4. I think most technologies enable me to do what I need to do.

Trusting Stance—General Technology (Adapted from McKnight et al. 2002):

1. My typical approach is to trust new technologies until they prove to me that I shouldn't trust them.
2. I usually trust a technology until it gives me a reason not to trust it.
3. I generally give a technology the benefit of the doubt when I first use it.

Appendix E - Informed Consent

Information Sheet and consent form

Introduction

The current study aims at exploring the concept of image recall. The researcher of the study is Niki Volonasi (n.volonasi@student.utwente.nl) supervised by Dr Simone Borsci (s.borsci@utwente.nl) from the University of Twente.

Description of Task

The overall goal of the study will be described to you at the end because we do not want to affect your answer in any way. However, we may anticipate that we will show you some pictures with faces of people, some scenes and some technological products multiple times. And you will be requested to answer some questions about the images you have seen, specifically which pictures you will remember better.

The experiment is divided into two phases through which these images will be presented to you. In between these two phases, a 30 minutes break will take place where you will be asked to watch a TEDx talk and answer some follow-up questions.

Duration and procedure

The duration of the study will be approximately 60 minutes.

While the experiment takes place, your eye movement activity will be recorded using eye-tracking technology.

Risks of participants

The pictures we are going to show you are not meant to provoke any reaction, and there is no material which can be considered disgusting or immoral, therefore there are no risks involved in this experiment

Rights of participants

This study is not aiming at assessing you in any way. There are no right or wrong answers In case of confusion or questions, please ask the researcher.

You have the right to quit the experiment at any time and in doing so, your data will also automatically be deleted from the dataset.

Your identity will remain confidential and anonymous and your data (eye-tracking recordings) will be securely stored in an encrypted repository.

Contacts

If you have any questions concerning your rights as a participant to this study, you may contact Dr Simone Borsci (s.borsci@utwente.nl)

Talking part in the study

	Yes	No
I have read and understood the study information I have been able to ask questions about the study and my questions have been answered to my satisfaction.		
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.		
I understand that taking part in the study involves a survey questionnaire to be completed		
I understand that taking part in the study involves the recording of eye-tracking activity		

Use of information in the study

	Yes	No
I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team.		
I agree to record my eye-tracking activity		

Signatures

------------------------------ ---------------------------------- ---------------

[Name of participant] [Signature] [Date]

Appendix F – Priors in R

```
priors_general <- c (set_prior("normal(1.51,0.39)", class = "b", coef = "DprimeTrust"),
                     set_prior("normal(1.51,0.39)", class = "b", coef = "DprimeUntrust"))
```

Appendix G – Models

Model to measure the effect of General Trustworthy and Untrustworthy stimuli on DprimePRE (memory ability):

```
MemoryAbility_General <- brm(formula = DprimePRe ~ DprimeTrust + DprimeUntrust,
          data    = D_1,
          seed    = 123,
          priors  = priors_general_trustun,
          save_all_pars = TRUE)

summary(MemoryAbility_ General)
```

Model to measure the effect of the contrast between DprimeUntrust and DprimtTrust on DprimePRE:

```
Trust_Untrust_Difference <- brm(formula = DprimeTrust ~ DprimeUntrust,
          data = D_1,
          seed = 123,
          save_all_pars = TRUE)
```

Appendix H – Memory Models

General Stimuli

```
MemoryAbility_General_Prior <- brm(formula = DprimePRe ~ DprimeUntrust +
DprimeTrust,
             data    = D_1,
             prior = priors_genera_trustun,
             seed    = 123,
             save_all_pars = TRUE)

posterior_summary(MemoryAbility_General_Prior)

##                    Estimate  Est.Error        Q2.5       Q97.5
## b_Intercept       0.9723361 0.12701633   0.71271633   1.2228490
## b_DprimeUntrust   0.2544038 0.09597873   0.06303166   0.4410292
## b_DprimeTrust     0.1486201 0.09463068  -0.04467990   0.3354000
## sigma             0.3683305 0.02952585   0.31603919   0.4315101
## Intercept         1.5086337 0.04104203   1.42954129   1.5915827
## lp__            -52.7912243 1.47034264 -56.42477032 -50.9782796
```

Bayes Model for the Untrustworthy Model

```
MemoryAbility_Untr <- brm(formula = DprimePRe ~ DprimeUntrust,
             data    = D_1,
             seed    = 123,
             save_all_pars = TRUE)

bayes_factor(MemoryAbility_Untr, MemoryAbility_Null)

## Estimated Bayes factor in favor of MemoryAbility_Untr over
MemoryAbility_Null: 82.53338

bayes_factor(MemoryAbility_Trust, MemoryAbility_Null)

## Estimated Bayes factor in favor of MemoryAbility_Trust over
MemoryAbility_Null: 12.31064
```

Trust & Untrust Difference

```r
Trust_Untrust_Difference <- brm(formula = DprimeTrust ~ DprimeUntrust,
             data    = D_1,
             seed    = 123,
             save_all_pars = TRUE)

posterior_summary(Trust_Untrust_Difference)

##                     Estimate  Est.Error        Q2.5       Q97.5
## b_Intercept        0.5155553 0.13226894   0.2505353   0.7739580
## b_DprimeUntrust    0.5860522 0.09216693   0.4092724   0.7709389
## sigma              0.4234593 0.03284866   0.3645482   0.4921060
## Intercept          1.3045317 0.04618665   1.2151157   1.3982297
## lp__             -52.9273461 1.20403821 -55.9827409 -51.5407013

bayes_factor(Trust_Untrust_Difference,Null_Trust_Untrust)

## Estimated Bayes factor in favor of Trust_Untrust_Difference over
Null_Trust_Untrust: 14.35216
```

Photos of Faces and Memory

```r
MemoryAbility_Faces_Priors <- brm(formula = DprimePRe ~ DprimeFU +
DprimeFT,
             data    = D_1,
             prior = priors_face,
             save_all_pars = TRUE,
             seed    = 123)

posterior_summary(MemoryAbility_Faces_Priors)

##                  Estimate  Est.Error        Q2.5       Q97.5
## b_Intercept   1.32830987 0.08712977   1.1587657   1.4943045
## b_DprimeFU    0.06096124 0.09223920  -0.1177941   0.2439879
## b_DprimeFT    0.14857772 0.08604970  -0.0192943   0.3231542
## sigma         0.39188082 0.03174826   0.3367768   0.4602007
## Intercept     1.50799357 0.04321469   1.4212424   1.5919239
## lp__        -59.51995394 1.49489876 -63.2447230 -57.6948841

bayes_factor(MemoryAbility_Trust_Face, MemoryAbility_Null)
```

```
## Estimated Bayes factor in favor of MemoryAbility_Trust_Face over
MemoryAbility_Null: 1.01052

bayes_factor(MemoryAbility_Untrust_Face, MemoryAbility_Null)

## Estimated Bayes factor in favor of MemoryAbility_Untrust_Face over
MemoryAbility_Null: 0.38858
```

Photos of Scenes

```
MemoryAbility_Scenes_Prior <- brm(formula = DprimePRe ~ DprimeSU +
DprimeST  ,
             data    = D_1,
             prior = priors_scene,
             save_all_pars = TRUE,
             seed    = 123)

posterior_summary(MemoryAbility_Scenes_Prior)

##                 Estimate  Est.Error        Q2.5       Q97.5
## b_Intercept   1.0061087 0.12413186   0.75737142   1.2475981
## b_DprimeSU    0.1950348 0.07695141   0.04485307   0.3428508
## b_DprimeST    0.2076039 0.08909552   0.03783448   0.3867086
## sigma         0.3716903 0.02984530   0.31801609   0.4344188
## Intercept     1.5085951 0.04033596   1.43008134   1.5899813
## lp__        -53.4764238 1.45064564 -57.21739008 -51.6811095
```

```
bayes_factor(MemoryAbility_Trust_Scene, MemoryAbility_Null)

## Estimated Bayes factor in favor of MemoryAbility_Trust_Scene over
MemoryAbility_Null: 8.15509

bayes_factor(MemoryAbility_Untrust_Scene, MemoryAbility_Null)

## Estimated Bayes factor in favor of MemoryAbility_Untrust_Scene over
MemoryAbility_Null: 12.84230
```

Photos of Devices

```
MemoryAbility_Devices_Prior <- brm(formula = DprimePRe ~ DprimeDU +
DprimeDT  ,
            data    = D_1,
            prior = priors_device,
            save_all_pars = TRUE,
            seed    = 123)

posterior_summary(MemoryAbility_Devices_Prior)

##                  Estimate  Est.Error         Q2.5       Q97.5
## b_Intercept   1.1480663 0.11876087   0.90687033   1.3770307
## b_DprimeDU    0.1967837 0.07959797   0.04395323   0.3544193
## b_DprimeDT    0.1234247 0.08503939  -0.04273530   0.2944145
## sigma         0.3873505 0.03191679   0.33243101   0.4563848
## Intercept     1.5092754 0.04108909   1.43010077   1.5909359
## lp__        -57.4673868 1.40608620 -60.98930440 -55.7326073

bayes_factor(MemoryAbility_Trust_Device, MemoryAbility_Null)

## Estimated Bayes factor in favor of MemoryAbility_Trust_Device over
MemoryAbility_Null: 0.56896

bayes_factor(MemoryAbility_Untrust_Device, MemoryAbility_Null)

## Estimated Bayes factor in favor of MemoryAbility_Untrust_Device over
MemoryAbility_Null: 2.24836
```

Appendix I – Faith in Technology & Trust Stance Models

Faith in Technology

```
FaithinTech_Device <- brm(formula = faith_percentage ~ DprimeDU + DprimeDT
,
             data                                          =          D_1,
             save_all_pars                    =                     TRUE,
             seed    = 123)

posterior_summary(FaithinTech_Device)

##                        Estimate  Est.Error        Q2.5        Q97.5
##   b_Intercept       29.7257503   1.2140330      27.354059      32.022719
##   b_DprimeDU        -0.4908056   0.8350259      -2.094526       1.177338
##   b_DprimeDT        -0.2144533   0.8521504      -1.887457       1.531601
##  sigma                3.8480542  0.3015787       3.308980       4.491860
##   Intercept          28.9288330   0.4220663      28.121436      29.791217
## lp__        -236.2975292 1.4536741 -239.995526 -234.526251

bayes_factor(FaithinTech_Trust, FaithinTech_Null)

##   Estimated   Bayes   factor   in   favor   of   FaithinTech_Trust   over
FaithinTech_Null: 2.30605

bayes_factor(FaithinTech_UnTrust, FaithinTech_Null)

##   Estimated   Bayes   factor   in   favor   of   FaithinTech_UnTrust   over
FaithinTech_Null: 2.50198
```

Trust Stance

```
TrustStance_Device <- brm(formula =  trust_percentage ~ DprimeDU + DprimeDT
,
             data                                          =          D_1,
             save_all_pars                    =                     TRUE,
             seed    = 123)
```

```
posterior_summary(TrustStance_Device)

##                        Estimate  Est.Error        Q2.5        Q97.5
##   b_Intercept      18.9549517   1.6076211      15.789084      22.030786
##   b_DprimeDU         0.3031219   1.1060314      -1.954517       2.482343
##   b_DprimeDT        -0.8544098   1.1461164      -3.190460       1.395987
##  sigma                5.0203890  0.3923775       4.316248       5.867439
##   Intercept         18.3482234   0.5563063      17.267847      19.456281
## lp__        -258.7716378 1.4766919 -262.376418 -256.941699

bayes_factor(TrustStance_Trust, Truststance_Null)

## Estimated Bayes factor in favor of TrustStance_Trust over
Truststance_Null: 3.47230

bayes_factor(TrustStance_Untrust, Truststance_Null)

## Estimated Bayes factor in favor of TrustStance_Untrust over
Truststance_Null: 2.55964
```

Appendix J – Personality Trait Models

Extraversion

```
ExtraversionPercentage_General <- brm(formula = extraversion_percentage ~ DprimeUntrust + DprimeTrust,
            data = D_1,
            save_all_pars = TRUE,
            seed    = 123)

posterior_summary(ExtraversionPercentage_General)

##                          Estimate  Est.Error         Q2.5         Q97.5
##  b_Intercept            73.2517345  3.7742139    65.822332     80.647542
##  b_DprimeUntrust       -2.5877548  2.9769048     -8.457391       3.267125
##  b_DprimeTrust         -0.7573226  2.9423259     -6.307591       5.078542
##  sigma                  10.8624445  0.8791605      9.322019     12.764165
##  Intercept              68.7803827  1.1863960     66.481000     71.105004
## lp__          -323.5897561 1.5196878 -327.523618 -321.719574

bayes_factor(Extraversion_Trust, Extraversion_Null)

## Estimated Bayes factor in favor of Extraversion_Trust over Extraversion_Null: 9.12554

bayes_factor(Extraversion_Untrust, Extraversion_Null)

## Estimated Bayes factor in favor of Extraversion_Untrust over Extraversion_Null: 13.77729
```

Agreeableness

```
Agreeableness_General <- brm(formula = agreeableness_percentage ~ DprimeUntrust + DprimeTrust,
            data = D_1,
            save_all_pars = TRUE,
            seed    = 123)

posterior_summary(Agreeableness_General)
```

```
##                         Estimate  Est.Error        Q2.5        Q97.5
## b_Intercept            81.053673  4.0597665   73.051453    88.981505
## b_DprimeUntrust        -4.131993  3.1720303  -10.210737     2.265386
## b_DprimeTrust           1.429773  3.0733211   -4.637979     7.412842
## sigma                  11.723674  0.9330494   10.067815    13.704011
## Intercept              77.355419  1.2861979   74.873954    79.898953
## lp__                 -330.032584 1.4186389 -333.717748 -328.222087
```

```
bayes_factor(Agreeableness_Trust, Agreeableness_Null)
```

```
## Estimated Bayes factor in favor of Agreeableness_Trust over Agreeableness_Null: 6.80049
```

```
bayes_factor(Agreeableness_Untrust, Agreeableness_Null)
```

```
## Estimated Bayes factor in favor of Agreeableness_Untrust over Agreeableness_Null: 15.39383
```

Conscientiousness

```
Conscientiousness_General <- brm(formula = conscientiousness_percentage ~ DprimeUntrust + DprimeTrust,
             data = D_1,
             save_all_pars = TRUE,
             seed    = 123)
```

```
posterior_summary(Conscientiousness_General)
```

```
##                         Estimate  Est.Error        Q2.5        Q97.5
## b_Intercept            72.344589   4.616526   62.951520    81.230063
## b_DprimeUntrust        -1.195886   3.668782   -8.334631     5.829530
## b_DprimeTrust          -1.640840   3.631566   -8.519529     5.614417
## sigma                  13.340192   1.069143   11.457753    15.570875
## Intercept              68.594916   1.444472   65.760927    71.416696
## lp__                 -340.899633   1.469970 -344.683566 -339.073266
```

```
bayes_factor(Conscientiousness_Trust, Conscientiousness_Null)
```

```
## Estimated Bayes factor in favor of Conscientiousness_Trust over Conscientiousness_Null: 9.96314
```

```r
bayes_factor(Conscientiousness_Untrust, Conscientiousness_Null)

## Estimated Bayes factor in favor of Conscientiousness_Untrust over Conscientiousness_Null: 9.53120
```

Emotional Stability

```r
EmotionalStability_General <- brm(formula = emotional_stability_percentage ~ DprimeUntrust + DprimeTrust,
          data = D_1,
          save_all_pars = TRUE,
          seed    = 123)

posterior_summary(EmotionalStability_General)

##                    Estimate  Est.Error         Q2.5        Q97.5
## b_Intercept       61.219785   5.064004   50.9517440    71.104176
## b_DprimeUntrust   -5.389491   4.022691  -13.2276033     2.369605
## b_DprimeTrust      7.267871   3.933726   -0.5496445    14.985162
## sigma             14.811188   1.180485   12.6261769    17.320622
## Intercept         63.441672   1.577393   60.4211726    66.536474
## lp__            -350.094990   1.467275 -353.6459381 -348.308846

bayes_factor(EmotionalStability_Trust , EmotionalStability_Null)

## Estimated Bayes factor in favor of EmotionalStability_Trust over EmotionalStability_Null: 19.60053

bayes_factor(EmotionalStability_Untrust , EmotionalStability_Null)

## Estimated Bayes factor in favor of EmotionalStability_Untrust over EmotionalStability_Null: 8.76012
```

Intellect Imagination

```r
IntellectImagination_General <- brm(formula = intellect_imagination_percentage ~ DprimeUntrust + DprimeTrust,
          data = D_1,
          save_all_pars = TRUE,
          seed    = 123)
```

```
posterior_summary(IntellectImagination_General)

##                              Estimate  Est.Error        Q2.5        Q97.5
##  b_Intercept               76.428813   3.592515   69.215568    83.673160
##  b_DprimeUntrust          -2.917159   2.779643   -8.427316     2.370465
##  b_DprimeTrust              2.052420   2.737372   -3.352603     7.680408
##  sigma                      10.456930   0.836785    8.947944    12.181224
##  Intercept                 75.177989   1.130996   72.956866    77.404834
## lp__          -320.270503  1.474509 -324.124540 -318.490668

bayes_factor(IntellectImagination_Trust , IntellectImagination_Null)

## Estimated Bayes factor in favor of IntellectImagination_Trust over
IntellectImagination_Null: 5.64394

bayes_factor(IntellectImagination_Untrust , IntellectImagination_Null)

## Estimated Bayes factor in favor of IntellectImagination_Untrust over
IntellectImagination_Null: 7.64834
```