Friedrich-Alexander-Universität Erlangen-Nürnberg

FAU
Friedrich-Alexander-Universität
Erlangen-Nürnberg

Master Thesis

How does conversational privacy affect user perceptions and behavior?

submitted by

Anna Leschanowsky

submitted

March 31, 2022

Supervisor / Advisor

Dr. Birgit Popp
Prof. Dr. Nils Peters

Reviewers

Prof. Dr. Nils Peters

AUDIO LABS

International Audio Laboratories Erlangen
A joint institution of the
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and
the Fraunhofer-Institut für Integrierte Schaltungen IIS.

Fraunhofer
IIS

Erklärung

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Erlangen, March 31, 2022

Anna Leschanowsky

Acknowledgements

I would like to thank my supervisors, Dr. Birgit Popp from the Fraunhofer IIS and Prof. Dr. Nils Peters from the International Audio Laboratories Erlangen for their support and guidance throughout the thesis. This work would not have been possible without the opportunities you provided. I am grateful for the knowledge and perspectives I've gained from working with you for the last months. I am especially thankful for your suggestions, constructive feedback and fruitful discussions. Overall, working on the project was fun!

Abstract

One of the key goals of the European General Data Protection Regulation (GDPR) and several other international data protection laws is to strengthen the control individuals have over their personal data. This becomes increasingly important with numerous digital devices entering peoples' homes and the associated large scale data aggregation. Security and privacy violations in Conversational User Interfaces (CUI) like Alexa raised global awareness for the topic and users reported increasing privacy concerns when engaging with those systems. Based on the outcome of a preceding qualitative study on peoples' privacy perceptions we argue that disclosing private information is subject to cognitive biases and often guided by heuristics rather than rational assessment. The Dual-Process Model of Cognition suggests that rational assessment can be triggered by embracing uncertainty in a controlled way. We apply those psychological insights to Conversational User Interfaces, to design them in a way that promotes rational and informed decision making about user data.

Contents

Acronyms

AI Artificial Intelligence. 26, 27, 36, 60, 61, 63

AIC Akaike Information Criterion. 67, 68, 72, 77–80, 84, 90

ANOVA Analysis of Variance. 47–49, 65, 67, 70, 74, 76–80, 84–86, 91

API Application Programming Interface. 24

ART Aligned Rank Transform. 47

CAGR Compound Annual Growth Rate. 6

CAI Conversational Artificial Intelligence. 5, 6

CBL Chatbot Language. 33, 34, 36, 42

CFA Confirmatory Factor Analysis. 40

CUI Conversational User Interface. 5, 6, 9, 10, 25, 27, 28, 31–33, 35, 43, 57, 93, 96–101, 157

CVSCALE Hofstede's five dimensions of cultural values measured on an individual level. 41

ECHR European Convention of Human Rights. 11, 12

EDPB European Data Protection Board. 12, 158

EFA Exploratory Factor Analysis. 51, 53

EU European Union. 6, 11–13

GDPR General Data Protection Regulation. 6, 9, 12–14, 23, 43, 97–99, 157, 158

HCI Human-Computer Interaction. 37, 42

HMI Human-Machine Interaction. 5, 31, 32

ICCPR International Convenant on Civil and Political Rights. 11

IoT Internet of Things. 6, 16–18

IPA Intelligent Personal Assistant. 16, 17

ITU-T International Telecommunication Union Telecommunication Standardization Sector. 42

IUIPC	Internet Users' Information Privacy Concerns. 40, 41, 52, 54, 58, 124, 145
IVA	Intelligent Virtual Assistant. 5
KMO	Kaiser-Meyer-Okin. 53, 72
MTurk	Amazon Mechanical Turk. 33, 41, 44
PANAS	Positive and Negative Affect Schedule. 37, 158
PriBot	Conversational Privacy Bot. 13, 14
PRU	Physicians' Reactions to Uncertainty. 38, 39, 45, 48, 52, 53, 58, 120
UDHR	Universal Declaration of Human Rights. 11
UK	United Kingdom. 16
UN	United Nations. 11
US	United States. 10, 16, 41
VUI	Voice User Interface. 5, 13, 14
VVA	Virtual Voice Assistant. 13

Chapter 1

Introduction

Since the launch of Apple's Siri in 2011 and Amazon's Alexa in 2014, voice assistants have seen tremendous growth in market share and adoption. Nowadays, they are integrated into our daily lifes by being implemented not only in mobile phones and smart speakers but also in cars, wearable devices and other home appliances. Voice assistants provide hands-free usage and allow for personalized experiences. In general, they can be seen as a form of *Conversational Artificial Intelligence (CAI)*. CAI uses automatic speech recognition, natural language understanding and machine learning to enable effective *Human-Machine Interaction (HMI)* by using natural language and dialogues [Cognigy, 2022]. While the aspect of conversation and the ability of chit-chat is important when talking about CAI, voice assistants may also allow users to carry out specific tasks or state commands that do not need further conversation e.g. "Turn off the lights." [Ottomatias Peura, 2020]. Whenever human-machine communication is carried out, a digital interface is necessary to enable the interaction e.g. voice assistants make use of *Voice User Interfaces (VUIs)*. However, *Conversational User Interfaces (CUIs)* do not only encompass voice or text-based communication interfaces but rather refer to a general design principle [AJ Burt, 2022]. Whenever direct human interaction is preferable, CUIs can be applied e.g. when recording field data for a researcher. While some refer only to text-based and task-oriented CUIs as chatbots [Cognigy, 2022], others use the term to describe a super category encompassing voice assistants, conversational text-based systems as well as purely task-oriented systems Adamopoulou and Moussiades [2020]. With advances in natural language processing and understanding, the market is moving from "process-centric conversational AI to customer-centric experience". *Intelligent Virtual Assistants (IVAs)* are capable of more human-like conversations and can be seen as a successor of traditional chatbots [Sonrat Priyanka, 2020]. Gartner Research distinguishes among chatbots and virtual assistants depending on their level of sophistication. The more complex and contextual the system is, the more it can be seen as a virtual assistant [Gartner Research, 2020]. But no matter how we refer to systems enabling seamless and human-like conversations with a machine, their market share and adoption is forcasted to rise drastically within the next years. The global Conversational Artificial Intelligence (CAI) market is predicted to reach \$32.62 billion by 2030, a 20% *Compound Annual Growth Rate (CAGR)* from 2021 to 2030 [Allied Market Research, 2021]. Moreover, adoption of voice assistants is forcasted to double with 4.2 billion devices being used in 2020 to 8.4 billion devices to be in use worldwide by 2024 [Juniper Research, 2020]. Similarly, chatbot adoption is expected to grow rapidly across all industries as they can significantly save business costs and time for both, businesses and consumers [Brain Code for Equity, 2021]. In the course of this thesis, we will mainly use CUIs as referring to both text- and voice-based applications with conversational character. CUIs and their integration into our daily life come with opportunities and benefits. However, in recent years privacy concerns have been raised and the introduction of data protection regulations worldwide such as the *General Data Protection Regulation (GDPR)* in the European Union (EU), have led to increased awareness among users regarding collection, processing and distribution of their data [der Sloot and De Groot, 2018, European Commission, 2016]. Although people might be aware of threats to their privacy due to data collection and processing, privacy-preserving actions can be time-consuming and might require technical knowledge leading to a lack of adoption [Leschanowsky et al., 2021]. Especially when using CUIs, changing privacy settings might not be straightforward. In many cases, it is not possible to use the same modality i.e. text or voice, to express privacy-related requests. Here, the concept of *Conversational Privacy* can help to make privacy settings and policies easily accessible and usable by presenting them in natural language Harkous et al. [2016]. Nevertheless, even with accessible privacy-preserving strategies in place, it is questionable if users engage in privacy self-management when using CUIs. Previous research on mobile applications, voice assistants and *Internet of Things (IoT)* devices has found that people are likely to express privacy concerns but not act upon them which is known as the *Privacy Paradox* [Williams et al., 2017, Konrad et al., 2020]. One reason for that could be that people decide intuitively upon disclosing personal information rather than thinking thoroughly about its implications [Leschanowsky et al., 2021]. Therefore, we argue that additionally to presenting privacy-related information via voice or text, strategies are necessary which interrupt intuitive decision-making and promote rational evaluation of risks and benefits. We carry out a literature research to find suitable debiasing strategies known from privacy research and other research disciplines. We provide theoretical background on legal, behavioural and cognitive concepts used in this thesis and show the results of the literature research in Chapter 2. Moreover, we evaluate cognitive forcing strategies known as a debiasing strategy in the context of CUIs and investigate its impact on user behaviour, privacy perceptions and usability aspects. Furthermore, we are interested whether those strategies can indeed interrupt intuitive decision-making by building on *Dual-Process Theory*. In Chapter 3, we describe the experimental setup and the results of a pilot and main study. We then discuss our results, their implications and future research directions in Chapter 4. Lastly, we summarize and conclude the thesis in Chapter 5.

Chapter 2

Theoretical Background

In this chapter, we want to introduce the fundamental theoretical background, foundations and concepts from the legal and cognitive science field that are being used in this thesis. In particular, we focus on multidisciplinary discourses on privacy and the evolution from the right to privacy to the General Data Protection Regulation (GDPR) in Section 2.1. The idea of Conversational Privacy, one way to convey complex legal information in CUIs, is put forward in Section 2.2. We then have a look at the Privacy Paradox in Section 2.3 and give a possible explanation based on the concept of Privacy Calculus in Section 2.4. Having investigated the model of a rational agent for privacy, we give an overview of theories incorporating the idea of biased risk assessment. We then present Kahneman's theory of slow and fast thinking in Section 2.5. Furthermore, in Section 2.6 we present an overview of debiasing strategies used in privacy research and other research disciplines. We put an emphasis on cognitive forcing strategies as we are going to investigate their impact on user perception and behaviour in the context of CUIs in the remaining parts of this thesis.

2.1 From the Right to Privacy and Informational Privacy to the GDPR

The origins of our current understanding of privacy are almost impossible to trace back to one single discourse and discipline. Masur provides an extensive overview of different theories on privacy. It becomes clear that multiple disciplines such as philosophy, law and later psychology, sociology and computer sciences have all contributed to our modern idea of privacy. Those discourses while often initiated around the same time did evolve independently from one another. Some of the approaches from the philosophical, legal and socio-technical field had an influence on privacy law but even more on the notion of privacy itself [Masur, 2019]. Due to the heterogeneity and multi-nationality of the discourses, it is not surprising that an exhaustive summary of privacy definitions and legal aspects is beyond the scope of this thesis. In this section, we will only highlight concepts and definitions of privacy that are helpful to understand the current legal state in Europe and its implications for the development of CUIs. We will first focus on the legal and philosophical foundation and development of the notion of privacy before having a look at societal-focused and socio-technical perspectives on privacy. Lastly, we will put forward an operational perspective and discuss the difference between privacy and data protection.

The publication of "The Right to Privacy" by Warren and Brandeis lays the foundation of privacy law in the United States (US) and of the conception of privacy as nonintrusion [Masur, 2019, Tavani, 2007]. They describe privacy in terms of "being let alone" [Warren and Brandeis, 1890]. However, one problem that comes with this definition is the confusion of privacy with liberty which are closely connected concepts yet distinguishable from one another [Tavani, 2007]. Liberty on one side allows people to have unpopular opinions in the first place while in a condition of privacy they can reveal those ideas to a certain group of people or other individuals without making it publicly known [Tavani, 2007]. Another often-cited description that avoids the confusion of privacy with liberty, stems from Westin. To him privacy is the "voluntary and temporary withdrawal of a person from the general society through physical [means] in a state of solitude" [Westin, 1967]. However, as Tavani notes, this formulation does not come without problems as it closely connects privacy to solitude and suggests that the level of privacy increases the more one is secluded from others. Further, he questions whether privacy can only be enjoyed when being alone. Nevertheless, from the descriptions given, we can conclude that the concepts of privacy were initially connected to limited accessibility. Similarly, privacy concerns were related to the intrusion into people's personal space [Tavani, 2007].

With computers made available to everyone, the discourse developed further by introducing the notions of choice and control [Masur, 2019]. Yet, the concept of accessibility did not diminish fully and therefore recent theories of privacy include both, access to as well as control over personal information [Tavani, 2007]. Prominent representatives of this new stream of *Informational Privacy* are e.g. Arthur Miller, Alan F. Westin and Helen Nissenbaum [Masur, 2019]. Importantly, those new conceptions of privacy are distinguishable from the definitions of liberty and solitude. Again one of the most influential definitions in the field comes from Westin who further refers to privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how and to what extent information about them is communicated to other" [Westin, 1967]. Additionally, Miller stated that "privacy is the individual's ability to control the circulation of information relating to him – a power that often is essential to maintaining social relationships and personal freedom" [Miller, 1971]. Although these definitions highlight the importance of control over one's personal information, they are not clear about the level of control and which kinds of personal information can be controlled after all [Tavani, 2007]. Regarding the kind of personal information that can be controlled by an individual, control theorists usually differentiate between sensitive or confidential data such as financial or medical records and public personal information [Masur, 2019, Tavani, 2007]. Here, public personal information refers to information about where a person works, lives, shops and so on [Tavani, 2007]. One can imagine that for someone not living in total isolation, control over certain public personal information is and has always been difficult.

Given the concept of Informational Privacy and control theory, privacy concerns were now associated with the flow of information rather than with intrusion [Tavani, 2007]. This was especially highlighted by Nissenbaum, in her work on *Contextual Integrity.* To her, privacy is directly related to an appropriate flow of information dependent on the social context. She argued that while privacy needs are individual to each person, they can be systematically related to social contexts. Therefore, the appropriateness of information flow will also vary depending on the social setting and the context. This may lead to the phenomenon that a privacy violation is specific to one social context, e.g. in a workplace, and is not seen as a violation at all in a different social setting, e.g. when being together with a group of friends. Contexts are described by Nissenbaum as "structured social settings characterized by canonical activities, roles, relationships, power structures, norms (or rules), and internal values (goals, ends, purpose)" [Nissenbaum, 2010]. While we will not go deeper into the conceptual framework of Contextual Integrity, it is useful to remember the importance of context, purpose and appropriateness when studying the sharing of information with a technical system.

So far we have given an overview of different conceptions of privacy and discourses around privacy in multiple disciplines. We will now focus on concrete legal instruments applicable to privacy. On an international level, the *Universal Declaration of Human Rights (UDHR)* from 1948 states in Article 12 that "no one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attack." [United Nations, 1948]. While the UDHR is not enforceable, in 1976 Article 17 of the *International Convenant on Civil and Political Rights (ICCPR)* entered into force which states in a more detailed manner what was written in the UDHR [United Nations (General Assembly), 1966]. The ICCPR was ratified by 173 states worldwide and signed but not ratified by six additional states among others by China and Cuba [United Nations (General Assembly), 2022]. Importantly, the *United Nations (UN)* General Assembly reaffirmed in its resolution in 2013 that the right to privacy is applicable in today's age of digitization and that the possible violation and abuse of human rights by modern technologies is increasingly concerning [Speaker Identifitcation Integrated Project (SIIP), 2018]. Moreover, mentioning Article 17 ICCPR all states are meant to respect and protect the right to privacy in all contexts, also in the context of digital communications and to take steps to tackle and prevent possible violations [Speaker Identifitcation Integrated Project (SIIP), 2018].

In the *EU*, a legal definition is missing for the right to privacy [Nautsch et al., 2019]. Nevertheless, at the level of the *Council of Europe* (47 European states), Article 8 of the *European Convention* *of Human Rights (ECHR)* explicitly addresses the human right to respect for private and family life, home and correspondence [Council of Europe, 1950]. More importantly to mention however, is the *Convention 108* as the first international legal instrument regarding the protection of personal data [Council of Europe, 1981]. In addition to the Council of Europe member states, some non-members became involved in the process and complied with the binding treaty [Council of Europe, 2021]. Convention 108 is specifically important as it recognizes the protection of personal data as a separate right and applies to the processing of such data by private and public actors [Council of Europe, 1981]. It is substantial to understand the difference between data protection regulation and privacy regulation. As seen above, privacy is not restricted to data exchange but can be subject to violations of the physical personal space. On the other hand, data regulation applies to processing of personal data even though the individual providing the data does not see privacy to be at stake. This difference is highlighted by Nautsch et al. when they state that data protection regulations apply whether gathering and processing of the data was subject to privacy violations or not. Hence, while all EU member states comply with European data protection regulations they might interpret the concept of privacy differently [Nautsch et al., 2019].

One of the recent regulations applicable in the EU member states from 25 May 2018 is the *General Data Protection Regulation (GDPR)* [European Commission, 2016]. In Article 25 GDPR the *Data Protection by Design and by Default* concept is mentioned. It is therefore required that data protection principles are implemented effectively and appropriate technical and organisational steps are taken to protect the rights of data subjects [European Commission, 2016]. Moreover, data protection by default implies that only personal data which is necessary for the purpose is processed [Speaker Identifitcation Integrated Project (SIIP), 2018]. In Article 5 GDPR data protection principles that apply directly to the data controller are summarized. One of them is the "lawfulness, fairness and transparency" principle [European Commission, 2016]. Lawful refers to the processing on a legitimate basis while fair processing requires the data controller to inform the user about rules, rights and risks regarding the processing [Speaker Identifitcation Integrated Project (SIIP), 2018]. Due to the transparency principle, this information must be easily accessible and understandable to the user [Speaker Identifitcation Integrated Project (SIIP), 2018]. Moreover, one can directly derive guidelines from the data subjects' rights [Speaker Identifitcation Integrated Project (SIIP), 2018]. One that should be highlighted in the following is the Right to Erasure ("Right to be Forgotten") in Article 17 GDPR [European Commission, 2016]. Additional to the purpose limitation and data minimisation principles which already restrict and limit the processing of the data, the data subject can at any time invoke his or her right to be forgotten and the data controller is obliged to delete the data accordingly [Speaker Identifitcation Integrated Project (SIIP), 2018].

More recently, the *European Data Protection Board (EDPB)* released their *Guidelines on Virtual Voice Assistants* [European Data Protection Board (EDPB), 2021]. Their document should help to interpret and apply GDPR in the context of *Virtual Voice Assistants (VVAs)*. Especially, they state that as the main interaction mode is voice, users of VVAs should be able to invoke their rights using voice commands. Moreover, designers and developers need to consider that they might process data that falls into special categories e.g. when performing voice identification or managing data related to health [European Data Protection Board (EDPB), 2021].

In conclusion, we recognize that multiple discourses from different fields have shaped our current understanding of privacy and influenced its notion as well as enforceable legal regulations. Thereby, we need to consider that current data protection regulations in the EU do not include the fundamental right to privacy. Instead, individual nations can interpret the concept of privacy independently while complying with the GDPR. Nevertheless, we need to acknowledge the overlap between privacy and data regulations and their impact on one another. For this thesis, both conceptualizations will remain important as we are investigating peoples' perception of privacy as well as taking steps to provide effective strategies to enhance transparency and fairness as required by GDPR and the recently released Guidelines on Virtual Voice Assistants.

2.2 Conversational Privacy

While the GDPR clearly states the users right of being informed about risks and rights of data processing in an understandable manner, in practice privacy regulations are often displayed as complex legal texts [Brüggemeier and Lalone, 2022]. Especially, when using voice-enabled interfaces the user is forced to switch modalities as the information is presented on the screen rather than in speech form [Brüggemeier and Lalone, 2022]. This may be problematic from a legal perspective but even more from a usability perspective as modality switching is known to increase cognitive load and user errors [Brüggemeier and Lalone, 2022, Sandhu and Dyson, 2012]. While privacy policies are still mostly presented in complex legal text forms and therefore require modality switching, just recently, a few companies providing VUIs have adopted techniques to change privacy settings by using voice commands. For example, at the time of writing this thesis, Google and Amazon allow users to delete their voice recordings by saying “Hey Google, delete everything I just said.” or “Alexa, delete everything I've ever said.” whenever voice deletion is activated in the Alexa app [Teague, 2021]. Along the same lines, Harkous et al. proposed the concept of *Conversational Privacy Bots (PriBots)* where a text-based dialogue system is used to present privacy policies and enable changing of privacy settings in natural language text [Harkous et al., 2016]. It can either be used as a primary method of delivering privacy policies by initiating a dialogue with the user (e.g. “Hello there! I'm a bot that you can chat with about our privacy policy. You can ask me questions like: 'Does your app share my location?'”) or as a complementary way where the privacy policy is displayed alongside with the chatbot. Furthermore, the user can initiate a dialogue to change privacy settings e.g. the visibility of a user's birthday on Facebook by asking "Who can see my birthday?" and the chatbot responding "All your friends can view your birthday. Do you want to change this?". They refer to the idea of expressing privacy-related information in dialogue form as *Conversational Privacy* [Harkous et al., 2016]. In their study on Conversational Privacy, Brüggemeier and Lalone state that this can be used for text as well as voice-based conversational agents and may provide an effective way of putting GDPR requirements into practice [Brüggemeier and Lalone, 2022]. At the same time, the approach may help in building trust with the users as it enables users to stay informed and in control. They investigated how Conversational Privacy may affect user perceptions and choices in different contexts. In particular, they tested different Conversational Privacy strategies, e.g. the offer to delete data and the possibility to gain information about one's rights under GDPR in a banking, tax and music scenario. They showed that the two above mentioned strategies positively affect users perceptions of privacy and security across scenarios with no influence on usability. As expected, context played an important role regarding increasing privacy perceptions in the tax and banking scenarios and elicited no significant changes in the music scenario. Importantly, in the experiment, the users were asked directly after carrying out a task whether they wanted to delete their data or gain more information [Brüggemeier and Lalone, 2022]. This is different to the Conversational Privacy approaches implemented in current VUIs or the PriBot. There, users need to specifically initiate the respective dialogues to change privacy settings or gain privacy-related information. This is difficult on multiple levels. First, many users may not initiate a privacy-related dialogue in the first place as they currently do not inform themselves well [Schaub et al., 2018]. However, to carry out informed decision-making and to give informed consent users need to be aware, informed and literate. It is, therefore, crucial to investigate easily usable strategies which raise awareness and trigger users to inform themselves about their data and rights [Brüggemeier and Lalone, 2022]. Second, user privacy concerns and attitudes do not necessarily translate to privacy-preserving behaviour, a discrepancy that is known as the Privacy Paradox [Barth and de Jong, 2017]. Conversational Privacy initiated by the conversational agents rather than by the users themselves can thus provide a way to bridge the gap and help the users to translate their concerns into actual behaviour. To better understand how exactly this could be done, we will now take a closer look at the Privacy Paradox itself as well as on possible explanations.

2.3 Privacy Paradox

"Herein lies the privacy paradox. Adults are concerned about invasion of privacy, while teens freely give up personal information. This occurs because often teens are not aware of the public nature of the Internet." [Barnes, 2006].

This quote stems from an article by Barnes and is seen as the origin of the term *Privacy Paradox* to describe the discrepancy between peoples' privacy concerns and their actual behaviour [Masur, 2019]. Although, in her study, the term Privacy Paradox refers to the different privacy behaviour of age groups, Norberg et al. later established the term in a broader context and it has since been used to describe the phenomenon that even people with strong privacy concerns might give away personal information more or less freely [Barnes, 2006, Norberg et al., 2007, Masur, 2019]. The dichotomy between privacy attitudes and concerns and actual behaviour has been heavily researched in contexts such as e-commerce and social networks with sometimes contradicting findings [Masur, 2019, Barth and de Jong, 2017]. While several studies provide evidence that supports the Privacy Paradox hypothesis others found that peoples' behaviour is indeed in line with their privacy concerns and attitudes [Kokolakis, 2017]. In his article, Kokolakis gives several possible explanations for the causes of these contradictory findings. Importantly, he points to the fact that privacy behaviour is inherently contextual and a comparison of results from studies conducted in different contexts can therefore easily lead to contradicting findings. Similarly, the sensitivity of personal information might vary depending on the individual and has often been neglected as a moderator in the existing literature. Moreover, it is necessary to keep different research methodologies in mind. While survey studies are reliable in measuring attitudes, experiments are more appropriate for measuring actual behaviour. However, one has to be cautious when generalizing the results as we can not expect people to behave the same during the experiment as they would normally do. In conclusion, he states that the Privacy Paradox should not be considered a paradox anymore as researchers have provided comprehensive models and explanations [Kokolakis, 2017]. In his recent article "The Myth of the Privacy Paradox" Solove goes one step further and argues that "it only appears to be a paradox because of conflated issues and flawed logic" [Solove, 2020]. In particular, he claims that the goal of aligning people's privacy attitudes and behaviour is a faulty approach. Privacy attitudes are mostly addressed in a general and even abstract manner, across many contexts, while behaviour is about risk evaluation of potential harm in one specific context. This is very different from the value people ascribe towards privacy. While some may exchange personal information in one specific context for benefits they might still value control over their privacy in general. One should therefore not conclude that people do not value their privacy because they show different behaviour. Moreover, privacy regulations should not be based on peoples' behaviour and on the outcome of those studies as it might lead to less regulation. On the other hand, Solove critizes that current regulations ask for more privacy self-management and vast and complex privacy controls that people are likely to fail in protecting their privacy. In this light, he argues for privacy policies to go beyond giving users more control and instead provide clear guidance on secure product design and system architecture. While one should not try to align attitudes and behaviour, researchers and developers should accept the existence of the gap due to ways of experimental setups and assessment [Solove, 2020]. Nevertheless, the discrepancy between privacy attitudes and privacy behaviour remains an open issue – especially in light of new emerging technologies and their associated contexts which pose new challenges to peoples' privacy self-management [Kokolakis, 2017]. Therefore, it is not surprising that the Privacy Paradox has extended to the world of *Internet of Things (IoT)* since the rise of smart speakers and other smart devices [Williams et al., 2017, Konrad et al., 2020].

Conversational agents appear in numerous ways and formats in today's life. First, users are mostly confronted with text-based conversational agents on websites, in educational or health care settings. They often use natural language processing to interact with a user to provide help or make recommendations in a conversational manner [Ischen et al., 2020]. While privacy concerns have been long researched with respect to the usage of websites and their effects on disclosure, little evidence exists for the use of chatbots [Ischen et al., 2020]. Ischen et al. compared users privacy concerns and behaviour between interactions with a human-like chatbot, a machine-like chatbot and a website [Ischen et al., 2020]. They found that a human-like chatbot increases perceptions of anthropomorphism such that people attributed friendliness and socialness to the respective chatbot. This then resulted in decreased privacy concerns and more willingness to share personal information compared to the machine-like chatbot. Additionally, they revealed a direct effect on privacy concerns when comparing the machine-like chatbot and the website. Here, users reported higher privacy concerns and showed less disclosure when being exposed to the website than to the chatbot [Ischen et al., 2020].

Much more research has been conducted in the field of voice-based conversational agents and *Intelligent Personal Assistants (IPAs)* such as Siri and Alexa regarding privacy attitudes and user behaviour. Similarly to chatbots, we are faced with an increasing number of possible contextual features, e.g. people use voice assistants at home, at work, in the car, in restaurants or during shopping [Vixen Labs Limited in partnership with Open Voice Network, 2021]. Context might play an important role for researchers to be considered, however, data protection regulation are independent of context and instead focus on special categories of data such as biometric data or medical data [European Data Protection Board (EDPB), 2021]. While Apple's Siri is primarily used on smartphones, Amazon's Alexa is mostly used on smart speakers in the United States (US), Germany and the United Kingdom (UK) [Vixen Labs Limited in partnership with Open Voice Network, 2021]. The distinction between device and agent is crucial as information disclosure varies accordingly [Ghosh and Eastin, 2020]. With smartphones being well established in today's world, smart speakers are still relatively new to customers and do not provide the same flexibility as smartphones as they are usually fixed to one location [Ghosh and Eastin, 2020]. Privacy concerns also differed depending on the device that was used. They were less prominent when people were interacting with a voice assistant on their smartphone than with a smart speaker [Ghosh and Eastin, 2020]. In another survey study, Lutz and Newlands examined different kinds of privacy concerns such as device privacy concerns, third party privacy concerns or government privacy concerns and found that they had limited impact on privacy preserving behaviour [Lutz and Newlands, 2021]. When conducting focus groups with IPA users and non-users it was found that privacy concerns are especially subject to risks of unexpected recordings and uncertainty of data collection [Vitak, 2020]. Because of the “always listening” mode users were uncertain when data is recorded and how the data is used or shared [Vitak, 2020]. Furthermore, users expressed that they might lack skills and knowledge to use existing privacy settings in the case of IPAs [Vitak, 2020]. Javed et al. could show that while users are concerned about their unintended recordings on the Alexa device they lack understanding on who can access the data and if and especially how one can delete Alexa’s voice recordings [Javed et al., 2019]. Interestingly, there was no significant difference neither between groups with and without technical background nor between groups expressing different levels of privacy concerns [Javed et al., 2019]. In a recently conducted interview study on privacy in human-machine interaction, we similarly found that uncertainty and vague risk assessment were not influenced by technical knowledge [Leschanowsky et al., 2021]. Instead, participants with and without technical education expressed their concerns and uncertainty regarding their disclosure towards machines and service providers [Leschanowsky et al., 2021]. Finally, research has shown that the Privacy Paradox seems to be more prevalent in IoT scenarios due to a lack of transparency and awareness [Williams et al., 2017, Konrad et al., 2020].

As seen above, the Privacy Paradox stimulates ongoing discussions and there have been many attempts to explain the gap between peoples’ attitudes and actual behaviour [Masur, 2019]. So far, there is neither an accepted theory about users online behaviour nor an agreement on underlying mental processes in the context [Barth and de Jong, 2017]. Barth and de Jong conducted a systematic literature review to categorize existing theories on the Privacy Paradox [Barth and de Jong, 2017]. Figure 2.1 shows their categorization of theories to explain the Privacy Paradox and the nature of decision-making. Two major streams can be identified. The first category consists of decision-making based on risk-benefit calculation whereas the second category comprises decision-making based on benefits solely with little to no risk assessment [Barth and de Jong, 2017]. In the following, we will further investigate decision-making based on a risk-benefit calculation. First, we will shortly introduce the idea of *Privacy Calculus*, the rational risk-benefit calculation, before taking a closer look at a risk-benefit assessment that is subject to cognitive biases.

2.4 Theories on the Privacy Paradox - Risk-Benefit Calculation

2.4.1 Rational Risk-Benefit Calculation

In general, theories on the Privacy Paradox can be divided into two streams [Barth and de Jong, 2017]. One stream of research is based on a rational view of decision-making. This goes back to the *Rational Choice Theory of Human Behaviour* which states that people apply rational

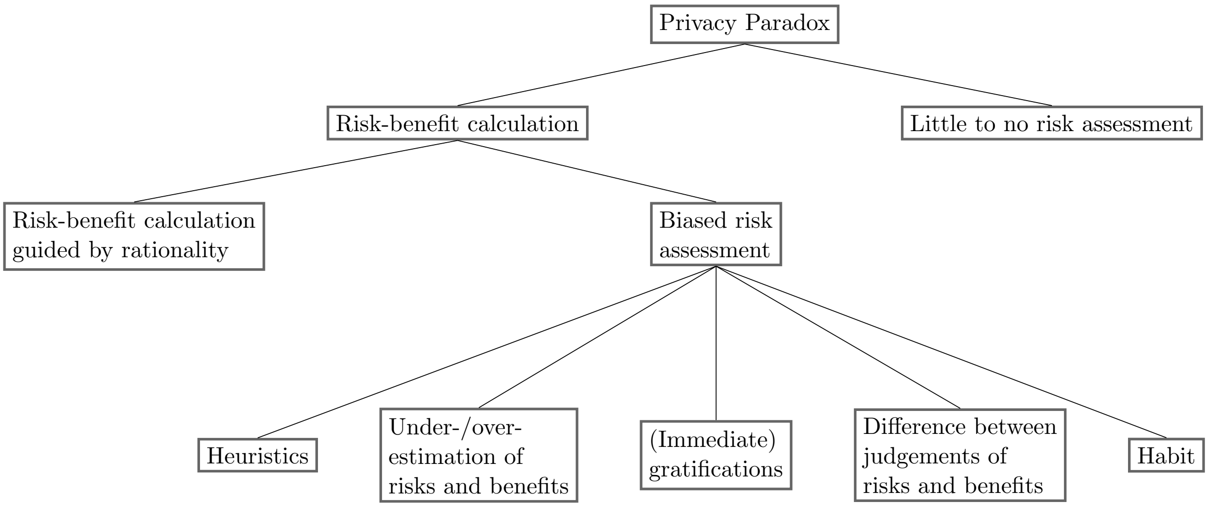

Privacy Paradox
Risk-benefit calculation
Little to no risk assessment
Risk-benefit calculation guided by rationality
Biased risk assessment
Heuristics
Under-/over-estimation of risks and benefits
(Immediate) gratifications
Difference between judgements of risks and benefits
Habit

Figure 2.1: Overview of theories on the Privacy Paradox adapted from [Barth and de Jong, 2017].

calculus by maximizing utility and minimizing risks in order to make decisions [Simon, 1955]. Therefore, decision-making is seen as a conscious and analytical process [Barth and de Jong, 2017]. Here, benefits have a positive contribution to disclosure intentions while costs reduce the likelihood to disclose information [Barth and de Jong, 2017]. Culnan and Armstrong were first to adapt the concept to privacy research in the context of e-commerce and coined the term Privacy Calculus [Culnan and Armstrong, 1999, Masur, 2019]. The way to explain the Privacy Paradox with Privacy Calculus theory simply stems from the observation that often the perceived benefits outweigh the costs and lead users to ignore their concerns and disclose information [Barth and de Jong, 2017]. The model was later extended to take the *Big Five Personality* as well as cultural factors into account [Pentina et al., 2016]. As seen earlier, the Privacy Paradox has been critized for relying on general privacy attitudes in comparison with context-dependent behaviour. Kehr et al. were able to show that disclosure is influenced heavily by situational Privacy Calculus as situational aspects can fully outweigh pre-existing privacy concerns and attitudes [Kehr et al., 2015]. Privacy Calculus has also been applied to conversational agents and IoT scenarios. However, studies revealed limited applicability by showing that privacy concerns have only little impact on privacy protection behaviours [Lutz and Newlands, 2021]. When examining IoT scenarios, perceived benefits positively impact peoples' disclosure, but perceived risks and information sensitivity did not [Kim et al., 2019]. Therefore, a growing stream of research has started investigating factors beyond costs and benefits to be included into the theoretical models and extended the Privacy Calculus model by including possible mediators and moderators such as self-efficacy [Kang and Oh, 2021].

While many approaches exist based on rational Privacy Calculus, the majority of theoretical frameworks that try to explain the Privacy Paradox include non-rational concepts and theories that influence decision-making [Barth and de Jong, 2017]. We will now give a short overview of approaches accounting for bias in risk-benefit calculation before taking a more detailed look at the *Dual-Process Theory* [Kahneman, 2011].

2.4.2 Biased Risk Assessment

Privacy theories in the category of biased risk assessment try to explain the paradoxical behaviour of users by adding biases such as time constraints, situational cues, habitual use or immediate gratification to the decision-making process [Barth and de Jong, 2017]. The concept differentiates largely from a rational cost-benefit evaluation as those biases usually influence decision-making subconsciously. Instead, the explanation of the Privacy Paradox stays the same as perceived benefits may outweigh perceived risks which are however subject to biases [Barth and de Jong, 2017]. Another important framework to mention in this context is the theory of *Bounded Rationality* [Simon, 1997]. In their literature review, Barth and de Jong state that Bounded Rationality heavily influences the decision-making process whenever options are infinite, consequences become unpredictable and uncertainty is reigning subconsciously. Especially in modern technological contexts individuals might have difficulties or are unable to process all the necessary information to make informed decisions [Barth and de Jong, 2017]. It is therefore obvious that even if from the user point of view rational assessment of risks and benefits takes place, decision-making will be limited based on the availability of time, information and cognitive resources [Barth and de Jong, 2017]. More so with all information available people might rely on simple heuristics or mental short cuts as the analysis of all aggregated information requires substantial cognitive involvement [Barth and de Jong, 2017]. Situational cues such as the design of an application or conversational agent could trigger affective-thinking resulting in a biased evaluation [van der Heijden, 2013]. Similarly, Kehr et al. could show that feelings and emotions influence risk perceptions and even outweigh rational factors [Kehr et al., 2015]. In particular, they found that risk perceptions of participants in a neutral affective state were influenced by the level of information sensitivity while information sensitivity did not influence risk perceptions of participants in a positive affective state [Kehr et al., 2015].

Furthermore, underestimation of privacy-related risks is a common phenomenon which goes hand in hand with overestimating the chances that others will experiences privacy breaches rather than oneself [Barth and de Jong, 2017]. This then lead to less engagement in privacy-preserving behaviour [Barth and de Jong, 2017]. Especially in social media networks, users think of themselves as taking only advantages of disclosure while others might also experience negative effects [Barth and de Jong, 2017]. The theory is based on *Third-Person Effect Theory* which states that people will perceive persuasive communication to affect others more greatly than themselves [Davison, 1983]. A larger corpus of research investigated the role of immediate gratification on disclosure behaviour [Barth and de Jong, 2017]. It has been shown in many contexts that people favour small benefits in the short term rather than larger benefits in the long term [Barth and de Jong, 2017]. Again, situational cues can override potential long term risks and their associated general privacy concerns by highlighting immediate gratification [Barth and de Jong, 2017].

Drawing on *Prospect Theory* and *Quantum Theory* researchers considered a difference between the judgement of risks and benefits [Barth and de Jong, 2017]. In Prospect Theory everything below a reference point, which could be specific to peoples' current situation, is counted as a loss and everything above is considered a gain [Kahneman, 2011]. The corresponding value function indicates that losses dominate gains which leads people to a more loss-averse behaviour whenever gains have moderate probabilities or chances to experience loss are low [Kahneman, 2011]. In contrast, people are risk-seeking whenever the chances to loose are only moderate or the probabilities of gains are small [Kahneman, 2011]. Related to privacy in the social media context, Prospect Theory was used to explain why users have decreased privacy risk perceptions and base their decisions on what they can gain rather than lose [Barth and de Jong, 2017]. Moreover, with privacy attitudes and risks being abstract concepts for many people, it is no surprise that concrete benefits may be more salient for evaluation [Barth and de Jong, 2017]. Another category of research, addresses the role of habit and repetitive behaviour which has been especially investigated in the area of social media and networks [Barth and de Jong, 2017]. Due to the important role of those sites in people's everyday life even direct privacy violations can not prevent users from engagement and benefits may again override privacy attitudes.

2.5 Dual-Process Theory: Thinking, fast and slow

Multiple of the theories used to explain biased risk-assessment, e.g. heuristics or Prospect Theory go back to research conducted by Daniel Kahneman and Amos Tversky [Kahneman, 2011]. In his book "Thinking, fast and slow", Kahneman, 2002 Nobel Prize winner in Economics, summarizes his current understanding of judgement and decision-making. Further, he describes the journey he and Tversky set out on when meeting in 1969 and how it has shaped the understanding of decision-making [Kahneman, 2011]. Of course, Dual-Process Theory, in particular the distinction between fast and slow thinking has been researched by many in the field [Stanovich and West, 2000, Evans, 2008, Kahneman, 2011]. However, to provide a basic understanding we will focus on explanations and descriptions provided in Kahneman's book "Thinking, fast and slow" [Kahneman, 2011]. He describes mental processes carried out by two agents, *System 1* and *System 2.* It is important to keep in mind that System 1 and System 2 are fictitious and should not be understood as systems in the traditional sense as groups of interacting elements. Neither can distinctive parts of the brain be associated with exactly one of those two systems. Nevertheless, treated as fictitious characters with personalities and traits we can easily understand their different roles in the decision-making process. Actions such as "detecting one object that is more distant than

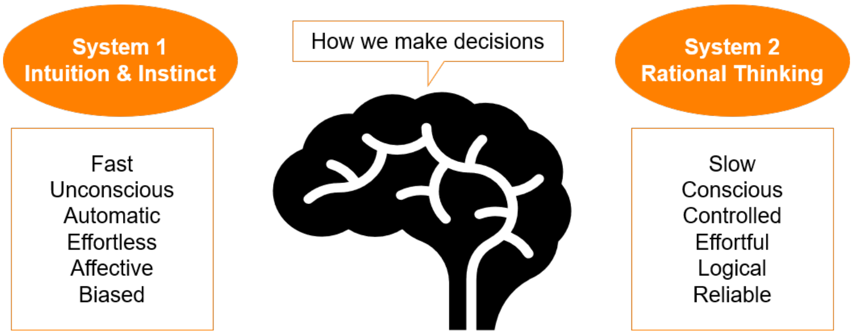


Figure 2.2: System 1 and System 2 Characteristics adapted from [Kahneman, 2011].

another", "orient to the source of a sudden sound" or "driving a car on an empty road" are examples of automatic processing and System 1 activity. In contrast, System 2 is in charge of effortful mental activities for example "focusing on a voice of a particular person in a noisy room" or "parking in a narrow space". Figure 2.2 provides an overview of characteristics of System 1 and System 2.

While most people will identify with System 2, the conscious system with rational thinking capabilities, we may not forget the importance and influence intuitive thinking has on our decisions and judgements. Understanding the existence and differences of the two systems also requires an understanding of their interactions. When we are awake both of the systems are active. If things go smoothly System 1 generates "impressions, intuitions, intentions and feelings" which are mostly excepted and turned into actions by System 2. This means that most of what we think and do has its origin in System 1. Only when things get difficult or violate the existing mental model of the world, System 2 is activated and takes over. This is especially true when people experience surprise, doubt or uncertainty. System 1 is not known for conscious doubt or thinking of alternative options. This will become important once we talk about how to trigger System 2 thinking.

System 2 normally runs in a low effort mode and as long as we experience *Cognitive Ease* there is no need to change that. Being in a good mood, repeated experiences, clear display or primed ideas are all things that can set us into a state of Cognitive Ease. Consequently, they lead to a feeling of familiarity, truth, well-being and low effort which then again can be cause of Cognitive Ease. Especially interesting is the effect of *Priming* which has been shown to be consistent and robust but not necessarily large in many different contexts. Priming occurs for example when being exposed to a word and consequently finding it easier to produce related words. In that case, words would induce Priming and at the same time be affected by Priming. However, priming effects are not only related to words but also to actual behaviour. In a study by the psychologist Bargh et al., young students who were primed with words associated with old age showed slower walking than those not primed. In his book, Kahneman describes that priming effects can take on even more forms such as actions, emotions and perceptions. On the other hand a variety of primers, things that induce priming, have been analyzed additionally to common primers like words or gestures. The Priming phenomenon that our actions are influenced by ideas or events is certainly a System 1 activity and happens without us being aware of it. Similarly, to keep System 2 in a low-mental effort state, System 1 is an expert in substitution and the core of heuristics and biases. In order to easily generate quick answers to difficult questions like “How much would you contribute to save an endangered species?” System 1 might simplify the task and finds answer to an heuristic question like “How much emotion do I feel when I think of dying dolphines?”. Additionally, System 1 has the skills to perform matching between the emotional scale and the scale of contribution and to come up with an answer to the original question. While System 2 surely has the ability to reject the intuitive answer or modify it accordingly, due to its laziness this often will not be the case. This effect becomes especially interesting when our likes and dislikes dominate our beliefs about risks and benefits. Somebody who dislikes smart speakers will probably believe that its risks are high and its benefits are low. Vice versa a person who is in favour of smart speakers will imagine its benefits to be huge and the risks to be neglectable. The *Affect Heuristic* is prominent for explaining biased cost-benefit calculation in various contexts.

The fact that humans are subject to heuristics and biases as described by Kahneman and Tversky does not mean that human decisions are irrational in general [Kahneman, 2011]. Instead, they claim that human behaviour is not explained well by a rational-agent model. For “true believers in human rationality”, freedom does not come with any cost as human decisions are not subject to mistakes and as long as nobody is harmed, people should be allowed to act as they want to. However, for behavioural scientists, decision-making is more complex and whenever society has to cope with peoples bad choices, freedom does indeed cost. In those cases, strategies, policies and institutions can help people to improve their judgments and make better decisions [Kahneman, 2011]. In the following, we will investigate strategies to overcome cognitive biases used in other fields and their applicability to privacy and conversational agents.

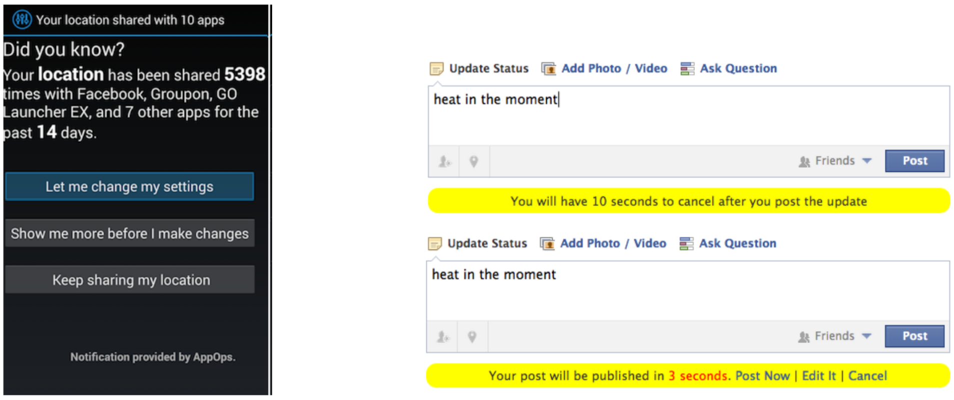
2.6 Nudges and Debiasing Strategies

From Section 2.5 it became clear that according to behavioural scientists, the human decision-making process is subject to heuristics and biases. One possibility to overcome biases and help humans to make informed decisions was presented in the book “Nudge” by Thaler and Sunstein. Following their idea of libertarian paternalism, people can be nudged by institutions or the state to make more accurate judgements regarding their long-term interests. They define the concept of a *Nudge* as follows:

“A nudge, as we will use the term, is any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their

economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates. Putting fruit at eye level counts as a nudge. Banning junk food does not." [Thaler and Sunstein, 2008].

Such a Nudge could also be a default option of being enrolled in a pension plan as most people will not devote the effort it needs to deviate from the default option [Kahneman, 2011]. Thaler and Sunstein also refer to Nudges when it comes to easily readable and understandable policies and contracts similar to what is required according to GDPR [European Commission, 2016]. These strategies are by no means intrusive but instead can help humans to make good and informed decisions considering possible cognitive flaws. In an article from 2013, Acquisti describes how a soft paternalism solution like nudging can be helpful in the privacy context [Acquisti, 2009]. He refers to privacy nudging in the social media context where people can post their dates of birth, data which can easily lead to inferences of sensitive data like a person's social security number. A strong but non-practical paternalistic approach would prohibit users to share this kind of data online while an approach focusing on usability and control will allow users to easily and intuitively change visibility settings for birth dates. A nudging approach would directly set the default settings to invisible or visually indicate how many other users can see the information [Acquisti, 2009]. Since then nudging approaches have been investigated in the field of mobile apps, app development or social media [Almuhimedi et al., 2015, Choe et al., 2013, Balebako and Cranor, 2014, Wang et al., 2013]. Almuhimedi et al. implemented a Nudge for mobile applications which displays privacy-relevant information such as how often a location was shared with different apps (see Figure 2.3a). Further, the users could choose from possible actions like being forwarded to the settings to change them. They found the reviewing app permission and privacy nudges to be effective with 95% of participants reviewing their permissions and 58% changing the corresponding settings [Almuhimedi et al., 2015]. In the context of social media, Wang et al. investigated three types of privacy nudges to encourage real-time adjustments when posting content on Facebook [Wang et al., 2013]. In particular, they used a picture Nudge, displaying a randomly chosen group of people who could view the post the user was about to write. The goal was to encourage users to consider their audience before posting content. In Section 2.4.2 we discussed the effect of immediate gratification on the assessment of risks and benefits. This effect might be especially prevalent in the context of social media. Therefore, a timer Nudge was designed that would delay the post and allow users to reflect and possibly cancel their action (see Figure 2.3b). A third Nudge did provide feedback on how other users might receive the post e.g. if the post may be received positive or negative. Overall, the timer Nudge was perceived positively by study participants as it provides the chance to correct typos, to post better quality content or even cancel unnecessary posts. Yet, the time delay was sometimes perceived as annoying. While the picture Nudge was judged as useful, the sentiment Nudge was not. Some of the reasons were that the algorithm did not predict the sentiments sufficiently well or that participants did not care about negative sentiment warnings and in the worst case



Your location shared with 10 apps

Did you know?

Your **location** has been shared **5398** imes with Facebook, Groupon, GO Launcher EX, and 7 other apps for the past **14** days.

Notification provided by AppOps.

heat in the moment

You will have 10 seconds to cancel after you post the update

heat in the moment

Your post will be published in 3 seconds. Post Now | Edit It | Cancel

(a) Privacy Nudge for Location implemented by Almuhimedi et al. [Almuhimedi et al., 2015].

(b) Timer Nudge implemented by Wang et al. [Wang et al., 2013].

Figure 2.3: Examples of privacy nudges investigated in the context of mobile apps and social media.

got annoyed with the Nudge [Wang et al., 2013]. So far, the proposed nudging strategies focus on the users and how to make them adopt privacy-preserving behaviour. However, there has been research on how to help the developers of those apps to overcome hurdles to good privacy decision-making [Balebako and Cranor, 2014]. In a qualitative study, Balebako and Cranor found that privacy guidelines are not effective in educating app developers and that privacy is not a priority that is focused on in the development process [Balebako and Cranor, 2014]. They suggest that platforms could provide privacy-preserving *Application Programming Interfaces (APIs)* rather than allowing developers to assess sensitive information by default and to put the documentation in places where it can be easily found by the developers. Moreover, cloud services, a commonplace for app developers to store their users' data, could require to set retention periods and remind developers to delete old data [Balebako and Cranor, 2014]. Hosseinzadeh. et al. describe an approach to easily create, implement and adapt privacy and security related specifications and policies whenever a data exchange between parties is required [Hosseinzadeh. et al., 2020]. Those automatic functions can include automatic deletion after a specified period or masking the data for the other party [Hosseinzadeh. et al., 2020].

In the medical field, a vast body of research on cognitive interventions based on the Dual-Process Theory exists. Studies suggest that 75% of diagnostic errors are subject to cognitive failures and not cases of insufficient knowledge [Lambe et al., 2016]. In there systematic review, Lambe et al. state that debiasing techniques to overcome mental shortcuts and avoid diagnostic errors have been broadly classified into educational strategies and workplace strategies. While educational strategies such as providing seminars to increase knowledge and awareness of reasoning styles could be interesting approaches in the field of conversational agents, in this work we will focus only on workplace strategies. Workplace strategies aim to enhance decision-making in the moment [Lambe et al., 2016]. This is suitable for our case as we want to implement techniques that can help the user to make better decisions during the interaction with the conversational agent.

Checklists are a very common tool to reduce cognitive failures as they provide consistency and ensure completeness of a task [Wikipedia, 2021]. Many use them daily in form of to-do lists, but they are also frequently applied during software development processes or to guarantee flight safety before departure [Wikipedia, 2021]. Diagnostic checklists or debiasing checklists have been investigated in the medical context resulting partly in fewer errors and an increased correction of errors [Lambe et al., 2016]. While diagnostic checklists usually state possible alternative diagnoses or special diagnoses that should not be missed, debiasing checklists provide step-by-step guidance to diagnosis [Ely et al., 2011]. Debiasing checklists can also include items on diagnostic timeouts e.g. to take time to stop and re-think [Ely et al., 2011]. So far, CUIs support the creation and management of checklists by the user such as grocery shopping lists or to-do lists. softengi.com offers a voice controlled checklist application based on voice recognition technology e.g. a voice controlled surgery checklist for anesthesiologists which ensures that critical safety steps are carried out and key points are performed in an appropriate order. On top of that, privacy-related checklists could be used in the context of CUIs e.g. before installing a new skill or when using a service. Such a checklist could include items on privacy settings, service type and alternative skills or applications. Moreover, users could specify their overall privacy requirements using a checklist which then can be used by the device to check that all requirements are met whenever installing a new application.

The idea of cognitive forcing strategies stems from the medical education field and the importance of metacognition [Croskerry, 2003]. Thus experts who are aware of their mental limitations, able to criticize themselves and their decision-making realistically and to select appropriate strategies to make improved decisions are less likely to commit diagnostic errors [Croskerry, 2003]. According to Croskerry "cognitive forcing strategies are a specific debiasing technique that introduces self-monitoring of decisionmaking [sic!].". He makes clear that clinicians should therefore not pursue intuitive pattern recognition during the diagnostic process but should broaden their view and consider other possible diagnoses. This is rather difficult and involves a multi-stage learning framework as cognitive processes are invisible and not easily accessible [Croskerry, 2003]. The original idea of cognitive forcing strategies requires clinicians to apply metacognitive steps and techniques consciously without any further help from outside but only by undergoing the respective training. However, cognitive forcing strategies have been applied in the medical context as workplace strategies which consist of instructions to consider alternatives or to reconsider diagnoses [Lambe et al., 2016]. It was found that when participants were asked to consider alternatives during the diagnostic processes compared to carrying out diagnoses based on first impression or without instructions, diagnostic accuracy increased [Lambe et al., 2016]. Similarly, reconsidering the diagnosis after details from the case outline were removed, improved the accuracy significantly [Lambe et al., 2016].

In the field of design, forcing functions – design aspects that put a physical constraint on people to prevent them from inappropriate behaviour – are common [Norman, 2002]. They are especially useful in safety-critical work processes as they prevent users from proceeding once failure was detected [Norman, 2002]. Forcing functions are used in everyday life e.g. starting a car requires to carry a physical key or in modern cars some kind of physical token for authorization [Norman, 2002]. As strong constraints can not be imposed everywhere, variations of forcing functions can be applied to various situations and the definition of forcing functions is not anymore exclusive to physical design aspects [Norman, 2002, Interaction Design Foundation, 2021]. We want to highlight one prominent example used in many computer applications and brought up by Norman in his book on "The design of everyday things". The message prompted to the user whenever he or she tries to exit an application without having saved the work can be interpreted as a forcing function. The prompt usually asks whether one wants to exit without saving, save the work or cancel the operation and can be used as a shortcut to easily save work before closing the application. Forcing functions are an effective tool to make it unnecessary for people to remember certain steps to take and therefore avoid memory-lapse errors [Norman, 2002]. More generally, they now refer to techniques which "force the users' conscious attention upon something" and therefore disrupt the intuitive, automated behavioural pattern [Interaction Design Foundation, 2021].

Since their emergence, the definitions of cognitive forcing strategies and forcing functions have changed and developed further. The way they are applied and interpreted nowadays is quite similar and also strongly connected to the nudging approach discussed above. However, they differ in how much freedom they give to the user to proceed without correcting for mental flaws and how much weight they put on self-monitoring. Let's consider a concrete example – a supermarket offering only fruit and unhealthy food. A forcing function imposing strong constraints would lead to banning unhealthy products altogether and to only offering fruit. A nudging approach as described above could put fruit at eye level and unhealthy products to higher or lower levels. Instead, a cognitive forcing strategy could put both, fruit and unhealthy products, at the same level but buyers of unhealthy products would be asked to consider buying fruits instead. The distinction between nudges and cognitive forcing strategies is however subtle and other studies might use the terms interchangeably. In a recent study cognitive forcing strategies were evaluated in the context of *Artificial Intelligence (AI)*-assisted decision-making [Buçinca et al., 2021]. The authors, Buçinca et al., refer to cognitive forcing strategies as "an umbrella term for interventions that elicit thinking at the decision-making time". They tested three different strategies and whether they had an influence on peoples' overreliance on the AI recommendations. One of the strategies consisted of asking the person to make a decision before the AI recommendation was shown. A second strategy introduced a delay before presenting the AI recommendation while the third one let people choose whether they wanted to see the AI advice at all. They could show that the cognitive forcing strategies had a significant effect on overreliance when compared to explainable AI approaches [Buçinca et al., 2021]. In addition, they investigated whether people with different levels of *Need for Cognition* benefit equally from the cognitive forcing strategies. The refer to Need for Cognition as “a stable personality trait that captures one’s motivation to engage in effortful mental activities”. People with a high level of Need for Cognition tend to prefer more complex user interfaces and to gather more information than those with a low level of Need for Cognition. Indeed, they could show that mostly people of high Need for Cognition did benefit from the cognitive forcing strategies in terms of performance improvements. However, this group of participants also trusted and preferred the system less compared to the simple AI approach, something that was not found statistically significant in the case of participants with low Need for Cognition [Buçinca et al., 2021].

Another technique that was investigated in the medical context to increase diagnostic accuracy is guided reflection [Lambe et al., 2016]. Guided reflection refers to a concept in “which the practitioner is assisted by a mentor (or ’guide’) in a process of self-enquiry, development, and learning through reflection” [Johns and Johns, 2010]. The reflective practice should lead to more critical thinking of the own decision-making process [Johns and Johns, 2010]. In many medical studies on guided reflection participants are given the task to follow a set of procedures to diagnose a case [Lambe et al., 2016]. Different to checklists where they might be reminded of possible alternative diagnoses, in guided reflection participants are given detailed instructions on what to consider e.g. “list findings that support this hypothesis” or “list alternative hypotheses if the first hypothesis proved to be incorrect” [Mamede et al., 2008]. This set of instructions serves as a guide and should induce reflective reasoning instead of stating the first diagnostic hypothesis that comes to mind [Johns and Johns, 2010]. In the medical context, reflective reasoning and guided reflection was found especially helpful when complex or unique problems needed to be diagnosed [Mamede et al., 2008]. In the context of CUIs various possibilities to adapt guided reflection practices are imaginable. First, the CUI itself could function as a guide to assist users in their development and decision-making process. This could not only be related to decisions which affect users’ privacy but to decisions on health, shopping behaviour or leisure activities. Second, regarding users’ privacy several privacy-related questions could be asked by the CUI to trigger reflective reasoning. Again, instead of simply listing privacy settings to be checked or alternative skills or applications as in the checklist approach, the CUI could ask users to find and list privacy-related information themselves. This could possibly lead to a state where guided reflection is no longer necessary and users automatically engage in reflective reasoning before interacting with services or disclose information to applications.

We now want to introduce one last category, similarly to guided reflection as a workplace strategy rather than an educational strategy. Instructions at test were used by researchers in the medical context to reduce diagnostic errors [Lambe et al., 2016]. Instructions varied widely from instruction for dual-process reasoning, instructions containing a list of clinical features, instructions for thoughtful diagnoses and instructions for quick diagnoses [Lambe et al., 2016].

While some studies found differences between the compared groups, others did not [Lambe et al., 2016]. Instructions were investigated not only in the medical context but also in the field of user acceptance. van der Heijden designed an experiment to test whether users can be primed towards either System 1 or System 2 thinking. They were shown a description and screenshot of an application and were then asked if they were willing to download the application. The application triggered a negative System 1 response by being designed in an "ugly" way while the description was supposed to trigger a positive System 2 response. The authors evaluated Priming questions shown right after displaying the application and description but before the evaluation took place. System 2 Priming questions such as "thinking carefully about the advantages and disadvantages of the system do you agree or disagree that the application is beneficial?" were asked. Even though the authors use the underlying concept of Priming we can easily see the similarity between the instructions of diagnosing thoughtfully and the question of "thinking carefully". They were able to show that users evaluating the application based on first impression were less likely to download it than users who were asked to carefully think about the benefits [van der Heijden, 2013]. This indicates that those kind of strategies are not only useful in the medical field but can be adapted to other disciplines.

An overview of the different strategies described in this section is given in Table 2.1 together with a CUI example application that is, however, unrelated to privacy. While some of the strategies have been applied in privacy-related contexts, to the best of our knowledge they have not yet been investigated with regards to privacy in CUIs. Moreover, we display an indicator for the applicability of the strategies in the context of CUIs. Their applicability to CUIs is based on our own understanding and view on those strategies i.e. ideas on implementing checklists or guided reflection are provided above. In this thesis, we will focus solely on cognitive forcing strategies and their usefulness for privacy-related decision-making when interacting with CUIs. For the remaining part of this thesis, we will refer to the implemented strategies as cognitive forcing strategies while acknowledging that they are designed in a way to nudge people into more privacy-preserving behaviour.

In the next chapter, we will introduce the conceptual framework of this thesis and present the conducted experiments in detail.

Strategies	Definition	CUI Example Application	Applicability in the CUI context
Nudges	"Any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" [Thaler and Sunstein, 2008]	Proactive virtual assistant nudging people towards energy-saving actions [He et al., 2021]	✓
Checklists	"List of actions or things that need to be done or thought about" [Wikipedia, 2021]	Voice Checklist for Anesthesiologists [softengi.com, 2022]	✓
Cognitive Forcing Strategies	"Debiasing technique that introduces self-monitoring of decision making" [Croskerry, 2003]		✓
Forcing Functions	Design aspects that impose strong constraints to avoid inappropriate behaviour		✓
Guided Reflection	"concept in which the practitioner is assisted by a mentor (or 'guide') in a process of self-enquiry, development, and learning through reflection" [Johns and Johns, 2010]	Conversational agent for workplace reflection Kocielnik et al. [2018]	✓
Instructions at test	variety of different instructions, especially interesting instructions are to diagnose carefully vs. diagnose quickly	Voice Assistant reminders for pain self-management Shade et al. [2020]	✓

Table 2.1: Overview of cognitive interventions suggested by behavioural economists (Nudges) [Thaler and Sunstein, 2008], designers to avoid inappropriate behaviour [Norman, 2002] and strategies used in the medical workplace to avoid diagnostic errors due to heuristics and biased reasoning [Lambe et al., 2016]. We display example applications for voice assistants and conversational agents. We could not identify applications specifically referring to cognitive forcing strategies or forcing functions in the field of CUI.

Chapter 3

Practical Experiments

In this chapter, we will give a detailed description of our conceptual framework, the experimental setup and results of the experiments. We start by introducing the framework and building blocks of the studies in Section 3.1 to Section 3.4. Further, the pilot study will be discussed in Section 3.5 before presenting the experimental setup on the experiment on cognitive forcing strategies and the corresponding results in Section 3.6.

3.1 Theoretical Framework

Our conceptual model (shown in Figure 3.1) is based on a previously carried out qualitative study and on the theoretical background provided in Chapter 2. Qualitative studies are especially useful whenever one aims to gather attitudinal data and understand users' underlying motivations in a specific context [Olson and Kellogg, 2014]. By conducting semi-structured interviews we aimed to better understand factors that influence privacy in human-machine interaction and peoples' decision-making about disclosing personal information [Leschanowsky et al., 2021]. Moreover, we were interested in factors which distinguish privacy in *HMI* from privacy in human-to-human interaction. We found that there are at least two main reasons for users to immediately reject a service. First, users could assess benefits clearly while costs remained elusive and difficult to understand. This could lead to an irrational benefit-risk assessment and decision-making. Here, decision-making can be heavily influenced by emotions as risks are usually associated with negative feelings. In consequence, participants were likely to refuse the service [Leschanowsky et al., 2021]. We argue that debiasing strategies as presented in Section 2.6 can promote rational evaluation and help users to make informed decisions in the HMI and specifically in the CUI context. Those strategies can support users in both ways whether their irrational assessment leads to risks outweighing benefits or vice versa. As shown in Figure 3.1 we make use of Dual-Process Theory by assuming that uncertainty can trigger System 2 thinking and could therefore lead to

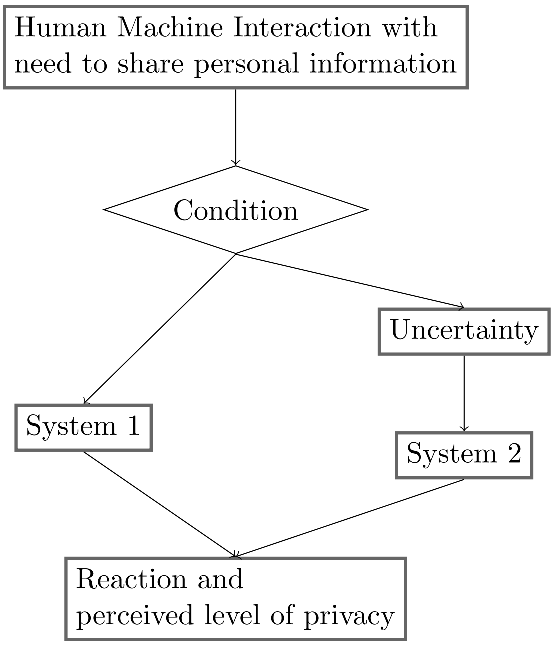

Human Machine Interaction with
need to share personal information
Condition
Uncertainty
System 1
System 2
Reaction and
perceived level of privacy

Figure 3.1: Conceptual Model of System 1 and System 2 behaviour in Human-Machine Interaction (HMI).

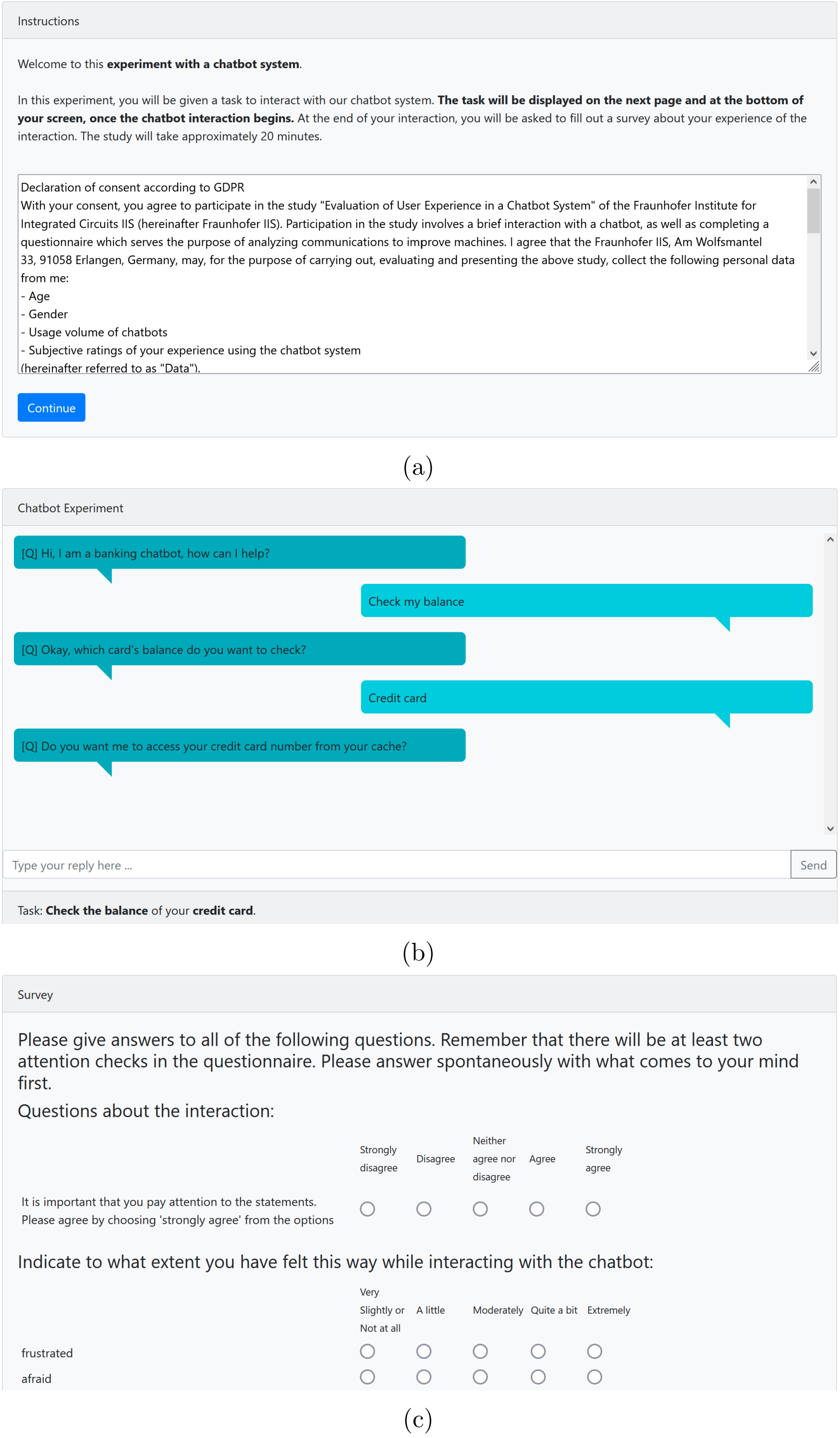
more rational assessment of costs and benefits in the context of CUI. We will use cognitive forcing strategies to bring users into a state of uncertainty while our control conditions will be designed in a way that supports System 1 thinking. Second, our qualitative study revealed that immediate rejection can be subject to peoples' need for protection [Leschanowsky et al., 2021]. While people can easily self-protect themselves in human-to-human interactions e.g. by concealing information, protective mechanisms are not easily accessible in an HMI context. Here, protective mechanisms are often not easily available or require technological knowledge[Leschanowsky et al., 2021]. Therefore, we design the cognitive forcing strategies such that they offer users the possibility to protect their previously shared personal information from further usage. By following the idea of Conversational Privacy – the conversational agent expresses privacy related information and user rights by using dialogue form – these protective strategies will be easily available even for users with little technical knowledge. Importantly, the developed strategies can be deployed independently of context and are suitable for human-machine interactions that require personal information to function correctly. Finally, our developed strategies aim to present users with easily accessible protective strategies and at the same time help them to overcome their cognitive biases regarding their data sharing.

Informed by the qualitative study, this thesis aims to test different debiasing strategies in two different contexts and evaluate whether they are able to trigger uncertainty as an indicator for System 2 thinking. Moreover, we are interested whether the strategies lead to differences in user behaviour or the level of perceived privacy. To reliably assess differences in behaviour we set up an experiment which we will describe in detail in the following sections. Moreover, a survey measuring participants' perceptions and attitudes was found appropriate as surveys provide statistically reliable metrics and allow to measure differences between groups [Olson and Kellogg, 2014]. While we presented the theoretical background how slow thinking could change the evaluation of risks and benefits, evaluating users perceived risks and benefits is out of scope of this thesis.

3.2 Chatbot Language (CBL)

We use *Chatbot Language (CBL)* to implement and test our strategies on a crowd sourcing platform[Brüggemeier and Lalone, 2021]. CBL is based on the high-level language JavaScript and offers the functionality to quickly develop CUIs and have them evaluated on *Amazon Mechanical Turk (MTurk)* [Brüggemeier and Lalone, 2021]. We will refer to the implementation as chatbots which use natural language to interact with a human via text [Brüggemeier and Lalone, 2021]. However, CBL allows experimenters to add audio to the chatbot such that the chatbot's responses are read out loud [Brüggemeier and Lalone, 2021]. Users, on the other hand, need to type in there responses as no automatic speech recognition is provided [Brüggemeier and Lalone, 2021]. Even if investigating only text-based CUI in form of chatbots, we expect the results of the study to be transferable to speech-based conversational agents due to the conversational aspect and similarity between chatbots and speech assistants. Rather than using a *Wizard of Oz* technique, which requires participants to take part in the experiment in a laboratory, crowdsourcing experiments can be conducted remotely [Buhrmester et al., 2011]. In a Wizard of Oz experiment, the intelligent system is simulated by an individual in the background [Brüggemeier and Lalone, 2019]. Surely, participants interacting with the machine are not aware of the human controller and thus the technique is widely used to test and develop user-friendly systems in a controlled environment [Brüggemeier and Lalone, 2019]. From the description one can easily imagine that Wizard of Oz experiments are labor and time-intensive. In contrast, crowdsourcing experiments can be conducted cost and time-efficiently and allow for a more diverse sample of participants as they are not restricted to a location [Buhrmester et al., 2011].

CBL is a rule-based conversational dialogue system. Experimenters can create dialogue flows and use regular expressions to detect keywords and trigger the corresponding response. It also offers the possibility to deal with unknown and unexpected user input and can guide the participant to fulfill the task. This allows experimenters to fully control the interaction and analyze the corresponding transcript later. Moreover, experimenters can design multiple conditions to be tested and CBL then randomly assigns different conditions to workers. We also used CBL to create instruction pages to inform workers about data protection regulations and the task they were going to complete (see Figure 3.2a). Afterwards, workers were exposed to the task and



Instructions

Welcome to this **experiment with a chatbot system**.

In this experiment, you will be given a task to interact with our chatbot system. **The task will be displayed on the next page and at the bottom of your screen, once the chatbot interaction begins.** At the end of your interaction, you will be asked to fill out a survey about your experience of the interaction. The study will take approximately 20 minutes.

Declaration of consent according to GDPR
With your consent, you agree to participate in the study "Evaluation of User Experience in a Chatbot System" of the Fraunhofer Institute for Integrated Circuits IIS (hereinafter Fraunhofer IIS). Participation in the study involves a brief interaction with a chatbot, as well as completing a questionnaire which serves the purpose of analyzing communications to improve machines. I agree that the Fraunhofer IIS, Am Wolfsmantel 33, 91058 Erlangen, Germany, may, for the purpose of carrying out, evaluating and presenting the above study, collect the following personal data from me:
- Age
- Gender
- Usage volume of chatbots
- Subjective ratings of your experience using the chatbot system
(hereinafter referred to as "Data").

(a)

Chatbot Experiment

[Q] Hi, I am a banking chatbot, how can I help?

Check my balance

[Q] Okay, which card's balance do you want to check?

Credit card

[Q] Do you want me to access your credit card number from your cache?

Task: **Check the balance** of your **credit card**.

(b)

Survey

Please give answers to all of the following questions. Remember that there will be at least two attention checks in the questionnaire. Please answer spontaneously with what comes to your mind first.

Questions about the interaction:

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
It is important that you pay attention to the statements. Please agree by choosing 'strongly agree' from the options	○	○	○	○	○

Indicate to what extent you have felt this way while interacting with the chatbot:

	Very Slightly or Not at all	A little	Moderately	Quite a bit	Extremely
frustrated	○	○	○	○	○
afraid	○	○	○	○	○

(c)

Figure 3.2: Implementation of the experiment using Chatbot Language (CBL), (a) shows information on data protection regulations, (b) shows an example of an interaction with a chatbot, (c) shows the survey displayed after the interaction with the chatbot.

one randomly assigned condition (see Figure 3.2b). In the end, they completed a survey on the chatbot interaction (see Figure 3.2c).

3.3 Scenarios

We investigate two chatbot scenarios, a banking chatbot and a chatbot asking for location information of the users. The banking chatbot is similar to the one used in a previous study on Conversational Privacy by [Brüggemeier and Lalone, 2022]. However, in their experiment participants were asked to check the balance of a credit card ending with 5678 while we ask participants for their personal credit card information. This was seen as a major limitation of the study as it does not represent a realistic scenario when users are asked to enter non-personal data. Moreover, it is likely that users perception will differ as they may be more concerned when disclosing real personal information [Brüggemeier and Lalone, 2022]. Similarly, the chatbot asking for the location of the user is designed as a pizza delivery chatbot where users are asked for real location data to get a pizza delivered to their current location. In contrast to financial information, location data has been largely researched in the context of mobile apps and privacy Nudges [Almuhimedi et al., 2015]. Yet, CUIs are closer to human behaviour and conversations and peoples' preferences to protect their location data when using chatbots might differ drastically from when using mobile apps.

In addition, the two scenarios are designed to ask only for information that is required to fulfill the task. This goes back to Nissenbaum's concept of Contextual Integrity and appropriateness of information flow as people are more likely to reveal information when reasons for disclosure are obvious [Nissenbaum, 2010]. Even more, current regulations and design guidelines suggest to gather only information truly necessary [European Commission, 2016, Leschanowsky et al., 2021]. Additionally, the two scenarios aim to cover two distinct levels of information sensitivity. Information sensitivity as shown in the context of the privacy paradox must be considered whenever evaluating peoples' privacy concerns and behaviour. Schomakers et al. analyzed information sensitivity across nations [Schomakers et al., 2019]. They found that financial account numbers and credit card numbers are perceived highly sensitive in the German and the US sample while GPS location data and home address were grouped into a medium sensitive data category with only little difference between the German and the US sample.

Dialogue trees for the two scenarios are provided in Appendix A for the pilot and in Appendix D for the main study.

3.4 Measurements and Survey Design

Our conceptual model is based on the assumption that cognitive forcing strategies will trigger uncertainty and lead to slow thinking (see Figure 3.1). Consequently, we aim to measure the level of uncertainty as well as whether the users find themselves in the state of System 2 thinking. Yet, most of the studies on debiasing techniques focus on user behaviour, e.g. how many errors were committed or whether users downloaded the application, and make conclusions on the underlying thinking process [van der Heijden, 2013, Lambe et al., 2016]. Buçinca et al. additionally looked at users' self-reports on mental demand and system complexity. They found that people who did not get *AI* assistance reported higher mental demand than those exposed to simple explainable AI or the cognitive forcing strategies. This makes sense in the context of explainable AI as they aim to assist users in decision-making tasks. However, mental demand could also be used as a measure of slow thinking activity [Kahneman, 2011, Buçinca et al., 2021]. While we remember that System 2 is usually not engaged whenever we experience Cognitive Ease, the experience of cognitive strain can trigger System 2 thinking [Kahneman, 2011]. The more strain we experience the higher the cognitive load and the more focused we are [Kahneman, 2011]. To assess cognitive strain or load, eye-tracking and pupil dilation has been frequently used, a measure not easily accessible in the context of crowdsourcing [Kahneman, 2011]. Other direct objective measures include brain activity measures or dual-task methodologies, where a form of distraction or a second separate task is presented while participants are required to stay focused [Martin, 2014]. Indirect objective measures refer to physiological approaches, analysis of speech and linguistic features or reaction and completion times [Martin, 2014]. Most of those measurements are not easy to implement when conducting crowdsourcing experiments. However, we did adapt CBL to capture users' reaction and completion times during the interaction with the chatbot. Moreover, subjective measures are still most prominent to investigate cognitive load [Martin, 2014]. Self-reports on mental effort, task difficulty and strain use Likert Scales to measure the experienced load retrospectively [Martin, 2014]. Nonetheless, it remains difficult as the scales tend to be unreliable and inconsistent and the relationship between mental effort and actual load is still unclear [Martin, 2014]. We will investigate mental demand and system complexity as measured by Buçinca et al. and evaluate its reliability and consistency in a pilot study. The survey of the pilot study including the individual items and scales can be found in Appendix B.

Due to the difficulty in assessing cognitive strain and load via a crowdsourcing platform, we will now present possible subjective measures for uncertainty. Moreover, we will introduce the scales on privacy perceptions and usability as well as the control variables used in this experiment.

3.4.1 Uncertainty

While we cannot measure the level of uncertainty directly during the interaction, we can assess uncertainty after the chatbot interaction was carried out. To the best of our knowledge, subjective measures of uncertainty have not yet been used in the context of conversational agents or chatbots. In the pilot study, we will therefore include multiple scales which were shown to be valid in different contexts to assess the level of uncertainty. Based on the reliability and validity tests and the results of the pilot study, we aim to reduce the number of scales for the main experiment on cognitive forcing strategies.

The first scale we want to include, focuses on the relationship between affective states and uncertainty. In a study by Smith and Ellsworth participants were asked to describe experiences related to 15 different emotions. After being interviewed on the experience itself, they answered a questionnaire on different cognitive appraisal dimensions such as pleasantness, control or certainty. The emotions could be described by a systematic variation along those dimensions. Of special interest for us, they found that fear, hope and surprise were significantly influenced by the dimension of certainty. Not only did Smith and Ellsworth find an impact of certainty on fear, hope and surprise, but could show that they could be distinguished from other emotions based on the dimension of certainty. As the study relies on peoples' emotional experiences, we can expect the distinction to be true for varying contexts. However, uncertainty was difficult to be rated consistently compared to the other dimensions and based on the described experiences they found uncertainty to describe two distinct events, either the violation of past expectations or uncertainty about future developments. While surprise was subject to violations of past expectations, hope and fear were related to uncertainty about future developments [Smith and Ellsworth, 1985]. While the study is not directly related to *Human-Computer Interaction (HCI)*, it certainly links cognitive appraisal to experienced emotions which can help us to identify whether participants found themselves in a state of uncertainty during the chatbot interaction. Based on those findings, we will ask participants to rate their affective state on fear and surprise using the *PANAS-X* scale [Watson and Clark, 1994]. The PANAS-X scale is an expanded version of the *Positive and Negative Affect Schedule (PANAS)*, a valid and reliable self-report questionnaire to measure positive and negative affect [Watson and Clark, 1994]. The PANAS-X scale usually consists of 60 items measuring 11 specific effects but single factors can be extracted and measured independently [Watson and Clark, 1994]. We focus on fear and surprise and exclude hope from the study as hope exerted only moderate uncertainty while fear and surprise were associated with maximal uncertainty [Smith and Ellsworth, 1985]. Additionally, to balance positive and negative affective items in the questionnaire, we investigate serenity. Moreover, serenity could provide insights into ease of use of the chatbot. This claim should however be critically examined in the pilot study. We measure fear (consisting of six items: afraid, scared, frightened, nervous, jittery and shaky), serenity (consisting of three items: calm, relaxed and at ease) and surprise (consisting of three items: amazed, surprised and astonished) on a 5-point Likert Scale (1=very slightly or not at all to 5=extremely).

Our dialogue trees are designed such that even if users disclose personal information no service is provided due to apparent technical difficulties or closure of the restaurant (see Appendix A). This is based on the assumption that a positive ending (e.g. providing a fake balance in the banking scenario or telling the user that the pizza is on its way) might lead to uncertainty related to the corresponding outcome and not the disclosure of information itself. The users might be unsure whether the fake credit card balance is actually correct or whether a pizza will be delivered to their location. To control for this effect, we do not provide service in both of the scenarios. However, the negative outcome could leave users with a feeling of frustration. Smith and Ellsworth found that frustration was fairly accompanied by uncertainty. The scenarios people described when being asked about frustration, were often related to failure when people expected to succeed. Here uncertainty was triggered as they usually did not understand causes for their failure [Smith and Ellsworth, 1985]. This means that our negative outcome scenario could also lead to a certain degree of uncertainty which is not related to the disclosure but to the outcome of the scenario. Therefore, we added the item "frustrated" to the beginning of the survey. Frustration was again measured on a 5-point *Likert Scale* (1=very slightly or not at all to 5=extremely).

The second scale related to uncertainty is the *Physicians' Reactions to Uncertainty (PRU)* scale [Gerrity et al., 1995]. The scale consists of four individual constructs namely anxiety due to uncertainty, concern about bad outcomes, reluctance to disclose uncertainty to patients and reluctance to disclose mistakes to physicians and was shown to be valid and reliable in a medical context [Gerrity et al., 1995]. For our survey, we will only include the two scales on emotional reactions (anxiety and concern) and will neglect the coping mechanisms (reluctance to disclose uncertainty or mistakes) as we focus on the experienced level of uncertainty. We rephrased the remaining items to adapt them to the context of chatbot interactions. It is important to notice that while the original scale measures uncertainty in general we aim to measure uncertainty in the specific chatbot interaction [Gerrity et al., 1995]. Therefore, we rephrased the items such that people were not asked to indicate their level of uncertainty on average but rather for this particular experience ("Uncertainty in patient care makes me uneasy" was rephrased to "Uncertainty in the chatbot interaction made me uneasy") [Gerrity et al., 1995]. Moreover, to provide consistency of the rating scales, we assessed the PRU scale on a 5-point Likert Scale rather than on the original 6-point Likert Scale ranging from 1=Strongly Disagree to 5=Strongly Agree [Gerrity et al., 1995]. While a 5-point Likert Scale presents users with an option to indicate neither agreement nor disagreement, the 6-point Likert Scale users need to commit to either the positive or negative side. As a neutral opinion on the questionnaire is legitimate in our case we prefer the 5-point Likert Scale over the 6-point Likert Scale. However, due to our changes of the PRU scale, we cannot take validity and reliability in the context of chatbots for granted and need to re-examine them based on the results of the pilot study. Nevertheless, regarding the anxiety due to uncertainty scale and the concern about bad outcomes scale one could argue that uncertainty leaves people with similar feelings independent of the context. Lastly, one should mention that while feelings are strong indicators for uncertainty to be present, it is questionable if the scales are valid in our contexts and can be applied retrospectively to an interaction. Nevertheless, if discussed critically they can provide valuable insight into how the interventions are perceived by the users and if they indeed trigger uncertainty or associated feelings.

In privacy research uncertainty is seen as an important factor and often considered to be a result of *Information Asymmetry*, the fact that the user has less information available compared to the provider of a service [Al-Natour et al., 2020]. It needs to be considered that privacy risks and privacy uncertainty are two distinct concepts that go back to economic debates on risks and uncertainty [Acquisti and Grossklags, 2005]. Risks are associated with possible outcomes with known probabilities while uncertainty or ambiguity is associated with events where probabilities remain unknown [Acquisti and Grossklags, 2005]. Al-Natour et al. showed that privacy uncertainty has a significant influence on users' intention to use a mobile app and the associated risks. They also distinguished among three subdimensions, namely collection, use and protection, regarding privacy uncertainty and modelled their items based on seller and product uncertainty scales developed by Dimoka et al.. To test their privacy uncertainty and Information Asymmetry measures, Al-Natour et al. conducted a card sorting exercise and found that all constructs are discriminable. Additionally, an overall privacy uncertainty scale was included as well as pre- and post-purchase uncertainty, distinguishing collection, use and protection of personal information. This scale seems to be most relevant as it addresses uncertainty concerning privacy and showed valid and reliable results. However, the scale was newly developed and has so far been tested in mobile contexts only. Moreover, as the user does not purchase a product in our case but shares personal information to use the service we focused on the post-purchase uncertainty and overall privacy uncertainty scale. The pre-purchase uncertainty scale was found not to be useful as we are assessing uncertainty retrospectively after the use of the service. Furthermore, we removed the last three items of the post-purchase protection uncertainty scale as they referred to possible vulnerabilities that could occur in the future after having purchased the application ("I am uncertain if the App Seller will fix information security vulnerabilities that may arise in the future after I start using the App", "I am uncertain if the App Seller will remain vigilant in managing the privacy of my information in the future after I start using the App", "I am concerned if the App Seller will spend enough effort, in the future after I start using the App, in managing the privacy of my information") [Al-Natour et al., 2020]. Similarly to the PRU scale, we rephrased the items to match the context of chatbot interaction. The post-purchase collection uncertainty scale (consisting of four items), the post-purchase use uncertainty scale (consisting of five items), the post-purchase protection uncertainty scale (consisting of three items) and the overall privacy uncertainty scale (consisting of four items) were all measured on a 5-point Likert Scale (1=Strongly Disagree to 5=Strongly Agree).

3.4.2 Privacy Perception and Usability

To measure privacy perceptions and usability we relied on scales that have been used before in privacy research and more specifically in the study by Brüggemeier and Lalone on Conversational Privacy. They showed satisfying reliability and validity results [Brüggemeier and Lalone, 2022]. We additionally added one item to the usability scale which was used in the study on cognitive forcing strategies by Buçinca et al.. The item refers to users motivation of using the chatbot more frequently. Again participants were asked to indicate their agreement on a 5-point Likert Scale (1=Strongly Disagree to 5=Strongly Agree).

3.4.3 Control Variables

We added several control variables which could have an impact on our results. First, we investigate trust in the chatbot as well as in the chatbot provider. Trust can influence users willingness to disclose personal information to a chatbot and can act as an uncertainty mitigator [Al-Natour et al., 2020, Pavlou et al., 2007]. Malhotra et al. investigated trust with respect to online companies on a 7-point Likert Scale. While they did ask for trust in online companies in general, we are interested in users' trust towards the chatbot with respect to the interaction that was carried out. Moreover, trust in the chatbot could be experienced differently from trust in the chatbot provider. Therefore, we rephrase the items accordingly and measure both constructs in the pilot study.

Second, we include a measure of privacy concerns as a control variable. In contrast to the control variable trust, we aim to measure privacy concerns as a trait-like characteristic rather than specific to chatbot interaction. Privacy concerns are known to influence users willingness to disclose personal information in varying contexts such as e-commerce, mobile applications and voice assistants and can influence users perceptions and behaviour [Al-Natour et al., 2020, Javed et al., 2019, Malhotra et al., 2004]. Privacy concerns are frequently measured using the *Internet Users' Information Privacy Concerns (IUIPC)* scale [Malhotra et al., 2004]. The IUIPC is divided into three factors – collection, control and awareness of privacy practices [Malhotra et al., 2004]. Having said that, Groß recently evaluated the original 10-items IUIPC scale in a *Confirmatory Factor Analysis (CFA)*. He confirmed the three-dimensionality but could not confirm the unidimensionality of the control and awareness subscale. Therefore, he reduced the IUIPC to an 8-item scale with improved construct validity and reliability by excluding one item of the control and one item of the awareness subscale [Groß, 2020]. Based on those findings we will measure privacy concerns on a 5-point Likert Scale (1=Strongly Disagree to 5=Strongly

Agree) using their newly validated IUIPC scale.

Third, we include privacy literacy as a control variable. Privacy literacy can be measured subjectively by using the scale provided by Masur. The scale was newly developed and showed internal consistency and reliability. While people might not be good at estimating their knowledge, compared to an objective measure of privacy literacy, self-assessment has the advantage of not being too exhaustive and prolonging the survey unnecessarily [Masur, 2019]. Privacy literacy could provide knowledge and skills to understand chatbot behaviour e.g. the impact of providing access to the system. Moreover, privacy literate people may be less influenced by cognitive functions as they are already well aware of possible privacy violations and therefore do not have to rely on cognitive forcing strategies to activate rational assessment.

Lastly, the level of perceived uncertainty might be affected by cultural differences such as uncertainty avoidance. Trepte et al. found that for people from cultures with high uncertainty avoidance privacy risks are more important than for people from cultures with low uncertainty avoidance. High uncertainty avoidance is strongly correlated with increased privacy concerns and could lead participants to an increased level of uncertainty when they are forced to reveal personal information [Trepte et al., 2017]. Moreover, they might react to cognitive forcing strategies stronger as they are meant to trigger uncertainty. This could impact how people perceive the usability of the system. In their study, Trepte et al. relied on Hofstede's measure of culture on a national level. Hofstede's metric consists of five dimensions namely power distance, uncertainty avoidance, individualism, masculinity and long-term orientation and has been heavily used in social sciences and cross-cultural studies but is also relevant to international business and consumer behaviour [Yoo et al., 2011]. Similarly to Trepte et al., we are not interested in an holistic view on cultural values but focus only on uncertainty avoidance. Moreover, we do not aim to investigate differences among countries but are interested whether an individual's cultural orientation regarding uncertainty has an impact on the perception of cognitive forcing strategies. Furthermore, when using MTurk it is rather difficult to restrict the sample to people with the same cultural background. While MTurk does allow to restrict samples to a specific region e.g. the US or Germany, but the cultural background might be distinct from peoples current location. Therefore, we decide to measure uncertainty avoidance on an individual level rather than on a national level using the *Hofstede's five dimensions of cultural values measured on an individual level (CVSCALE)* developed by Yoo et al..

All control variables explained above were assessed at the end of the survey before asking for demographic information to ensure that no Priming of participants for privacy occurs. In the end, we included questions on gender, age, language proficiency and frequency of use of chatbots [Brüggemeier and Lalone, 2022, Al-Natour et al., 2020].

Finally, we want to point out that while all control variables were originally measured on a 7-point Likert Scale, we plan to use only a 5-point Likert Scale. Therefore, we will investigate the scales reliability and validity in the pilot study and possibly adjust the scales accordingly. By using 5-point Likert Scales, we ensure that constructs included for control are easily comparable to the constructs we are mainly interested in, i.e cognitive state, privacy and usability. Moreover, as we conduct crowdsourcing tests we are keen on keeping the questionnaire clear and simple. Literature suggests that the 5-point Likert Scale is less confusing to be interpreted and simple and easy to use for respondents [Hayes and Hayes, 1992, Neumann, 2016]. Moreoever, it is sufficient for participants to express their views and perceptions [Marton-Williams, 1986]. Further, we adapted CBL to display all corresponding Likert Scale labels as they can guide participants to make final decisions and serve as anchors for participants to use the scales similarly to one another.

3.4.4 Screening Questions

In addition to the items assessing peoples' perceptions and attitudes we included three screening questions in our survey (see Appendix B). In a crowdsourcing experiment, screening questions are included to check the reliability of submitted responses and whether participants paid attention to the survey questions [ITU-T P.808, 2021, Lehigh University, 2022]. While there is – to the best of our knowledge – no standard procedure for including attention checks into crowdsourced HCI experiments, different versions of screening questions are frequently used by researchers in the field [Brüggemeier and Lalone, 2022, Groß, 2020, Javed et al., 2019]. When conducting crowdsourced listening experiments the *International Telecommunication Union Telecommunication Standardization Sector (ITU-T)* gives a recommendation for including screening questions (in the context of listening experiments called gold standard questions) [ITU-T P.808, 2021]. We follow their recommendations and design our screening questions such that crowdworkers can easily provide correct answers when they read the survey items consciously. Moreover, they are not easily recognizable as attention checks because they are visually similar to the other questions in the survey. Additionally, it is recommended that screening questions motivate crowdworkers while making them aware of the importance of their work [ITU-T P.808, 2021]. This is why our screening questions point to the importance of paying attention to the questionnaire items. Our three screening questions are randomly positioned between the other survey items. We plan to exclude participants who do not pass the attention checks from further analysis. Nevertheless, they will not be denied payment as they can still provide valuable insights and feedback.

3.5 Pilot Study

The goal of the pilot study was to compare suitable measures and analyze their reliability and validity in the context of chatbots. As our interactions were supposed to force people to reveal private information we realized that common applications use two distinct ways to gather information. Some let users enter the information necessary while some ask for access to users' devices to gather information. This is especially prominent in the case of location data. Therefore, in the pilot study, we were also interested whether the two strategies are perceived differently by the users and to decide which one to use when testing the cognitive forcing strategies. We will shortly give some background on the two conditions in Section 3.5.1 before presenting the results of the pilot study and their implications for the main study in Sections 3.5.2 and 3.5.3.

3.5.1 Access vs. Enter

Asking users to permit access to personal data has been heavily researched in the context of mobile applications and recently CUIs [Tsavli et al., 2015, Degirmenci et al., 2013, Lentzsch et al., 2021]. To use applications, users are required to give their consent even though they might not be able to fully understand why this data is needed and how it is handled [Tsavli et al., 2015, Beresford et al., 2011]. Even though data regulations such as the GDPR require to only access information necessary for the service to function and advocate for giving users control over their data, application providers also collect data to create profiles and understand users' needs and behaviour [European Commission, 2016, Tsavli et al., 2015]. Those profiles can later be used for the growing stream of market research, advertisement and analytics. Lentzsch et al. found that 23.3% of the Alexa skills ask for sensitive data, such as an individuals' full name, device address or postal code, but do not point out the access to those data types in their privacy policies. Nonetheless, we need to consider that many applications require access to personal information to function correctly and that users appreciate the usage of free applications by giving away more information than needed [Tsavli et al., 2015]. Moreover, developers might lack awareness of privacy measures and simply use default options to access information [Balebako and Cranor, 2014]. The permission model currently used in the mobile application and CUI context seems to lack transparency for both users and developers [Tsavli et al., 2015, Balebako and Cranor, 2014, Lentzsch et al., 2021]. Furthermore, users have no option to negotiate permissions, instead, they need to decide whether to grant permission and use the application or to not use it at all [Beresford et al., 2011]. On some platforms, users can choose between always providing access or being asked for granting access every time they run the application [Beresford et al., 2011]. Because this is not the standard procedure we designed our dialogues in a way that the user can only use the service once access to the information is granted. In the case of the location chatbot, location can be accessed through sensory information of the device. In the banking scenario, we assume the credit card information to be locally stored in the cache which the system wants to access. While we pretend that the system can access the information during the interaction, practically this is not possible. At the end of the questionnaire, we drew participants' attention to the fact that we indeed did not access any information and will only analyse the information provided by the user during the interaction.

To the best of our knowledge, no study exists which investigates the difference of users' perceptions and behaviour between granting access to personal information and entering personal information in the context of chatbot interactions. Nevertheless, e-commerce frequently uses online forms to ask users for personal information when purchasing a product or when having something delivered. While users might be more aware of the disclosure when entering their data, they might still be forced to provide sensitive data not necessarily needed for the service e.g. the need to enter a phone number and an e-mail address when purchasing a product online while an e-mail address would suffice. Similarly to the access condition we give users the choice to either enter information or stating "no" to abort the interaction. Again, participants can only use the service if the information is entered. Additionally, participants can protect themselves in the enter condition by providing misinformation. To be able to assess the number of participants who stated incorrect information, we asked them at the beginning of the survey whether they shared true information or misinformation. While this together with manual analysis of the entered statements will serve as a sufficient indicator of the level of disclosure there is still some risk that users provide misinformation and answer the question with "no misinformation provided" or vice versa.

The corresponding dialogue trees can be found in Appendix A. We aim to test the impact of those disclosure strategies on peoples' behaviour and perception and find a suitable strategy for our main study. Participants can either disclose or refuse to reveal personal information in both of the conditions. However, our cognitive forcing strategies rely to some extent on the user to provide access to the information in the first place. If users choose to terminate the interaction in general in one of the conditions it will not make sense to expose them to the cognitive forcing strategies. For the pilot study we collected data from 200 participants. Therefore, 50 participants per condition and scenario took part in the study, i.e. banking/access, banking/enter, location/access and location/enter. The participants were paid $2 for taking part in the study. As they spent on average 12 minutes to complete the experiment, the average hourly pay calculates to $10.

3.5.2 Results

The location chatbot experiment was conducted on MTurk with 100 participants. The banking chatbot experiment was released one week later, again with 100 workers participating. Users

Demographic and experimental data	Banking	Location
# conditions	2	2
# participants	100	100
# excluded participants	7	16
# accepted participants in the access condition	50	41
# accepted participants in the enter condition	43	43
Mean (SD) age of participants in years	32 (9)	33 (9)
# Gender (female/male/diverse/not provided)	62/31/0/0	37/47/0/0
# Native English speakers (yes/no)	92/1	83/1
# Usage (weekly/monthly/less than once a month/never)	6/41/25/21	16/31/22/15

Table 3.1: Summary of demographic and experimental data for the banking and location scenario in the pilot study

were allowed to participate once in one of the experiments only. However, users with multiple worker identification numbers could have participated more often in the experiment without being recognized. Small errors found in the location experiment were resolved for the banking chatbot experiment such as adding one missing item to the PRU scale or indicating that comments are optional. Participants who failed one or more screening questions were excluded from the analysis. An overview of the total number of participants, the number of participants who passed the screening questions and their distribution across the conditions as well as demographic data is provided in Table 3.1.

In the location scenario, the question on misinformation was not captured correctly. Therefore, the entered addresses were analyzed manually. Once the address information was sufficient to be used for delivery and existed on Google Maps, we denoted that correct information was shared. This error was corrected for the banking scenario such that the data shown in the following for the banking scenario is based on the questionnaire answers.

In Figure 3.3 one can see that most of the participants exposed to the access condition gave access to their data while only roughly 40% shared true information when exposed to the enter condition. Interestingly, this is true for both scenarios. However, people in the enter condition could also state that they did not want to enter personal information but could not specifically state that in the questionnaire. Therefore, manual analysis was necessary to extract the percentage of people who did not want to reveal their address or credit card number in the enter condition. The results are shown in Figure 3.4. Especially interesting is the fact that in the banking scenario participants specifically expressed that they did not want to enter personal information while none did so in the location scenario. Sometimes they did even state that credit card information is personal information as a reason for non-disclosure. While there is a context difference visible, one has to be careful with interpreting the results as many of the classified misinformation provided in the location scenario only contained state information or expressions like “my location” and therefore do not reveal any kind of personal information. Moreover, we accepted all kinds of information

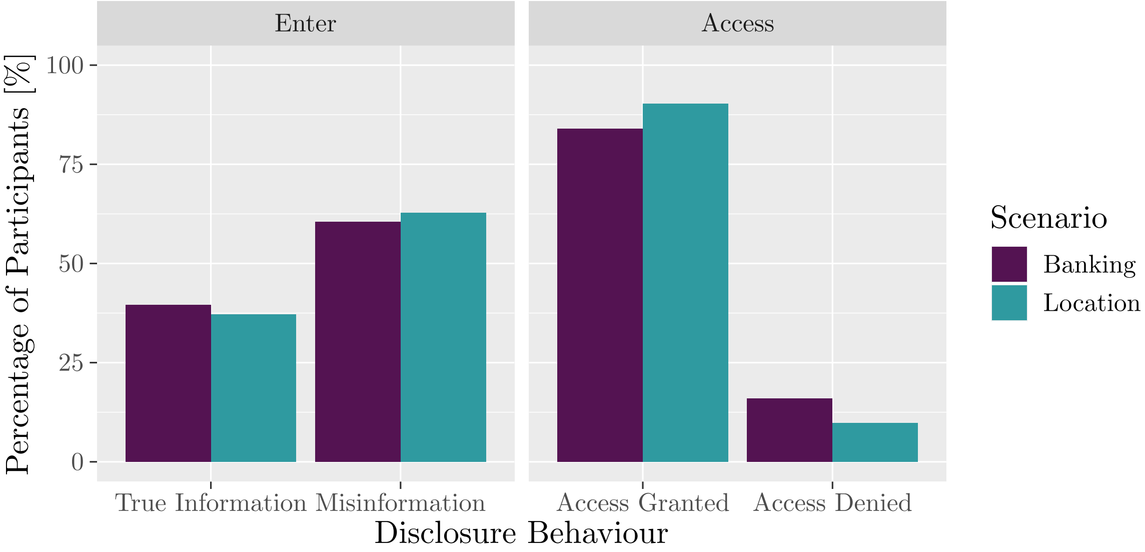


Figure 3.3: Percentage of participants providing misinformation and true information in the enter condition and granting or denying access in the access condition, displayed for both scenarios.

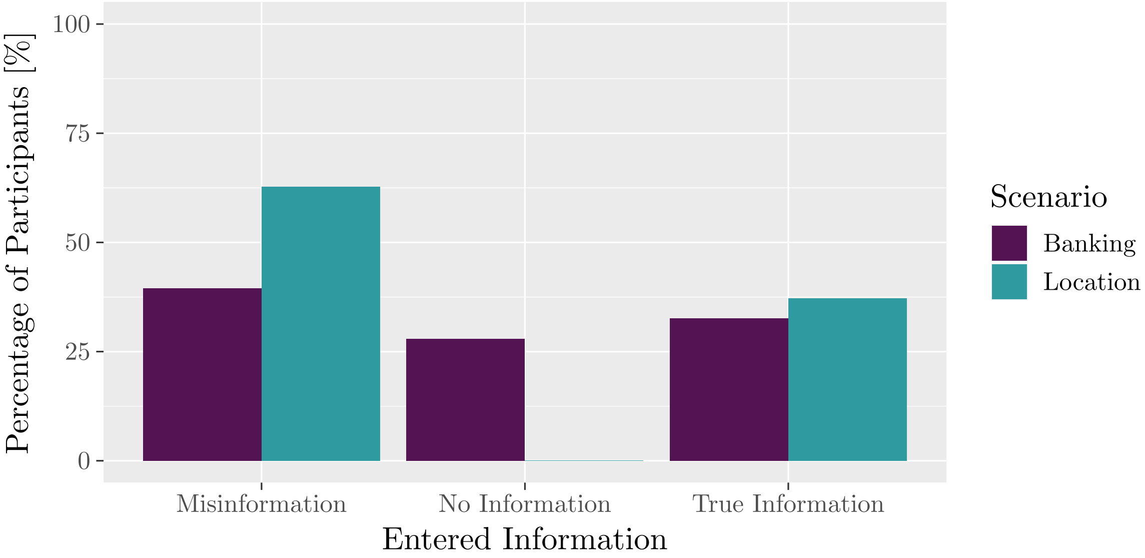


Figure 3.4: Percentage of participants who entered misinformation, no information, or true information, displayed for both scenarios.

in the enter condition in the location scenario while the banking scenario explicitly wanted the user to enter numbers or an expression of refusal. Participants who entered explicitly "no" in the banking scenario answered the question on misinformation differently with the number being almost equally sub-divided towards "misinformation" and "true information".

To check whether a statistical difference is observable regarding the disclosure behaviour of

	Df	Chisq	$Pr(> Chisq)$
(Intercept)	1	13.83	0.0002 ***
scenario	1	0.3518	0.553
condition	1	37.4854	9.21e-10 ***
scenario:condition	1	0.7098	0.3995

Table 3.2: Type III ANOVA of a binary logistic regression model on particpants' disclosure behaviour based on the scenario and condition they were exposed to.

participants, we rely on non-parametric testing. The underlying dataset violates the assumptions of a 2-way factorial *Analysis of Variance (ANOVA)*. A Shapiro-Wilk test reveals non-normality and Levene's test is used to test for equal variances but does not hold [Robertson and Kaptein, 2018]. While *Aligned Rank Transform (ART)* can be used whenever the experiment includes more than one factor with at least two levels and follows a between-subject design, ART was found not to be suitable for dichotomous variables [Luepsen, 2021]. Instead, we conduct binary logistic regression which is useful for multifactor analysis and a sufficient sample size [Robertson and Kaptein, 2018]. In logistic regression, the logistic function

$$P(y=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+...+\beta_p x_p)}} \tag{3.1}$$

is applied to a linear combination of variables $\beta_0 + \beta_1 x_1 + ... + \beta_p x_p$ to model the probability of the outcome of the dependent variable e.g whether participants disclosed information. Thus, instead of interpolating between the classes of the dependent variable i.e. in the binary case between 0 and 1, as done in linear regression, logistic regression outputs meaningful probabilities of being in one category versus being in the reference category. The weights $\beta_0, ..., \beta_p$ in logistic regression can be interpreted as odds ratios. For a binary dependent variable this means that when changing from the reference category to the other category the odds change by a factor of $exp(\beta_j)$ [Molnar, 2019].

After having set up the model, a Type III ANOVA with testing based on a Wald Chi-Square test reveals a statistically significant effect of condition on disclosure behaviour as shown in Table 3.2. Before computing Type III ANOVA in R, we converted scores using sum-to-zero contrast coding and checked for missing cells. As a related measure of effect size, we can compute the odds ratio, which calculates to 3.35 for the condition variable. This means that the odds of disclosure is 3.35 times greater for the access condition than for the enter condition. This translates to a Cohen's d of 0.67 and therefore a medium effect size.

All the scales used in the survey were measured on a 5-item Likert Scale. Such scales can usually be analyzed using ordinal logistic regression [Robertson and Kaptein, 2018]. While binary logistic regression gives the probability of being in one category versus being in the reference category, ordinal logistic regression output probabilities of changing from one level of the ordered variable to the next. Again, we carry out a Type III ANOVA based on the aggregated data with contrasts set appropriately. While there is no effect of scenario or condition on fear, serenity or surprise, we find a significant effect of condition on frustration ($\chi^2(1) = 48.01, p = 4.24e^{-12}$ and an odds ratio of 3.45). This means that for participants exposed to the access condition the odds of being frustrated is 3.45 times that of participants exposed to the enter condition. An explanation for this finding cannot be easily given as we do not see any differences in usability and cognitive demand measures. Moreover, the manual inspection of dialogues and comments does not show differences that would explain this effect either. Possibly, participants give access to data relying on System 1 thinking, that is without critical thinking, and then experience a feeling of frustration when they become aware that they disclosed personal information. Importantly though, we did not expect frustration to have an impact on uncertainty. This seems to be confirmed by the fact that only the factor scenario shows a significant effect on all uncertainty scales (Anxiety due to uncertainty: $\chi^2(1) = 3.88, p = 0.048$, Concern about bad outcomes: $\chi^2(1) = 8.11, p = 0.0044$, Collection Uncertainty: $\chi^2(1) = 7.44, p = 0.006$, Use Uncertainty: $\chi^2(1) = 9.93, p = 0.0016$, Protection Uncertainty: $\chi^2(1) = 5.79, p = 0.016$, Overall Uncertainty: $\chi^2(1) = 5.26, p = 0.022$).

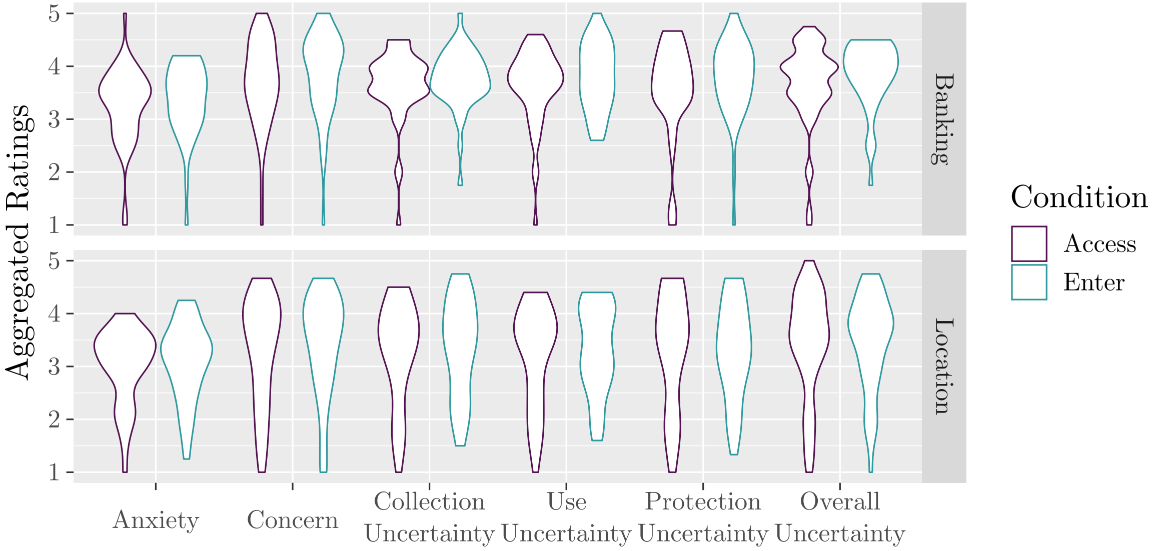


Figure 3.5: Violin plots of aggregated uncertainty ratings for the banking and location scenario and both conditions. Anxiety and Concern refer to the two subscales of the PRU scale [Gerrity et al., 1995], the remaining uncertainty scales were adapted from Al-Natour et al.. The mean was taken over all subscale items to compute the aggregated ratings per participant. Performing ordinal logistic regression and a Type III ANOVA we found that the factor scenario shows a significant effect on all uncertainty scales.

Figure 3.5 shows uncertainty ratings for each sub-scale averaged over the individual items for each participant. In Table 3.3 the calculated odds ratios for the individual uncertainty measures are shown. On average, for participants exposed to the location scenario the odds of expressing uncertainty were 29% lower than for participants in the banking scenario. This difference can be

Uncertainty Sub-scale	Odds Ratio for Scenario
Anxiety due to Uncertainty	0.77
Concern about Bad Outcomes	0.68
Collection Uncertainty	0.69
Use Uncertainty	0.66
Protection Uncertainty	0.72
Overall Uncertainty	0.74

Table 3.3: Calculation of odds ratios for the ordinal logistic regression models on all uncertainty sub-scales. Type III ANOVA showed a significant effect of scenario on uncertainty. The odds ratios are calculated for location versus banking which means that for people being exposed to the location scenario the odds of reporting uncertainty on the anxiety due to uncertainty sub-scale is (1-0.77)*100% = 23% lower than for people in the banking scenario. On average, for participants exposed to the location scenario the odds of expressing uncertainty were 29% lower than for participants in the banking scenario.

due to the varying level of information sensitivity across scenarios, as we expect participants to perceive banking information more sensitive than location information.

No significant effect was found for measures on privacy perception, usability and demand. The two items supposed to capture mental demand to make conclusions whether participants experienced cognitive strain or ease showed high variance in ratings. Therefore, we conclude that two items are not sufficient to capture the underlying construct and plan to remove the scale for the main study. When conducting statistical testing for the control variables, no difference is observable. Important to mention, no difference in ratings is found for trust in the chatbot and trust in the chatbot provider suggesting that one of the scales is sufficient.

3.5.2.1 Reliability and Validity

Reliability and validity are important concepts to determine whether a particular empirical indicator, such as a certain subjective measurement scale, represents an underlying theoretical concept sufficiently [Carmines and Zeller, 1979]. Reliability can be estimated using different methods. In a retest method the same test is carried out multiple times and correlation between the scores at different time instances is obtained as a measure of reliability [Carmines and Zeller, 1979]. Split-Halves method can be used whenever reliability needs to be estimated without multiple testing [Carmines and Zeller, 1979]. Here, the set of items is divided into halves and the correlation between the two sets is calculated [Carmines and Zeller, 1979]. To determine the overall reliability, a correction is applied to the split-half correlation [Carmines and Zeller, 1979]. One of the split-halves methods limitation is that the reliability estimate varies depending on how the items are divided. Therefore, the most prominent way of assessing reliability is the internal consistency method where retesting or splitting up items is not required. One of the

Scale	Cronbach's Alpha (Location)	Cronbach's Alpha (Banking)	Cronbach's Alpha (combined)	Omega Total	Omega Hierarchical
PANAS-X					
Fear	.94	.91	.92	.95	.88
Serenity	.5	.68	.59	.6	.08
Surprise	.76	.51	.66	.7	.62
PRU					
Anxiety due to Uncertainty	.76 (4 items)	.69 (5 items)	.69 (4 items)	.73	.1
Concern about Bad Outcomes	.83	.75	.81	.81	.01
Al-Natour et al. Uncertainty Scales					
Collection Uncertainty	.84	.84	.77	.8	.74
Use Uncertainty	.87	.75	.84	.86	.8
Protection Uncertainty	.84	.74	.81	.81	.02
Overall Uncertainty	.88	.77	.84	.88	.8
Privacy Perception	.74	.83	.8	.87	.75
Usability	.73	.77	.75	.83	.68
Demand	.84	.77	.81	.81	.81
Trust Chatbot	.73	.72	.72	.77	.6
Trust Provider	.75	.79	.77	.82	.67
IUIPC					
IUIPC Control	.43	.4	.41	.41	.41
IUIPC Awareness	.53	.22	.39	.39	.39
IUIPC Collection	.63	.38	.54	.56	.57
IUIPC total	.71	.6	.66	.75	.45
Privacy Literacy	.57	.38	.48	.57	.4
Uncertainty Avoidance	.62	.62	.62	.69	.51

Table 3.4: Reliability analysis for the measures used in the pilot study. Cronbach's Alpha was computed for both scenarios individually while the sub-scale on anxiety due to uncertainty was missing one item in the location scenario, thus only consisting of only four items compared to five items in the banking scenario. Computations of Cronbach's Alpha for the combined dataset, Omega total and Omega hierarchical only consider four items of the anxiety due to uncertainty sub-scale as provided in the location chatbot dataset. Calculations of Cronbach's Alpha, Omega total and Omega hierarchical are performed using the "psych" package in R [Revelle, 2021]. Details can be found in Section 3.5.2.1

most frequently used internal consistency measures is *Cronbach's Alpha.* Cronbach's Alpha can be interpreted as the average of all possible split-half reliability estimates performed on a set of items [McNeish, 2018].

$$\alpha = \frac{N}{N-1}\left(1 - \frac{\sum_{i=1}^{N}\sigma_{Y_i}^2}{\sigma_X^2}\right) \tag{3.2}$$

N refers to the total number of items, $\sigma_{Y_i}^2$ is equal to the individual item variances and σ_X^2 is equal to the variance of the total scale. Cronbach's Alpha takes on values between 0 and 1 with higher values indicating increased consistency and therefore increased reliability. Here, a value of 0.8 is usually considered reliable [Carmines and Zeller, 1979].

While Cronbach's Alpha is still the most popular measure when reporting reliability, researchers have argued that assumptions for the calculation of Cronbach's Alpha are unlikely to be met in real-life scenarios and that alternative reliability estimates should be investigated McNeish [2018]. Therefore, we calculate *Omega* as another measure of internal consistency [Revelle, 2017]. Because Omega serves as an estimate of the general factor saturation of a test an Exploratory Factor Analysis (EFA) is usually performed [McNeish, 2018]. We use the R package "psych" to compute both, Cronbach's Alpha and Omega to assess reliability in the pilot study [Revelle, 2021]. The Omega calculation proposed here differs slightly from the original computation as it uses a more sophisticated variance decomposition [McNeish, 2018]. As mentioned above an EFA is preformed using oblique rotation[Revelle, 2021]. However, Schmid-Leiman rotation is applied to the factor solution yielding a bifactor model with one general factor and several minor factors [McNeish, 2018]. The variance of a test score can thus be divided into one part related to the general factor, one part related to a set of group factors, one part related to specific factors for individual items and random error [Revelle, 2021]. Omega total can then be computed by squaring the loadings of the general factor λ_{gi} and the ones of the group factors λ_{fi} with k being the total number of items, F being the total number of group factors and k_f referring to the number of items relevant to the corresponding group factor [McNeish, 2018]. V_X refers to the total variance after rotation [McNeish, 2018].

$$w_t = \frac{(\sum_{i=1}^{k}\lambda_{gi})^2 + (\sum_{f=1}^{F}\sum_{i=1}^{k_f}\lambda_{fi})^2}{V_X} \tag{3.3}$$

While Omega total estimates give an overall reliability estimate due to a general factor and possible lower-level factors, Omega hierarchical gives an estimate for the variance due to the general factor only [Revelle, 2017]. Thus, the equation simplifies to

$$w_h = \frac{(\sum_{i=1}^{k} \lambda_{gi})^2}{V_X}. \tag{3.4}$$

Zinbarg et al. compared reliability estimates and found that Omega hierarchical performs best. However, they acknowledged that Omega total is preferred when one is interested in the scales variance due to all common factors and not only in the variance due to the general factor [Zinbarg et al., 2005]. Thus, it make sense to investigate those two additional reliability measures.

For the analysis, we use only data of participants who passed the screening questions and computed Cronbach's Alpha for the results of the location chatbot, the banking chatbot and lastly for the combination of them. From Table 3.4 it can be seen that the PANAS-X scale for "Fear" and the uncertainty scales are reliable. However, this is not the case when it comes to the PANAS-X scale for "Serenity" and "Surprise" and the adopted scales of privacy concern, privacy literacy, and uncertainty avoidance. Two remarks can be made related to the PRU scale and the IUIPC collection sub-scale with respect to Cronbach's Alpha for the combined scenarios. For the PRU scale it should be noted that if the item on "being comfortable" was dropped, Cronbach's Alpha would increase to 0.79. Furthermore, the specific item correlates only with 0.19 and there is a decrease for the correlation coefficient visible suggesting to exclude the item for further studies. In the case of the IUIPC collection sub-scale it should be considered that if the item on "It usually bothers me when being asked for personal information" was dropped, Cronbach's Alpha would increase to 0.6. As above, the specific item correlates with only 0.23 and the average correlation coefficient also decreases in this case.

All of the calculations for Omega were performed with the default three-factor assumption except for the 2-item scale measuring demand and the individual IUIPC sub-scales where only one factor was used. In this case, the two Omega estimates are naturally the same. Here, one needs to pay attention to the two PRU sub-scales, where Omega hierarchical is low. As Omega hierarchical can be interpreted as the precision with which a score assesses a single construct, we conclude that not the general factor but rather related group factors of the two sub-scales are the source of the variance in the ratings [Watkins, 2017]. This is different to the other uncertainty sub-scales where Omega hierarchical mostly calculates to a higher value such that the general factor assesses the underlying construct precisely.

When we talk about validity, we mostly refer to *Construct Validity* which is central when measuring abstract theoretical concepts [Carmines and Zeller, 1979]. "Construct Validation focuses on the extent to which a measure performs in accordance with theoretical expectations" and it is thus not simply confirmed by a single measurement outcome but requires multiple consistent findings across researchers and studies [Carmines and Zeller, 1979]. Therefore, a lack of Construct Validity could also be subject to an incorrect theoretical framework, inappropriate experimental set-up or unreliability of other variables [Carmines and Zeller, 1979]. Nevertheless, we can apply factor analysis to assess Construct Validity and find factors that underlie the data set. EFA only shows patterns of correlations among the questionnaire items and it is thus important to correctly interpret the outcome e.g. whether the scale truly reflects multiple underlying constructs or whether the difference is subject to wording and phrasing of the items [Carmines and Zeller, 1979]. Before carrying out the EFA, we conduct *Kaiser-Meyer-Okin (KMO)* measure of sampling adequacy to check if the number of participants is sufficient to carry out a reliable factor analysis [Hof, 2012]. The scale on privacy literacy shows a mediocre result with a value of 0.59 while all other scales show values between 0.7 and 0.95 indicating good to superb sampling adequacy [Hof, 2012].

Factor analysis was conducted using R's function "factanal" on the combined data set of location and banking chatbot and only participants who passed the attention checks. When performing factor analysis, one hypothesizes that the number of factors chosen for the analysis fits the data well [Hof, 2012]. In case of rejection, more factors should be considered [Hof, 2012]. According to Hof, if variables hold loadings lower than 0.3 they are considered to have a non-significant impact on a factor and can be ignored. We used oblique rotation ("promax") as we assume the factors to correlate within one scale in most of the cases [Hof, 2012]. Orthogonal rotation ("varimax") was only used whenever correlation was not sufficient. The detailed results of the EFA can be found in Appendix C.

A factor analysis using an orthogonal rotation of the PANAS-X scale showed that three factors are sufficient at an α-level of 0.01. However, the third factor makes up only 2.5% of the variance in the data set and is highly related to the item "nervous". We can conclude that the construct of "Serenity" and especially "Fear" is represented by the items quite well. Yet, the underlying construct of "Surprise" is not captured sufficiently well in our data.

When analyzing the PRU scale, we found that items related to the anxiety due to uncertainty scale make up the first factor and items related to the concern about bad outcomes scale make up the second factor. Only the item on "being comfortable" does not fit well, similarly to what we saw when performing the reliability analysis.

EFA of the remaining uncertainty subscales (collection, use, protection and overall uncertainty) reveals that three factors are sufficient at an α-level of 0.05 while a model with two factors is rejected at all common α-levels. When carrying out an EFA of the combination of the PRU scale and the uncertainty scales taken from Al-Natour et al., we found a model with four factors has low correlation coefficients and the third and fourth factor explain only 5.5% and 4.3% of the variance in the dataset. However, a model of three factors is rejected at an α-level of 0.05 but can be excepted at an α-level of 0.01. This suggests that the scales measure similar constructs and thus, further usage of only one of the scales is preferred. When combining the PRU scale and the PANAS-X scale and conducting a factor analysis using orthogonal rotation, five to six factors are at least necessary to explain the variance in the data set. This indicates that the two scales measure different underlying constructs while the two uncertainty scales measure similar constructs.

The null hypotheses for the two factor analyses on perception of privacy and usability cannot be rejected on an α-level of 0.05 ($\chi^2_{Privacy}(9) = 15.95, p = 0.068$; $\chi^2_{Usability}(14) = 12.63, p = 0.556$). Thus one factor for the scale measuring perception of privacy as well as for the one measuring usability fit the data well.

When exploring Construct Validity for the trust scales, we combined the ratings of the two sub-scales measuring trust in the chatbot and trust in the chatbot provider. Interestingly, a factor analysis with two or three factors and oblique rotation was found to not be sufficient. Only a model with four factors cannot be rejected at an α-level of 0.05. When splitting up the scales for the analysis, one factor describes the variance in the individual data sets sufficiently. Possibly, one could therefore conclude that using one of the scales will be sufficient to measure the underlying construct of trust. While whenever both scales are used, the underlying construct is more complex and would need further analysis.

Based on the outcome of the factor analysis of the IUIPC scale, we can hardly confirm the three-dimensionality of the scale divided into control, awareness and collection. Nevertheless, when dropping the fifth item as suggested by the reliability analysis, a model with one factor cannot be rejected at an α-level of 0.05. As we do not rely on the three-dimensionality aspect of the IUIPC scale, we can conclude that the modified scale captures the underlying construct of general privacy concerns sufficiently well.

The factor analysis on the privacy literacy scale shows that one factor is sufficient on an α-level of 0.01 ($\chi^2(2) = 7.4, p = 0.0247$) which shows that the scale could be enhanced to fulfill uni-dimensionality criteria. Nevertheless, it shows sufficient validity to be included as a control variable into the main study. For the scale on uncertainty avoidance one factor suffices and fits the data well ($\chi^2(5) = 8.25, p = 0.143$).

3.5.2.2 Reaction and Completion Times

Additionally to the analysis of participants' behaviour and ratings on the survey questions, we had a closer look at reaction and completion times. Participants spent roughly 5.7 minutes on average on the interaction with the location chatbot and 4.5 minutes with the banking bot. In the location scenario, they spent 6.6 minutes on average on the survey, while some took up to 35 minutes to complete the survey. In the banking scenario, they spent 7.1 minutes on average on the survey, again some took up to 30 minutes to complete the whole survey.

We did not only capture timestamps when the interaction started and ended but also whenever the chatbot sent responses (intermediate time stamps t_c) or participants started typing (type stamps t_p). In the case of the location scenario, we had to exclude 36 sets from the analysis as type stamps were either not captured for all user inputs or there were more type stamps than user interactions. Similarly, we excluded 22 sets for the banking scenario. Although we captured additional browser and device information when conducting the banking chatbot experiment, we could not identify why several type stamps were not saved correctly. We compared the time participants took to answer to the question to disclose information and the average time they needed to reply to the remaining questions asked by the chatbot according to the following formula:

$$\Delta t = (t_{pN} - t_{cN}) - \frac{1}{N-1}\sum_{i=1}^{N-1}(t_{pi} - t_{ci}) \tag{3.5}$$

N denotes the number of interactions an individual had with the chatbot in the experiment. The difference $(t_{pi} - t_{ci}) > 0$ as the chatbot sends questions first before participants respond to them. The average is computed for all interactions except the Nth one, as the question asked to disclose personal information was the last one participants replied to in the interaction (see Appendix A). The time difference Δt, the difference between this baseline and the time taken to respond to the condition-specific question, was computed for the sets available. To check whether a difference is due to the time taken to search for the correct information, we differentiated between the ones who entered misinformation and correct information. Participants who said that they do not want to provide information in the dialogue with the banking chatbot are distributed similarly across groups that indicated having provided misinformation or true information respectively.

The results are shown in Figure 3.6 for the two scenarios and conditions, in the enter condition subdivided into participants entering misinformation and the ones entering correct information. Positive values indicate that participants took more time to respond to the chatbot's prompt to enter or access data in the condition compared to the interactions before while negative values indicate that participants took less time to respond to those prompts than to prior prompts (see Appendix A). This is especially interesting as it shows, that in the case of the access scenario participants took less time before typing their response while when entering they needed more time to think before starting to write. This is true whether they provided correct information or misinformation. We carried out a Kruskal-Wallis test for the individual scenarios as the time differences did not fulfill the assumptions of parametric testing. Kruskal-Wallis extends to factors with more than two levels and is therefore suitable to compare the differences between the three conditions (access, enter misinformation and enter true information) [Robertson and Kaptein, 2018]. The results for the location scenario and the banking scenario indicate differences among groups ($\chi^2_{Location}(2) = 9.76, p = 0.008$ and $\chi^2_{Banking}(2) = 8.01, p = 0.018$). We further analyzed the group differences by conducting Wilcoxon rank-sum tests for pairwise comparison between the conditions. To account for multiple comparison and to adjust the α-level accordingly, we

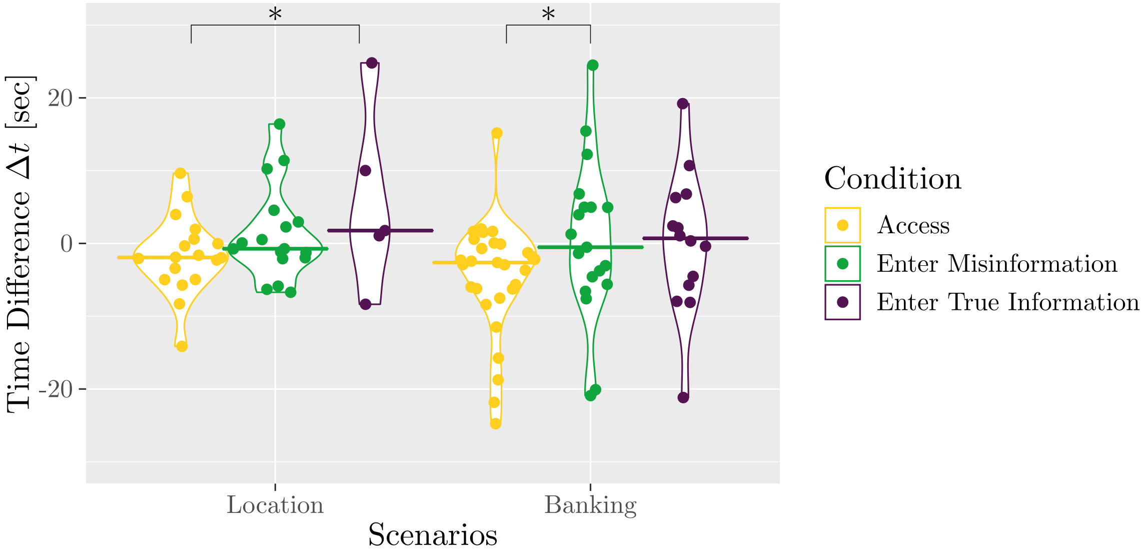


Figure 3.6: Time Differences between the time taken to respond to the condition question and the baseline provided by the average time difference over all interactions before for individual users (see Equation 3.5). 17 outliers are not visible in this figure in order to allow a close-up comparison of group differences. Median values are indicated by colored lines. Condition differences highlighted with an asterisk have a p-value lower than 0.025 in the pairwise comparison of conditions in each scenario as we apply Bonferroni correction to account for multiple testing.

apply the Bonferroni method. Pairwise comparison suggests a significant difference between the access and the enter condition when people entered correct information in the location scenario ($p = 0.0084$) and between the access and the enter conditions when people entered misinformation in the banking scenario ($p = 0.022$). No difference was found between the two enter conditions in both scenarios. Given the analysis of the reaction times we could conclude that people are more likely to use mental shortcuts when answering in the access condition. Contrarily, the enter condition seems to trigger slow thinking. This tends to be the case whether participants shared correct or incorrect information and seems to be independent of the information they need to provide.

3.5.3 Discussion and Implications

We assumed that the ambiguities regarding the systems' capabilities to access information will lead to refusal of granting access. Participants might be unsure whether the system restricts access to the information required or whether additional data is collected. However, we found that the opposite is true. While only 40% revealed true information when they were asked to enter personal data, roughly 80% granted access to their data. Despite the fact that we worked with peoples' real data in the study setup, it can be questioned whether participants believed that the chatbot would be indeed capable to access the information. It is important to keep in mind that reaction times significantly differed between conditions with participants showing faster reaction times when exposed to the access condition. If participants were certain, that their data can not be accessed by us, their immediate responses could be interpreted as being care free, because they knew that their data is save. However, accessing location data is common practice when interacting with technology, including interactions with CUIs. Thus, we believe that participants may not have been certain that we will not access their data. Hence, their fast responses, mostly allowing data access, may not be a sign of well-informed and care-free decision making, but rather careless decision making, potentially in System 1.

One could argue that the differences in reaction times could be due to the fact that information e.g. a credit card number needs to be looked up in the enter condition. However, we assume that this would only be true for the banking scenario as participants are likely to know their current address and that there is no need to look up this information. Instead, our results show differences across scenarios. Moreover, reaction times differ between participants' granting access and those who shared misinformation. As entering misinformation does not require participants to look up information, we would not expect to see an increase in waiting time. Thus, we conclude that the time differences can not be explained by looking up information. Instead, the enter condition seems to be more likely to trigger slow and effortful thinking.

In addition, we found that people who give access to their data are more likely to express frustration than those entering personal information. We assume that rather fast and carelessly given consent to have the system access the data can result in a feeling of frustration. Contrarily, people who took more time to think about their decision, were less likely to provide data and less frustrated after the interaction with the chatbot. Furthermore, our results show that participants feel more uncertain when being exposed to the banking scenario. This could be a result of different levels of information sensitivity associated with the two scenarios, as credit card numbers are considered to be more sensitive than location or home address. Thus, participants in the banking scenario experience a feeling of uncertainty after the interaction with the chatbot. Moreover, we neither found privacy perceptions nor usability to be influenced by condition or scenario.

We aimed to find a suitable strategy for people to reveal personal information for our main study. This is because in our main study we implement cognitive forcing strategies that rely to some extent on participants volunteering to share information. Our results show, that significantly more people share information in the access rather than in the enter condition, independent of the scenario. Thus, we decide to use the access condition within our main study. We believe that the enter condition might already serve as a cognitive forcing strategy by slowing down the decision to disclose personal data. This might explain why participants when asked to enter personal information take more time before responding, are less likely to disclose and less frustrated after the interaction with the chatbot.

Another goal of the pilot study was to test the individual scales on their reliability and validity and reduce the size of the questionnaire for the follow-up experiment. The PANAS-X scale was supposed to serve as an indicator for the dimension of “Uncertainty”. However, our results did not confirm this as only the single item “frustrated” was affected. Moreover, the reliability and validity tests showed only promising results for the underlying construct of “Fear”. While fear is related to the uncertainty of future events, surprise is related to violations of past expectations [Smith and Ellsworth, 1985]. Therefore, we conclude to keep the items on “Fear” and “Frustration” but exclude the items on “Serenity” and “Surprise” for future experiments.

Further, we aimed to reduce the number of uncertainty scales used in the questionnaire. While all of them were similarly significantly affected by scenario, the PRU scale showed worse reliability and validity results than the other scales. Because of this, we exclude the PRU scale from further surveys and only include the collection, use, protection and overall uncertainty scale. Although we could not confirm the four-dimensionality of the scale, we argue that reducing the scale further could possibly lead to lower reliability, validity and sensitivity. As we do not have enough experience in using the scale in the context of chatbots, we will keep the whole scale as tested in the pilot study for the following experiment.

In addition, we remove the two items assessing mental demand as ratings showed high variance and were not useful in assessing differences based on scenario or condition. As the ratings on the two trust scales (trust in chatbot, trust in chatbot provider) did not differ, we agree upon using the scale, that measures trust in the chatbot for future experiments only. Regarding the IUIPC scale, reliability analysis suggested removing the fifth item to increase Cronbach’s Alpha. While we could not confirm the three-dimensionality of the scale, it might still be useful to measure general privacy concerns as a control variable.

Lastly, we were able to identify a few usability problems with the implementations of chatbots used in the pilot study by analysing the individual transcripts. Even though we did not receive negative feedback on the chatbot interactions we took care of several problems for the main study. Major changes included acceptance of credit card types such as Visa or American Express in the banking scenario and accepting “ok” or “okay” in addition to “yes” or “sure” as an answer.

In our power analysis we found that metrics differ in the number of participants required for an appropriately powered analysis. For example we found that differences in the uncertainty scales require between 400 and 750 participants, whereas differences in privacy perception and usability scales require more than 1000 participants. Scales differed in the necessary participants between 400 and 1600. As we want to avoid both under- and over-powered comparisons, we decided to continue to assign 50 participants per condition and scenario. This results in a total number of 500 participants which lies in the range of participant numbers that we found for appropriately powered comparisons.

3.6 Main Study

After having found a good strategy to make people disclose personal information during the chatbot interaction, i.e. the access condition, we now set out to test several cognitive forcing strategies in the main study. We will investigate in which way cognitive forcing strategies influence users' perceptions and behaviour while relying on the revised, reliable and valid questionnaire. In Section 2.6, we gave an overview of different nudging and debiasing strategies. As indicated there, some of the strategies used in this thesis might not fall into the category of cognitive forcing strategies in other disciplines. Nevertheless, they are all set up to overcome biased thinking and trigger rational assessment of risks possibly leading to a more privacy-preserving behaviour. At the same time, they are designed in a way to not restrict users but leave them with all choices possible. We will now introduce the cognitive forcing strategies used in this thesis as well as the corresponding control conditions in Section 3.6.1. Afterwards, we will present our hypotheses based on the theoretical background and conceptual framework in Section 3.6.2 and show results of the main study in Section 3.6.3.

3.6.1 Control Conditions and Cognitive Forcing Strategies

Our cognitive forcing strategies are designed such that they include the aspect of Conversational Privacy. In particular, we focus on presenting users with an offer to delete their data. Brüggemeier and Lalone found that this significantly affected peoples' perception of privacy and their behaviour in a banking scenario. Moreover, companies like Amazon or Google recently started to let users delete their recordings via simple voice commands [Teague, 2021]. This is especially important in the light of data regulations where users are given the right to have their data deleted at any time [European Commission, 2016]. Again we test our cognitive forcing strategies in the two scenarios, location and banking. We implement the same dialogue trees as were used in the access condition of our pilot study. After granting or denying access to their data, participants are exposed to either one of the three cognitive forcing strategies or to one of the two control conditions. The corresponding dialogue trees can be found in Appendix D for the banking and the location chatbot followed by the five conditions. Additionally, the questions, user response options and their meaning for each of the five conditions are shown in Table 3.5. Both of the scenarios include the five conditions and thus based on the power analysis we collected data from ten groups á 50 participants. The participants were paid $ 1.40 for taking part in the study. Because participants took roughly 11 minutes on average to complete the experiment, this translates to an average hourly pay of $7.60.

The first control condition is taken from a previous study on Conversational Privacy and supposed to resemble a common interaction with a chatbot nowadays, that may end with the chatbot saying "Is there anything else I can do for you?". Moreover, this control condition can serve

Condition	Question	User Response Options	Meaning of User Responses
Control 1	Is there anything else I can help you with?	Yes/No	Help/No Help
Control 2	I will save your data for future interactions now, okay?	Yes/No	Save/Delete
Slow Down	I will save your data for future interactions now, okay? I'll give you 20 seconds to think about it.	Yes/No	Save/Delete
Alternative	Do you want me to delete your data from this interaction or have it saved for future interactions?	Delete/Save	Delete/Save
Reconsider	Do you want me to delete your data from this interaction now?	Yes/No	Delete/Save

Table 3.5: Question, user response options and their meaning for the five conditions, including two control conditions and three cognitive forcing conditions. The corresponding dialogue trees can be found in Appendix D.

as a baseline of "yes" and "no" answers. In addition, we add another control condition that is more specific to the behaviour we want to investigate, i.e. permanent storage or deletion of data. The second control condition was designed to give people the opportunity to state whether they wanted to have their data saved or not saved while at the same time nudging them into a more disclosing behaviour. Similar strategies, called dark patterns are used in interface design e.g. when designing cookie banners [Bermejo Fernandez et al., 2021]. Here, interfaces are designed such that individuals make decisions which favour the data collectors rather than themselves [Bermejo Fernandez et al., 2021]. In the case of cookie banners colours are chosen in a way that users are more likely to click on "save" or "accept" rather than on "settings" or "'deny" [Bermejo Fernandez et al., 2021]. While we expect people to think of an implicit deletion of their data once they answer with "no" in the second control question, we have no influence on their mental models. Nevertheless, the second control condition is meant to serve as a baseline of save and deletion requests.

We implement three different cognitive forcing strategies, some of them similar to what was used in other studies on privacy or explainable AI. The cognitive forcing strategies are applied at the time the decision takes place or shortly after during the chatbot interaction to disrupt heuristic reasoning or reconsider biased decisions. They are designed to make users engage in System 2 thinking. While this does not necessarily mean that informed decision making takes place it can nevertheless support the process of a rational cost-benefit analysis. Based on the second control condition we implement a slow down condition. Here people are given 20 seconds time to think about their response before it is sent to the chatbot. Importantly, the interaction can not be terminated earlier but participants have to wait for 20 seconds until their answer is processed. Similarly to the timer Nudge investigated by Wang et al. or the delay condition tested by Buçinca et al. our slow-down condition is designed to trigger slow thinking and give users the time to reflect and possibly reconsider their decision. 20 seconds waiting time is chosen based on the two studies with 30 seconds waiting time before the AI suggestion was shown and 20 seconds visual delay before the Facebook post was published [Wang et al., 2013, Buçinca et al., 2021]. Our second strategy is based on cognitive forcing strategies applied in the medical context where participants are asked to consider alternatives instead of deciding intuitively or without instructions [Lambe et al., 2016]. While we do not know peoples' underlying mental models on how their data is collected, processed and stored when interacting with the chatbot, we use an alternative question to confront people with the option of having their data saved or deleted. Importantly, users can only proceed by answering "delete" or "save" and therefore, explicitly need to state their decision compared to simpler "yes" or "no" answers. Lastly, we use a cognitive forcing strategy where users can reconsider their decision of having disclosed personal information by asking them whether they want their data to be erased. The strategy of reconsideration is also known from the medical context and has shown improved accuracy on erroneous decisions [Lambe et al., 2016]. The same question on deletion was used by Brüggemeier and Lalone in their research on Conversational Privacy where they found significant differences in privacy perceptions when users were given control over their data.

In Table 3.6 we give a short overview of the metrics and variables included in the main study and their corresponding levels. We are especially interested in the main effects of scenario and condition on the dependent variables, i.e. users' behaviour and perceptions.

3.6.2 Hypotheses

Given the theoretical background on cognitive forcing strategies and first experimental results of the pilot study, we hypothesize that participants when exposed to our strategies will engage more in privacy-preserving behaviour than when being exposed to the control conditions. Particularly, they will delete their data more often in conditions when slow thinking is triggered.

Hypothesis 1 *Participants will delete data more often in conditions with cognitive forcing strategies than in control conditions*

As we have seen in Section 2.5, System 2 thinking is engaged in states of doubt, surprise and uncertainty. Therefore, the goal is to bring users into a controlled state of uncertainty during the chatbot interaction which should lead to more rational risk-benefit assessment and thus changes in decision-making. We investigate uncertainty by assessing the associated feeling of "Fear" and by using dedicated uncertainty scales. Accordingly, we state that participants will report higher levels of fear and uncertainty when being exposed to the cognitive forcing strategies. As seen

Variable	Levels
Independent Variables	
Scenario	Banking, Location
Condition	Control 1, Control 2, Slow Down, Alternative, Reconsider
Dependent Variables	
Behaviour	Binary Response (“Yes”/“No”, “Save”/“Delete”)
Frustration	5-point Likert-Scale
Fear	5-point Likert-Scale
Collection Uncertainty	5-point Likert-Scale
Use Uncertainty	5-point Likert-Scale
Protection Uncertainty	5-point Likert-Scale
Overall Uncertainty	5-point Likert-Scale
Privacy Perception	5-point Likert-Scale
Usability	5-point Likert-Scale
Control Variables	
Trust in the Chatbot	5-point Likert-Scale
Privacy Concerns	5-point Likert-Scale
Privacy Literacy	5-point Likert-Scale
Uncertainty Avoidance	5-point Likert-Scale
Usage	Categorical Response (“not at all”/“less than once a month”/“2-4 times a month”/“more than once a week”)

Table 3.6: Overview of variables assessed and manipulated in the main study. Further details on the variables and their individual items measured in the questionnaire can be found in Appendix E. We are especially interested in the main and possible interaction effects of scenario and condition on the dependent variables, i.e. users’ behaviour and perceptions.

in the pilot study, participants showed longer reaction times when entering information than when granting access to their data. This was true whether people shared true information or misinformation. We concluded that the enter condition could possibly be seen as a cognitive forcing strategy which brings people into a state of cognitive strain and slower thinking. Generally, cognitive forcing strategies are supposed making the user to think more thoroughly about costs and benefits. Because analytical thinking is usually slower than relying on heuristics, we can assume that engaging in cognitive forcing strategies will lead to longer reaction times.

Hypothesis 2a *Participants will report higher levels of "Fear" when exposed to cognitive forcing strategies than to control conditions.*

Hypothesis 2b *Participants will report higher levels of uncertainty when exposed to cognitive forcing strategies than to control conditions.*

Hypothesis 2c *Participants will show longer reaction times when exposed to cognitive forcing strategies than to control conditions*

Previous research showed that similar questions on offers to delete shared data increased the level of perceived privacy and security [Brüggemeier and Lalone, 2022]. Similarly, we hypothesize that people explicitly asked whether they want to delete their data or given more time to think about saving their data, will report higher levels of privacy.

Hypothesis 3 *Participants will report higher levels of perceived privacy when exposed to cognitive forcing strategies than to control conditions.*

An important aspect when designing cognitive forcing strategies in the context of conversational AI, is the usability of these strategies. Buçinca et al. found that the implemented functions reduced the usability and acceptability of the system. As the interventions are meant to trigger analytical thinking and therefore put users into a state of cognitive strain, the chatbot interaction might be perceived as more exhausting and less usable for future tasks.

Hypothesis 4 *Participants will report lower levels of usability when exposed to cognitive forcing strategies than to control conditions.*

3.6.3 Results

Based on results from our power analysis, we ran both scenarios with 250 participants each. This constitutes 50 workers being exposed to each condition. Experimental and demographic data can be found in Table 3.7. Although we had to exclude almost a fifth of participants in the location scenario due to incorrectly answered screening questions or the chatbot asking two cognitive

Demographic and experimental data	Banking	Location
# conditions	5	5
# participants	250	250
# excluded participants	26	56
# accepted participants in the different conditions (Control 1/ Control 2/ Slow Down/ Alternative/ Reconsider)	48/50/41/44/41	43/36/34/35/45
# accepted participants' disclosure behaviour (Granting Access/ Denying Access)	184/40 (82%/18%)	166/28 (86%/14%)
Mean (SD) age of workers in years	34 (10)	35 (10)
# Gender (female/male/diverse/not provided)	87/137/0/0	79/114/0/0
# Native English speakers (yes/no)	219/5	183/10
# Usage (weekly/monthly/less than once a month/never)	44/66/60/54	21/65/76/31

Table 3.7: Summary of demographic and experimental data for the banking and location scenario in the main study

forcing strategy questions at the same time, participants are still distributed rather equally over the five conditions (see Table 3.7 for the number of participants in the different conditions). We see a similar pattern of disclosure behavior as we saw in the pilot study. In both scenarios more than 80% of the workers granted access to their data before they were exposed to one of the control conditions or cognitive forcing strategies.

3.6.3.1 Analysis of Peoples' Behaviour

Similar to the pilot study, we exclude participants who did not pass at least one of the screening questions. First, we investigate response behaviour of participants exposed to individual conditions. The first control condition serves as a baseline for ' ´yes"-"no" response behaviour while the second control condition serves as a baseline for a Nudge to saving data as a default. Therefore, we expect distinct response behaviour between the two control conditions, although both can be answered using the same expressions i.e. "yes" and "no". Notably, only two out of three cognitive forcing conditions allow users to respond with a simple "yes" or "no". The alternative condition question requires participants to respond using either "save" or "delete" to express their opinions. For now, we focus solely on "yes"-"no"-decisions and therefore exclude the alternative condition from the following analysis (see Table 3.5 for user response options and meaning of the condition questions). In Figure 3.7 we show peoples' response behaviour for the first control, the second control, the slow down and the reconsider condition. One needs to keep in mind that responding to the conditions with "yes" has distinct meanings depending on the condition question. Unrelated to data processing, "yes" in the first control condition refers to asking for additional help. Importantly, participants answering "yes" in the second control or

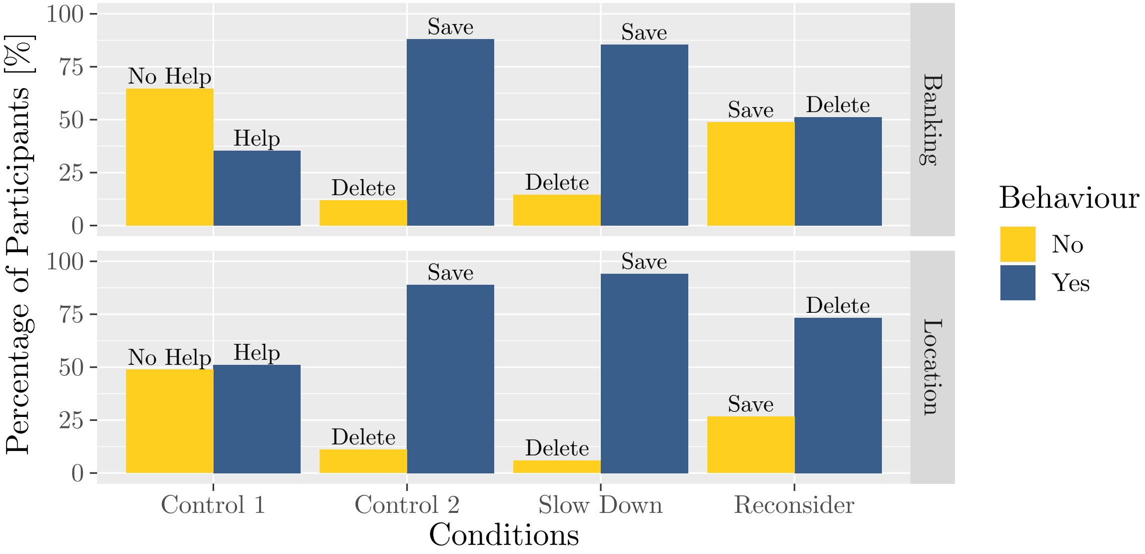


Figure 3.7: Response behaviour of participants in the individual condition questions. Only conditions are displayed where participants could answer with “yes” and “no”. Meaning of participants’ answers are shown above the bars. (see Table 3.5 for user response options and their meanings for all conditions).

slow down condition wanted to have their data saved while in the reconsider condition responding with “yes” refers to a deletion request. We can see that people exposed to the second control or slow down condition are more likely to agree to the suggestion of having their data saved – only 6% to 15% of participants in the respective conditions and scenarios denied. Contrarily, we can see varying answer behaviour for the first control and reconsider condition. Roughly half of the participants asked for further help in the location scenario while only 35% did so in the banking scenario. When exposed to the reconsider condition, roughly half of the participants agreed to deletion of their data in the banking scenario while 74% did so in the location scenario. This suggests that the reconsider condition, but not the slow down condition, affects user behaviour by making it more likely to delete data. Although we can see some variation between scenarios, the described effect is visible across scenarios. Thus, reconsidering data sharing shows some robustness in increasing the likelihood to delete personal data.

We perform binary logistic regression for the “yes”-“no” behaviour while excluding the alternative condition. We compare two models, one including the main effects of conditions and scenarios and a second one including both predictors and their interactions. Comparison of the two models is carried out via a Chi-Square test which shows that the model with interactions does not perform significantly better ($\chi^2(3) = 1.24, p = 0.74$). Further, we test whether there is a statistical significance of the coefficients in the additive model by conducting an ANOVA Type III test using Wald statistics. The results are shown in Table 3.8. One can see that scenario, as well as condition, have a significant effect on the response behaviour. This suggests that while the effect of condition does not depend on the value of scenario as an additive model performs sufficiently well, both variables impact peoples' response behaviour. Nevertheless, we can confirm that response behaviour shows some robustness across scenarios and thus, our tested strategies are likely to yield similar results when applied in varying contexts. We will now report results of the additive model in detail (parameter estimates are shown in Table 3.9). We find that holding all other parameters constant, the odds of responding with "yes" are twice as high as for participants exposed to the location scenario than for participants in the banking scenario. Differences between scenarios are most pronounced for the first control condition and the reconsider condition. For the second control and slow down condition behaviour does not vary significantly between scenarios. This suggests that the first control condition as well as the reconsider condition are sensitive to scenario-specific factors. One factor that may differ between the banking and the location scenario is information sensitivity, with banking data being perceived more sensitive than location data as suggested by a study by Schomakers et al.. However, if banking data is indeed perceived as more sensitive than location, one may expect that more people would opt to delete banking data by responding "yes". However, we saw that almost three-quarters of participants wanted their data to be deleted in the location scenario while fewer did so in the banking scenario. This is the opposite of what one would expect based on the presumed sensitivity of banking and location data. Manual inspection of the transcripts does not reveal an explanation for this behaviour. Therefore, we can only suspect that participants might be more prone to believing that we were able to access their location but might have doubted that we could access banking data (like their credit card number). While location data has been largely researched in the context of mobile apps and privacy Nudges Almuhimedi et al. [2015], studies on accessing participants' real financial data are rare. Thus, we know that people are likely to reassessing location permission requests but do not have similar information on user behaviour regarding financial information. Although many of today's browsers can store credit card numbers in their history or cache, it can be questioned whether all of our participants are aware of the technical term and its implications. Thus, they might have doubted that we could access their credit card information. Additionally, our conversational design could have led to the observed differences in participants' response behaviour between scenarios. In the location scenario, participants are advised to come back tomorrow due to the closure of the restaurant whereas in the banking scenario we pretend to have technical difficulties that keep us from checking the balance. Possibly, participants assumed that no data was transferred in case of the banking scenario due to the outcome and that their data will be safe either way.

Moreover, we encounter that the odds of agreeing are 10.2 times and 10.9 times higher for participants in the second control condition and slow down condition compared to the participants in the first control condition. For people in the reconsider condition the odds of responding with "yes" are only twice as high as for the baseline condition. We further investigate whether there are differences in the coefficients between conditions not compared to the first control

	Df	Chisq	Pr(> $\|Chisq\|$)
(Intercept)	1	4.66	0.03 *
Scenario	1	7.01	0.008 **
Condition	3	50.44	6.44e-11 ***

Table 3.8: Type III ANOVA of the binary logistic regression model on participants' "yes"-"no" response behaviour based on the scenario and condition they were exposed to. Both scenario and condition show a significant effect on the response behaviour.

condition. The chi-squared test statistic indicates that there are statistically significant differences between the coefficients of the second control condition and the reconsider condition and between the slow down and reconsider condition while there is no difference found between the second control condition and the slow down condition ($\chi^2_{Control2/Reconsider}(1) = 15.4, p = 8.6e^{-05}; \chi^2_{SlowDown/Reconsider}(1) = 14.6, p = 0.00013; \chi^2_{Control2/SlowDown}(1) = 0.022, p = 0.88$). The first control condition was supposed to serve as a baseline for "yes"-"no" response behaviour while the second control condition was supposed to nudge participants towards permanent storage of their data. Considering this, differences between the two conditions reveal that we were successful in nudging people to agree to the chatbot's suggestion of saving data. While the slow down condition was supposed to make people rethink their decision, results suggest that an additional time delay does not lead to reconsideration but similarly to the second control condition nudges people to have their data saved. We conclude that given the "yes"-"no" response behaviour, only the reconsider condition successfully affects user behaviour. First, we find that response behaviour in the reconsider condition is distinct from the behaviour seen in the first control condition. This is interesting as it may indicate that people behave differently once there is something at stake, i.e. they are asked to make a decision about their personal data rather than accepting an offer of help from a virtual assistant. Second, we see that response behaviour differs between the reconsider and the second control condition. Notably, in the reconsider condition a "yes" refers to agreeing to delete personal data, whereas a "yes" in control 2 refers to agreeing to save personal data. Thus, the same word ("yes") has opposite effects in reconsider and control 2. If the reconsider condition would merely reverse the Nudge to deleting data rather than having it saved, we would expect a similar distribution of "yes"-"no". However, we do not see a similar distribution between the second control and reconsider condition. Thus, we conclude that the reconsider condition does not nudge people into deletion of their data as the second control condition nudges them into saving their data.

We move on to include our control variables into the model and check whether this leads to an improved model fit. Therefore, we carry out an *Akaike Information Criterion (AIC)* based model selection. AIC uses the maximum likelihood estimate of the model L and the number of parameters in the model K to measure its information value:

Parameter		Estimate	Std. Error	z value	Pr(> $\|z\|$)
(Intercept)		-0.54	0.25	-2.16	0.03 *
Scenario	Location	0.72	0.27	2.65	0.008 **
Condition	Control 2	2.32	0.40	5.74	9.51e-09 ***
	Slow Down	2.39	0.44	5.5	3.85e-09 ***
	Reconsider	0.71	0.31	2.27	0.02 *

Table 3.9: Outcome of the binary logistic regression model for "yes"-"no" response behaviour. The intercept estimate refers to the logit of the probability of participants responding with "yes" who are exposed to the banking scenario and the first control condition, as they serve as references. The parameter estimate for "Location" represents the effect of being exposed to the location scenario compared to the banking scenario. The parameter estimates for "Control 2", "Slow Down" and "Reconsider" represent the effect of being exposed to this condition compared to being exposed to the first control condition.

$$AIC = 2K - 2\ln(L) \tag{3.6}$$

While it rewards the goodness of fit of a model it generally selects the one which explains the outcome best with the fewest number of independent variables. AIC is a relative measure with a lower AIC score indicating a better fit [Akaike, 1998].

The simple binary logistic regression model including only the main effects of scenario and condition holds an AIC of 354.3. We now use the "step" function in R to choose the model with the lowest AIC value in a stepwise algorithm. A trivial model, a binary logistic regression model without predictors, serves as a lower bound and a model including condition, scenario and all control variables as an upper bound. We only check for main effects and neglect possibly interaction effects. Our control variables consist of subjective ratings on privacy concern, privacy literacy, uncertainty avoidance, trust in the chatbot and usage of chatbots. Indeed the best model fit found by the stepwise algorithm resulting in a minimal AIC, is the model we analyzed above. We, therefore, conclude that the response behaviour of participants is not influenced by any of the control variables but only by scenario and condition.

In the following, we compare participants' intention to delete across scenarios and conditions. We now exclude the first control condition from the analysis as it does not offer participants to delete or save their data. Instead, we include the alternative condition where participants had to explicitly state their storage or deletion request by answering "save" or "delete" (see Table 3.5 for condition questions and user response options). In Figure 3.8 we show peoples' intention to delete and save for both scenarios and conditions. We can clearly see that people exposed to the second control or slow down condition are more likely to save their data. In the banking scenario, roughly half of the people exposed to the alternative condition or the reconsider condition wanted their data to be saved while the other half wanted their data to be deleted. In

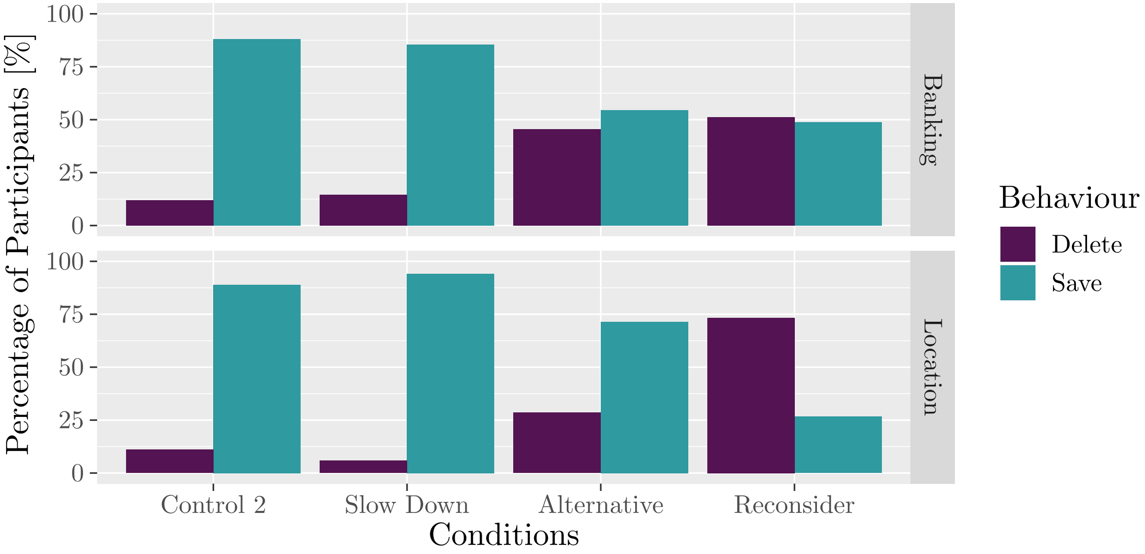


Figure 3.8: Response behaviour of participants in the individual condition questions. Only conditions are displayed where participants could decide whether they wanted to save or delete their data (see Table 3.5 for user response options and their meanings).

the location scenario, participants' intention to delete varies for the alternative and reconsider condition. Here, only roughly 29% of participants asked for deletion of their data when exposed to the alternative condition while 74% did so in the reconsider condition. Again, this suggests that two of our cognitive forcing strategies i.e. the alternative and reconsider condition, affect user behaviour while the slow down condition does not. Although we see variations between scenarios, in the reconsider and alternative condition participants are generally more likely to delete their data. Thus, reconsidering data sharing and active decision between alternatives show robustness across scenarios.

Again, we compare two binary logistic regression models including one with main effects of condition and scenario and one with main and interaction effects. Comparison of the two models is carried out via a Chi-Square test which shows that the model including the interaction is significantly better ($\chi^2(3) = 8.5, p = 0.04$). Therefore, we will report on the model including an interaction effect as this helps to explain more variability among participants behaviour. The results are shown in Table 3.10. We found that the odds of someone asking for deletion of their data is 6.1 times higher when exposed to the alternative condition and 7.7 times higher when exposed to the reconsider condition compared to the second control condition. Although not significant, the odds of deleting personal information when given additional thinking time, i.e. in the slow down condition, is 1.3 times higher compared to the second control condition. This suggests that the slow down condition slightly increases the likelihood that participants adapt their decision based on the information they shared. But the additional time delay is not as useful as the other tested cognitive forcing strategies. Further, we investigate whether there are

Parameter		Estimate	Std. Error	z value	Pr(> $\|z\|$)
(Intercept)		-1.99	0.44	-4.58	4.69e-06 ***
Scenario	Location	-0.09	0.69	-0.13	0.9
Condition	Slow Down	0.23	0.62	0.37	0.71
	Alternative	1.80	0.53	3.41	0.0006 ***
	Reconsider	2.04	0.54	3.81	0.0001 **
Location:Slow Down		-0.90	1.09	-0.84	0.4
Location:Alternative		-0.65	0.84	-0.77	0.44
Location:Reconsider		1.05	0.83	1.30	0.20

Table 3.10: Outcome of the binary logistic regression model for deletion behaviour with interaction effect. The intercept estimate refers to the logit of the probability of participants wanting to delete their data when exposed to the banking scenario and second control condition. This serves as a baseline. The parameter estimate for "Location" represents the effect of being exposed to the location scenario compared to the banking scenario. The parameter estimates for "Slow Down", "Alternative" and "Reconsider" represent the effect of being exposed to this condition compared to the second control condition. Lastly, the interaction estimate gives a multiplicative effect of location and condition.

	Df	Chisq	Pr(> $\|Chisq\|$)
(Intercept)	1	20.96	4.689e-06 ***
Scenario	1	0.02	0.9
Condition	3	23.25	3.588e-05 ***
Scenario:Condition	3	8.13	0.04 *

Table 3.11: Type III ANOVA of the binary logistic regression model for deletion behaviour. While scenario does not show a significant effect on peoples' deletion behaviour, condition does. In addition, we find a significant interaction effect between scenario and condition. For further analysis of the interaction effect, pairwise comparison results are shown in Table 3.12.

differences in the coefficients between conditions not compared to the second control condition. While the chi-squared test statistics did not reveal a statistically significant difference between the coefficients of the alternative and the reconsider condition, statistically significant differences were found among the coefficients of the slow down and alternative as well as slow down and reconsider condition ($\chi^2_{Alternative/Reconsider}(1) = 0.28, p = 0.6; \chi^2_{SlowDown/Alternative}(1) = 8.7, p = 0.0032; \chi^2_{SlowDown/Reconsider}(1) = 11.2, p = 0.0008$). Similarly to what was found when analysing participants' "yes"-"no" response behaviour, deletion behaviour in the second control condition is alike to the behaviour in the slow down condition. In contrast, the alternative and reconsider condition lead to a significant shift in behaviour. Although we do not find a difference when investigating the main effects between alternative and reconsider, from Figure 3.8 it becomes clear that behaviour in those conditions varies across scenarios. Thus, we analyse the interaction effect in detail.

An ANOVA Type III test using Wald statistics shows that conditions and the interaction show a

	Estimate	SE	z-value	p-value
Banking Scenario				
Control 2 - Alternative	-0.33	0.09	-3.8	0.0008 ***
Control 2 - Reconsider	-0.39	0.09	-4.3	0.0001 ***
Control 2 - Slow Down	-0.03	0.07	-0.37	0.98
Alternative - Reconsider	-0.06	0.11	-0.53	0.95
Alternative - Slow Down	0.31	0.09	3.31	0.005 **
Reconsider - Slow Down	0.37	0.1	3.83	0.0007 ***
Location Scenario				
Control 2 - Alternative	-0.17	0.09	-1.89	0.23
Control 2 - Reconsider	-0.62	0.08	-7.4	<.0001 ***
Control 2 - Slow Down	0.05	0.07	0.8	0.86
Alternative - Reconsider	-0.45	0.1	-4.44	0.0001 ***
Alternative - Slow Down	0.23	0.09	2.63	0.04 *
Reconsider - Slow Down	0.67	0.08	8.73	<.0001 ***

Table 3.12: Results of the post-hoc analysis for pairwise comparison within the individual scenarios. The Tukey method is applied for p-value adjustment due to multiple comparison. The estimated differences in probabilities are shown along with the standard errors, and the outcome of the hypothesis tests. We can identify differences among the two scenarios. In particular, the alternative and reconsider condition do not show significant differences in the banking scenario but for the location scenario. This makes sense given the results shown in Figure 3.8 where almost opposite behaviour is visible for people exposed to the alternative and reconsider condition in the location scenario.

significant effect on peoples' intention to delete (see Table 3.11). Therefore, we further analyse the interaction effect, calculate the estimated marginal means and carry out a post-hoc analysis for pairwise comparison between the conditions in the individual scenarios. We compute pairwise comparison using the "emmeans" package in R where the Tukey method is applied to adjust p-values due to multiple comparisons [Lenth, 2022]. We ensure that predictions are on the response scale and thus can be interpreted as probabilities. The results are provided in Table 3.12.

While we see consistency in differences between the conditions among the two scenarios, we can identify two major differences. First, in the banking scenario, deletion behaviour of participants varies significantly between the second control and alternative condition. However, this difference was not found to be significant in the location scenario. Here, considering alternatives increases the likelihood to ask for deletion compared to the second control condition only slightly. Second, in the banking scenario, no significant difference in behaviour is found between the alternative and reconsider condition. However, when exposed to the location scenario, behaviour between those two conditions varies significantly, considering that both conditions are supposed to make people think more thoroughly about permanent disclosure of their personal information. Thus, depending on the individual risk-benefit evaluation participants decide upon deletion or storage of their data. Consequently, decisions are highly subjective and can be influenced by peoples' technical knowledge, privacy concerns and other trait-like as well as situational factors. Moreover, participants were only exposed to one combination of scenario and condition. If both conditions trigger rational risk-benefit assessment and we had exposed the same group of people to both of the conditions in the scenarios, we would expect to see similar behaviour. However, due to the influence of personal traits and individually perceived risks and benefits in a System 2 thinking state, differences in peoples' intentions to delete are not surprising. Instead, this can be seen as an indicator that both conditions have an influence on peoples' decision-making process by taking into account individual perceptions and intentions.

We carry out the same stepwise selection process based on AIC for predicting participants' deletion behaviour. The model above including scenario, condition and their interaction results in an AIC value of 338.7. In the selection process, we include scenario and conditions as predictors as well as their interaction and privacy concern, privacy literacy, uncertainty avoidance, trust in the chatbot and usage as additional predictors. We only consider possible main effects of control variables, not their interactions. The best model according to the stepwise selection process is the model we analyzed above including scenario, condition and their interaction as predictors. We conclude that the above fitted model for predicting deletion behaviour is the preferred one compared to any model including control variables. Because of this, we claim that the designed strategies are the main source of influence on peoples' intention to have their data deleted and can outweigh main effects of trait-like specifics. Nevertheless, we have seen that the alternative and reconsider condition let participants express their intention to delete or save their data freely and possibly dependent on individual perceptions and characteristics.

Assuming that people who granted access to their data in the first place might show different behaviour than those who denied it, we analyse the deletion behaviour considering only people who granted access to their personal information. However, we found that the data including only people who gave access did not differ significantly from the combined dataset.

3.6.3.2 Reliability and Validity

Before analyzing the results of the survey in detail, we test reliability and validity of the scales. To assess reliability, we compute Cronbach's Alpha, Omega total and Omega hierarchical for the combined dataset. We show results of the analysis in Table 3.13. Given the Cronbach's Alpha and Omega total values, one can say that reliability is in general acceptable for all of the scales. Only the scale on protection uncertainty holds a low Omega hierarchical similar to what was found in the pilot study. This suggests that a general factor explains only little proportion of the variance and instead group factors account for the variance.

To assess Construct Validity we apply factor analysis on the combined dataset of location and banking. KMO measure of sampling adequacy suggests that most of the scales have good

Scale	Cronbach's Alpha	Omega Total	Omega Hierarchical
Fear	.94	.95	.9
Collection Uncertainty	.69	.76	.58
Use Uncertainty	.81	.84	.74
Protection Uncertainty	.74	.74	.04
Overall Uncertainty	.79	.82	.74
Privacy Perception	.78	.83	.68
Usability	.78	.83	.66
Trust Chatbot	.74	.76	.66
IUIPC	.7	.79	.65
Privacy Literacy	.64	.71	.58
Uncertainty Avoidance	.71	.78	.49

Table 3.13: Reliability analysis for the measures used in the main study. Calculations of Cronbach's Alpha, Omega total and Omega hierarchical are performed using the "psych" package in R [Revelle, 2021]. Details on the calculation can be found in Section 3.5.2.1.

sampling adequacy with values between 0.73 and 0.95, only the scale on privacy literacy shows a mediocre result with a value of 0.68. Thus, a reliable factor analysis can be performed (see Appendix F for details). For the scale measuring the underlying construct "Fear" we find that one factor is not sufficient at an α-level of 0.05 ($\chi^2(9) = 19.77, p = 0.019$). Therefore, we apply a factor analysis with two factors and find that one factor mainly captures properties of nervousness while the second factor relates to anxiety. Moreover, we show that five factors are needed to describe the privacy uncertainty scale. However, we found that the third, fourth and fifth factor explain only 7.5%, 5.2% and 5% of the variance in the data set. Moreover, we could not identify another distinct pattern of sub-constructs of uncertainty and thus will use the original subdivision into collection, use, protection and overall uncertainty for further analysis.

The null hypothesis for a factor analysis on perception of privacy using one factor is rejected on an α-level of 0.05 ($\chi^2(9) = 19.52, p = 0.021$). Instead, we find that two factors are sufficient with the second factor being related to the item of the chatbot showing concern for the users' privacy. While one factor was sufficient to explain the variance within the usability scale in the pilot study, now we find that three factors describe the variability sufficiently well ($\chi^2(3) = 3.03, p = 0.387$). Further analysis shows that we can distinguish between items referring to the understandability and controllability of the chatbot constituting to the first factor, while the second factor is determined by the simplicity of use of the chatbot. The third factor mainly refers to structure and content of the chatbot. We acknowledge that those three factors constitute to the overall usability of the chatbot and thus provide reasonable insights into peoples' perception of usability.

Further, one factor was found to be sufficient for the scale measuring trust in the chatbot. Similarly to the pilot study, we can hardly confirm the three-dimensionality of the scale measuring privacy concern divided into control, awareness and collection. Nevertheless, factor analysis with three factors performed sufficiently well with a tendency of the division into the three subscales ($\chi^2(3) = 5.74, p = 0.125$). Moreover, factor analysis on the privacy literacy scale shows that one factor is not sufficient suggesting that the scale can be enhanced ($\chi^2(2) = 13.36, p = 0.001$). Lastly, for the scale on uncertainty avoidance two factors suffice with the second factor relating solely to the item of "Instructions for operations are important"($\chi^2(1) = 0.58, p = 0.445$). Thus, as the reliability analysis suggests, the general factor predominately explains variability in the ratings.

3.6.3.3 Analysis of Perceptions

To analyse peoples' perceptions, we investigate their ratings of subjective scales in the questionnaire they filled out after interacting with the chatbot. Our analysis is based on the aggregated data of the individual scales. Participants' aggregated ratings on frustration, fear and the uncertainty scales are shown in Figure 3.9. One can see that frustration and fear have large variances compared to the uncertainty scales for both scenarios and conditions. Further analysis showed that frustration ratings were rather uniformly distributed across the rating scale. This could be a result of measuring frustration using only one single item. Moreover, the aggregated fear ratings show a bimodal distribution with some people indicating low fear ratings and others indicating high fear ratings.

Furthermore, we show participants' aggregated ratings on perception of privacy and usability of the chatbot in Figure 3.10. While there is almost no difference of participants' ratings in the banking scenario, we can see some differences in participants' ratings among conditions in the location scenario.

As we are interested in possible effects of condition and scenario on the individual scales, we perform ordinal logistic regression. We compare models including only the main effects and models including main and interaction effects to predict the aggregated ratings. Similarly to analyzing peoples' behaviour, we use the first control condition and the banking scenario as a reference. We found that collection and use uncertainty are significantly affected by scenario when fitting a model with scenario and condition as predictors, while all other subjective measurements are not influenced by scenario nor condition. The models including interaction effects did not succeed in explaining more variability among collection or use uncertainty ratings and we will therefore report on the simpler model. The model coefficients are provided in Table 3.14 and Table 3.16. Moreover, the results of the Type III ANOVA on the models are shown in Table 3.15

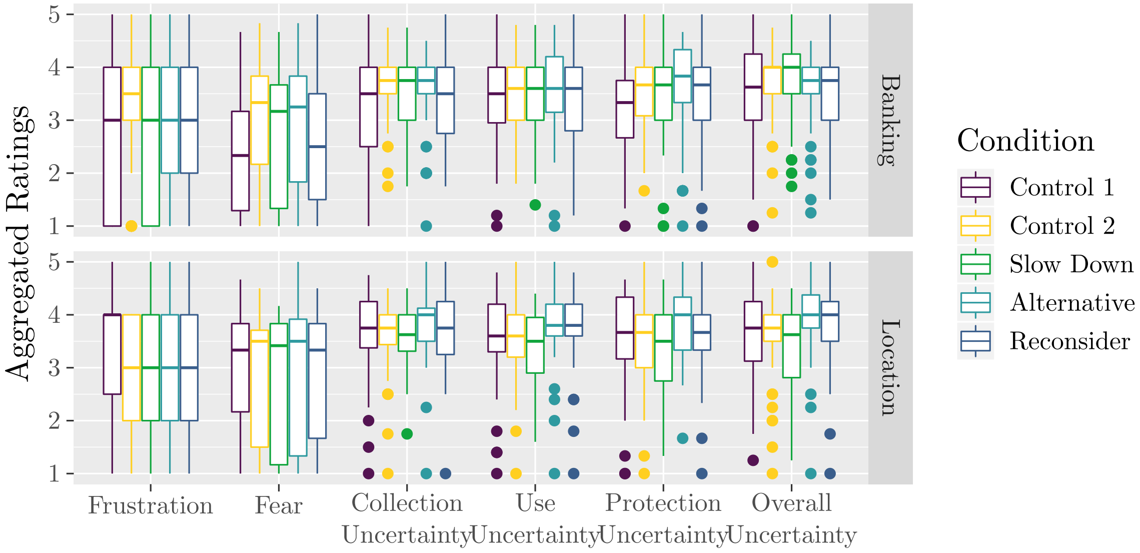


Figure 3.9: Boxplot of aggregated frustration, fear and uncertainty ratings for both scenarios and conditions. The mean was taken over all subscale items to compute the aggregated ratings per participant.

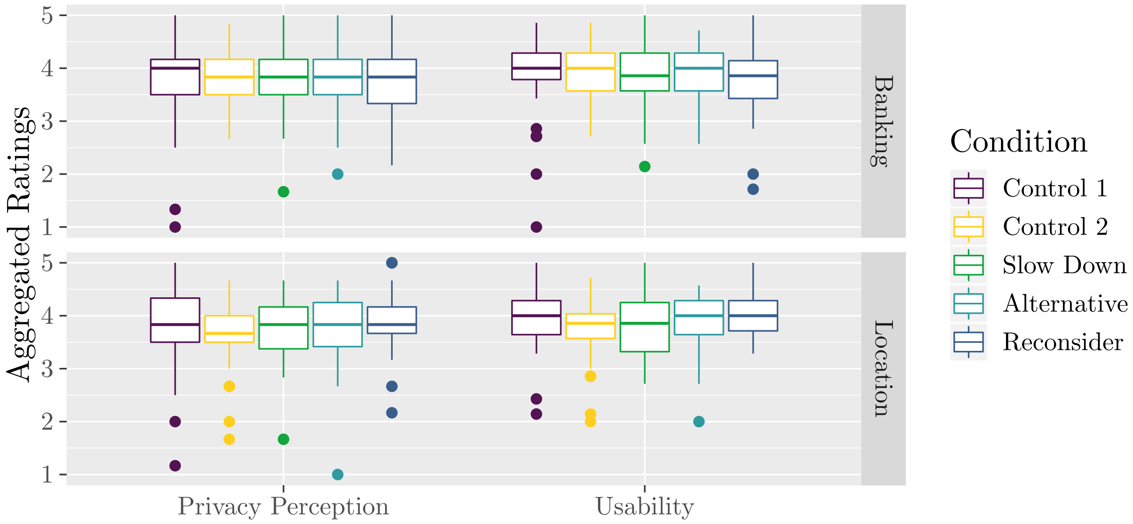


Figure 3.10: Boxplot of aggregated perception of privacy and usability ratings for both scenarios and conditions. The mean was taken over all subscale items to compute the aggregated ratings per participant.

Parameter		Estimate	Std. Error	t value	p value
Scenario	Location	0.37	0.17	2.15	0.03 *
Condition	Control 2	0.15	0.27	0.55	0.58
	Slow Down	-0.01	0.28	-0.02	0.98
	Alternative	0.35	0.28	1.25	0.21
	Reconsider	0.02	0.27	0.08	0.93

Table 3.14: Outcome of the ordinal logistic regression on collection uncertainty. Only the main effects of scenario and condition are considered.

	Df	Chisq	Pr(> $|Chisq|$)
Scenario	1	4.64	0.03 *
Condition	4	2.38	0.67

Table 3.15: Type III ANOVA of the ordinal logistic regression model on collection uncertainty. While scenario shows a significant effect on peoples' collection uncertainty ratings, condition does not.

Parameter		Estimate	Std. Error	t value	p value
Scenario	Location	0.41	0.17	2.37	0.02 *
Condition	Control 2	0.08	0.26	0.30	0.76
	Slow Down	-0.12	0.27	-0.45	0.65
	Alternative	0.49	0.27	1.79	0.07
	Reconsider	0.25	0.27	0.92	0.36

Table 3.16: Outcome of the ordinal logistic regression on use uncertainty. Only the main effects of scenario and condition are considered. There is a tendency that use uncertainty ratings for the alternative condition differ from ones provided by participants exposed to the first control condition.

	Df	Chisq	Pr(> $|Chisq|$)
Scenario	1	5.64	0.02 *
Condition	4	5.81	0.21

Table 3.17: Type III ANOVA of the ordinal logistic regression model on use uncertainty. While scenario shows a significant effect on peoples' use uncertainty ratings, condition does not.

and Table 3.17.

For people exposed to the location scenario the odds of feeling more uncertain regarding collection and use uncertainty are 1.5 times higher than for people exposed to the banking scenario. Because there is a tendency for participants in the alternative condition to report higher use uncertainty compared to the ones in the first control condition, we also investigate the odds ratios in this case. The odds of reporting higher uncertainty values regarding use uncertainty is 1.6 times higher in the alternative condition compared to the first control condition.

When fitting a model to the aggregated ratings of protection uncertainty using scenario and condition as predictors, we discover that there is a tendency of condition to influence participants' ratings ($\chi^2_{Scenario}(1) = 2.17, p = 0.14, \chi^2_{Condition}(4) = 8.3, p = 0.08$). The model coefficient of the alternative condition shows a significant difference compared to the first control condition ($\beta = 0.68, p = 0.01$). The odds of being more likely to report protection uncertainty are 1.98 times higher when exposed to the alternative condition compared to the first control, the simple assistance question. Although little differences are visible in perceived privacy and usability ratings among people exposed to the location scenario (see Figure 3.10), neither scenario nor condition nor their interaction effects showed a significant impact on perceived privacy or usability.

To further investigate the effects found by fitting simple models consisting of only scenario and condition, we now add control variables to the models and conduct a stepwise selection based on the AIC values. As control variables we include ratings on privacy concern, privacy literacy, trust in the chatbot, uncertainty avoidance and usage. When including control variables, we still find that collection and use uncertainty are significantly influenced by scenario as well as by privacy literacy, privacy concern, uncertainty avoidance and usage or trust in the chatbot. Model coefficients and results of the Type III ANOVA are provided in Table 3.18 and 3.19 for collection uncertainty and in Table 3.20 and 3.21 for use uncertainty. The AIC values for the extended model of collection uncertainty calculate to 1896.3 compared to the AIC value of 1984.64 for the simple model investigated above. Similarly, for the model of use uncertainty the AIC value decreases from 2282.36 to 2192.44. For both extended models the AIC values decrease by more than two AIC units compared to the simpler models (1984.64 - 1896.30 = 88.34 units and 2282.36 - 2192.44 = 89.92 units). Thus, they can be considered significantly better.

Similar to the simple model, the odds of reporting higher collection uncertainty ratings and use uncertainty ratings are 1.5 times and 1.7 times higher for people exposed to the location scenario than to the banking scenario. Moreover, for every one unit increase in privacy literacy the odds of being more likely to report collection uncertainty increases by 43%, i.e. is multiplied 1.43 times. When it comes to use uncertainty, an increase of 72% with every one unit increase in privacy literacy is visible. In addition, privacy concern has an effect on collection uncertainty. Here, for every one unit increase in privacy concern the odds calculate to 11.6 times higher likelihood of reporting collection uncertainty and to 14.3 times higher likelihood of reporting use uncertainty. Furthermore, with every one unit increase in uncertainty avoidance the odds of being more likely to report collection uncertainty are $(1 - 0.37) * 100\% = 63\%$ lower. Similarly, the odds of being more likely to report use uncertainty reduces by $(1 - 0.31) * 100\% = 69\%$ for every increase in uncertainty avoidance. Given those results, we conclude that trait-like characteristics heavily influence peoples' reports on collection and use uncertainty. As expected, participants who are more literate and more concerned about privacy in online environments are more likely to report uncertainty with regard to collection and usage of data. However, cultural differences also play a role as participants with a high need of avoiding uncertainty are generally less likely to report uncertainty with regard to collection and use. Lastly, we found that usage of the chatbot has

Parameter		Estimate	Std. Error	t value	p value
Scenario	Location	0.43	0.18	2.44	0.015 *
Privacy Literacy		0.36	0.18	1.96	0.05 *
Privacy Concern		2.45	0.31	7.9	2.73e-15 ***
Uncertainty Avoidance		-0.99	0.24	-4.13	3.69e-05 ***
Usage	"less than once a month"	-0.19	0.24	-0.77	0.44
	"2-4 times a month"	-0.33	0.24	-1.37	0.17
	"more than once a week"	-0.72	0.29	-2.5	0.013 *

Table 3.18: Outcome of the extended ordinal logistic regression model on collection uncertainty. The model resulting in a minimal AIC value includes scenario, privacy literacy, privacy concern, uncertainty avoidance and usage of the chatbot as predictors. Similarly to the simple model, location as opposed to the banking scenario is associated with a higher likelihood of reporting higher collection uncertainty ratings. With one unit increase in privacy literacy or privacy concern the log odds of reporting higher collection uncertainty ratings increase by 0.36 and 2.45, both statistically significant effects. Contrarily, with one unit increase in uncertainty avoidance the log odds of reporting higher collection uncertainty decrease by 0.99. For people using chatbots more than once a week as opposed to participants having never used a chatbot before we find a significant effect for the likelihood to reporting lower collection uncertainty ratings.

	Df	Chisq	Pr(> $\|Chisq\|$)
Scenario	1	5.99	0.01 *
Privacy Literacy	1	4.0	0.05 *
Privacy Concern	1	63.15	1.915e-15 ***
Uncertainty Avoidance	1	17.079	3.587e-05 ***
Usage	3	6.75	0.08

Table 3.19: Type III ANOVA of the extended ordinal logistic regression model on collection uncertainty. We find that scenario, privacy literacy, privacy concern and uncertainty avoidance have a significant effect on peoples' collection uncertainty ratings.

an impact on collection uncertainty. Compared to participants who are not using chatbots on average, the likelihood of reporting collection uncertainty decreases the more often participants are using chatbots $(1 - 0.83) * 100\% = 17\%$ less likely for participants using chatbots less than once a month, $(1 - 0.72) * 100\% = 28\%$ less likely for participants using chatbots 2-4 times a month and $(1 - 0.48) * 100\% = 52\%$ less likely for participants using chatbots more than once a week). Moreover, use uncertainty is influenced by peoples' trust in the chatbot. Thus, the likelihood of reporting on use uncertainty decreases by $(1 - 0.66) * 100\% = 34\%$ for every one unit increase in trust.

When controlling for trait-like characteristics, we now find that protection uncertainty is significantly influenced by scenario, privacy literacy, privacy concern, uncertainty avoidance and trust in the chatbot. As seen before, there is a tendency that condition impacts protection uncertainty. Again the model is found by performing stepwise selection based on the AIC values. The model

Parameter		Estimate	Std. Error	t value	p value
Scenario	Location	0.54	0.17	3.14	0.002 **
Privacy Literacy		0.54	0.2	2.76	0.006 **
Privacy Concern		2.66	0.24	-4.9	1.17e-16 ***
Uncertainty Avoidance		-1.16	0.24	-4.85	1.22e-06 ***
Trust in the Chatbot		-0.41	0.18	-2.3	0.02 *

Table 3.20: Outcome of the extended ordinal logistic regression model on use uncertainty. The model resulting in a minimal AIC value includes scenario, privacy literacy, privacy concern, uncertainty avoidance and trust in the chatbot as predictors. Similarly to the simple model, location as opposed to the banking scenario is associated with a higher likelihood of reporting higher use uncertainty ratings. With one unit increase in privacy literacy or privacy concern the log odds of reporting higher use uncertainty ratings increase by 0.54 and 2.66, both statistically significant effects. Contrarily, with one unit increase in uncertainty avoidance or trust in the chatbot the log odds of reporting higher use uncertainty ratings decrease significantly.

	Df	Chisq	Pr($>|Chisq|$)
Scenario	1	9.9	0.002 **
Privacy Literacy	1	7.62	0.006 **
Privacy Concern	1	72.08	<2.2e-16 ***
Uncertainty Avoidance	1	23.73	1.109e-06 ***
Trust in the Chatbot	1	5.2	0.022 *

Table 3.21: Type III ANOVA of the extended ordinal logistic regression model on use uncertainty. We find that all parameters included in the extended model have a significant effect on peoples' use uncertainty ratings.

Parameter		Estimate	Std. Error	t value	p value
Scenario	Location	0.35	0.17	2.05	0.04 *
Condition	Control 2	0.54	0.27	2.03	0.04 *
	Slow Down	0.35	0.27	1.28	0.2
	Alternative	0.8	0.28	2.84	0.004 *
	Reconsider	0.51	0.27	1.93	0.05 *
Privacy Literacy		0.44	0.20	2.19	0.03 *
Privacy Concern		2.65	0.31	8.45	2.87e-17 ***
Uncertainty Avoidance		-1.21	0.24	-5.13	2.97e-07 ***
Trust in the Chatbot		-0.37	0.18	-2.1	0.04 *

Table 3.22: Outcome of the extended ordinal logistic regression on protection uncertainty. The model resulting in a minimal AIC value includes scenario, condition, privacy literacy, privacy concern, uncertainty avoidance and trust in the chatbot as predictors. Again, location as opposed to the banking scenario is associated with a higher likelihood of reporting higher protection uncertainty ratings. More importantly, the second control, the alternative and the reconsider condition as opposed to the first control condition show significant differences and are associated with higher likelihood of reporting higher protection uncertainty ratings. Similarly to the extended model on use uncertainty ratings with one unit increase in privacy literacy or privacy concern the log odds of reporting higher protection uncertainty increase by 0.44 and 2.65, both statistically significant effects. Contrarily, with one unit increase in uncertainty avoidance or trust in the chatbot the log odds of reporting higher protection uncertainty ratings decrease significantly.

	Df	Chisq	Pr(> $\|Chisq\|$)
Scenario	1	4.24	0.04 *
Condition	4	8.9	0.06
Privacy Literacy	1	4.84	0.03 *
Privacy Concern	1	73.06	<2.2e-16 ***
Uncertainty Avoidance	1	26.29	2.939e-07 ***
Trust in the Chatbot	1	4.46	0.035 *

Table 3.23: ANOVA Type III of the extended ordinal logistic regression model on protection uncertainty. We find that scenario, privacy literacy, privacy concern, uncertainty avoidance and trust in the chatbot have a significant effect on peoples' protection uncertainty ratings. Condition does not show an overall significant effect, however, we find significant differences between distinct conditions as shown in Table 3.22.

coefficients and the results of the statistical analysis are provided in Table 3.22 and Table 3.23.

The results of the ordinal logistic regression show a significant effect on protection uncertainty between the first control condition and the second control condition as well as between the first control condition and the alternative and reconsider condition. However, a chi-squared test did not reveal statistically significant differences between the second control condition and conditions supposed to trigger cognitive forcing. We find that the odds of being more likely to articulate protection uncertainty are 2.2 times higher for participants exposed to the alternative condition

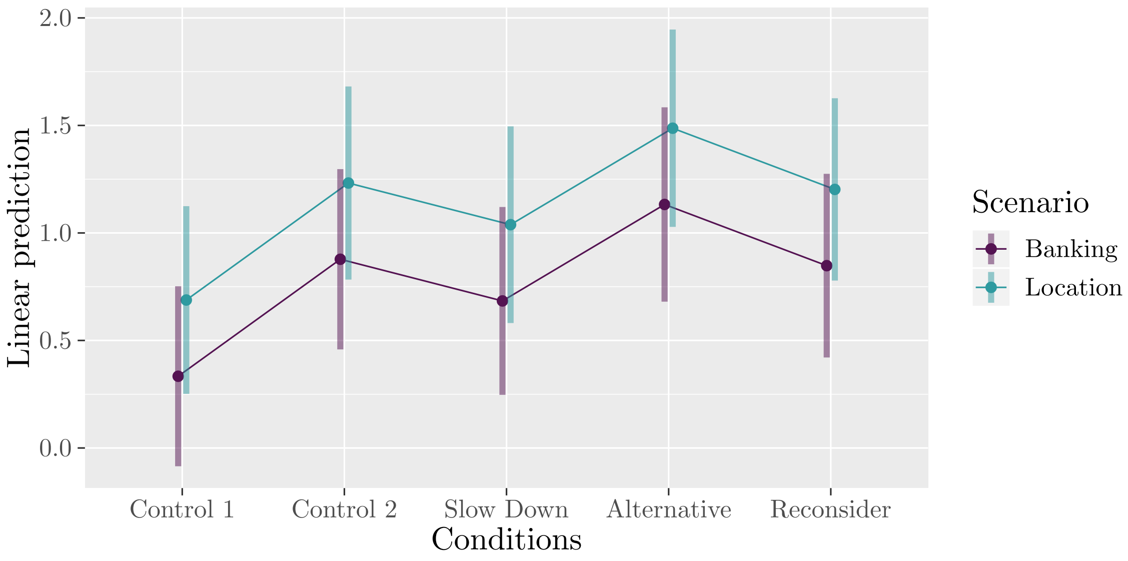


Figure 3.11: Effect plot of condition and scenario on participants' aggregated ratings on protection uncertainty. We can clearly see a difference in scenario among the uncertainty ratings with higher uncertainty perceived in the location scenario. In both scenarios participants' rate highest on protection uncertainty when exposed to the alternative condition. Moreover, the second control condition, the alternative and the reconsider condition are significantly associated with a higher likelihood of reporting higher protection uncertainty values as opposed to the first control condition.

compared to those exposed to the first control condition. However, the odds calculate to 1.7 for participants in the second control condition and in the reconsider condition and to 1.4 in the slow down condition compared to the first control condition. In addition, the effects plot in Figure 3.11 shows that highest uncertainty regarding protection of data is experienced by people exposed to the location scenario and the alternative condition. Moreover, we clearly see an increase in protection uncertainty when moving from the first control condition to any of the other tested conditions. This suggests that all of our questions pointing towards data collection increase uncertainty of the users compared to the simple assistance question. We have to keep in mind though that only the alternative and reconsider condition change users' behaviour drastically. As we measure uncertainty ratings retrospectively it might be possible that we need to distinguish between privacy uncertainty in general and privacy uncertainty which triggers System 2 thinking in the moment of decision-making. It could be that participants exposed to the second control condition decide intuitively at the time of decision-making but feel uncertain about data protection when being reminded of their decision in the questionnaire. Instead, we might suspect that participants exposed to the alternative or reconsider condition already experience uncertainty at the time of the decision-making which leads them to varying behaviour based on the information they shared. Further research is therefore needed which measures uncertainty and System 2 thinking when decision-making takes place, for example using physiological measures such as pupil dilation. Comparable to the previous investigated uncertainty scales, we find that the odds of reporting protection uncertainty is 1.4 times higher for people exposed to the location scenario compared to the ones exposed to the banking scenario. Moreover, for every unit increase in privacy literacy the likelihood of claiming protection uncertainty increases by 56%. An increase in privacy concern leads to 14 times higher likelihood of reporting protection uncertainty. Moreover, we find that people are less likely to report protection uncertainty for every unit increase in uncertainty avoidance (a decrease of $(1 - 0.3) * 100\% = 70\%$) and for every unit increase in trust in the chatbot (a decrease of $(1 - 0.67) * 100\% = 33\%$).

When investigating participants deletion behaviour we found that people exposed to the alternative and reconsider condition showed significantly different behaviour compared to the ones exposed to the control 2 condition. Based on the difference in behaviour we now analyse perceptions of people exposed to those conditions only. Figure 3.12 shows the aggregated ratings on the fear, frustration and uncertainty scales and Figure 3.13 those of privacy perception and usability in detail. Especially, in the location scenario we can see small differences among participants' ratings in the three conditions. For further analysis, we again carry out ordinal logistic regression to find differences between the perceptions of people exposed to the three conditions.

This time, we only report on the models including control variables as we saw that they can significantly influence peoples' perception and usually help to explain the variability in the data set. Again, we find that use uncertainty is significantly influenced by scenario but not by condition ($\chi^2(1) = 7.04, p = 0.008$). The model on collection uncertainty which holds a minimal

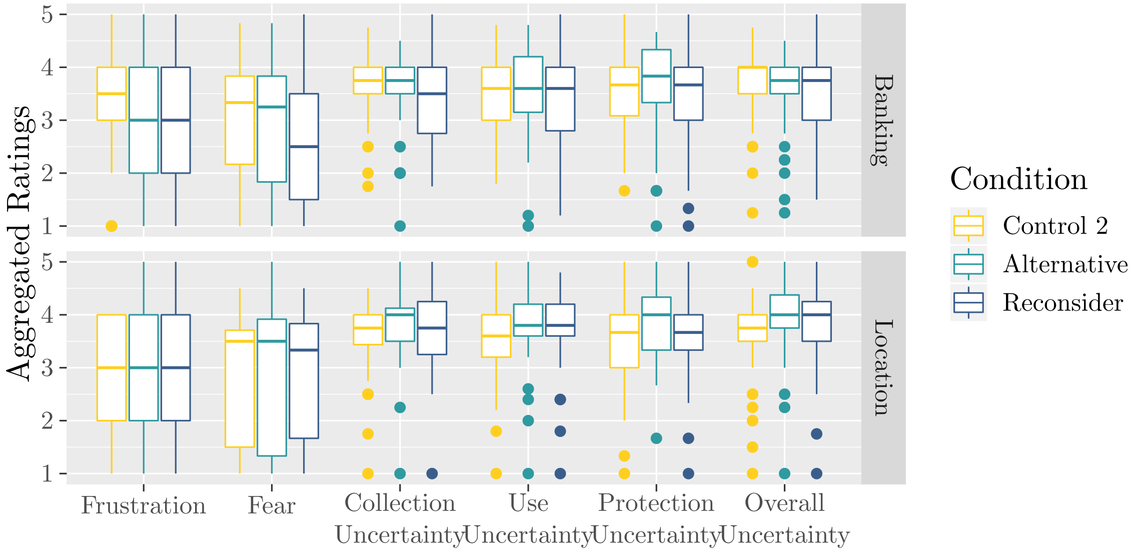


Figure 3.12: Boxplot of peoples' frustration, fear and uncertainty ratings for both scenarios and conditions. Only participants' perceptions are displayed who were exposed to the second control condition, the alternative and reconsider condition.

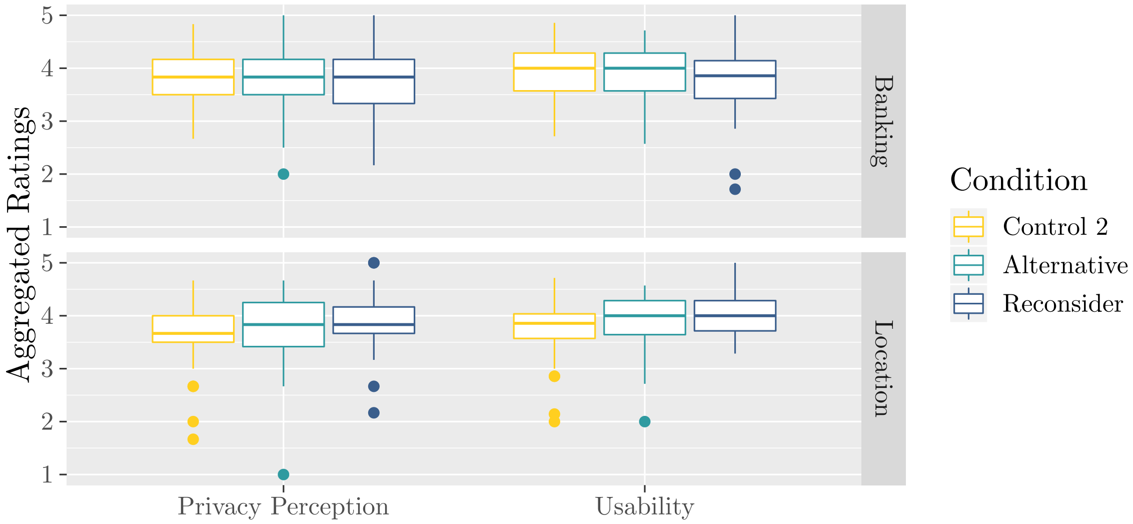


Figure 3.13: Boxplot of peoples' perception of privacy and usability. Only participants' perceptions are displayed who were exposed to the second control condition, the alternative and reconsider condition.

AIC value includes scenario as a predictor but without significant impact. While the model on protection uncertainty is not influenced by scenario or condition, a model to predict participants' ratings on overall uncertainty holding a minimal AIC value includes both condition and scenario as well as their interaction effect. Detailed results are shown in Table 3.24 and Table 3.25. Similarly to the previously fitted models on uncertainty, we find that privacy literacy, privacy concern and uncertainty avoidance significantly predict overall uncertainty. More interesting though, we find a significant crossover interaction effect as shown in more detail in Figure 3.14. While overall uncertainty increases for people exposed to the alternative condition and the location scenario compared to the second control condition, overall uncertainty decreases for participants exposed to the alternative condition and the banking scenario. Given that almost three-quarters of participants exposed to the alternative condition and the location scenario wanted their data to be saved, the cross-over interaction effect regarding overall uncertainty seems to be surprising. One could argue that the more participants save their data when exposed to the alternative condition, the more they experience uncertainty when asked about it retrospectively. However, this argument should also hold for the second control condition where more than 85% of participants in both scenarios decided to have their data saved. Instead, people report medium levels of overall uncertainty when exposed to the second control condition. Therefore, we argue that the overall uncertainty was not solely triggered by the questionnaire but by the condition itself. Nevertheless, scenario differences remain. Possible reasons are manifold. The overall uncertainty reported by participants in the location scenario and the alternative condition could be a sum of the uncertainty experienced due to the cognitive forcing and the uncertainty experienced when reporting retrospectively. While their decision to have their data saved might have been more rational in the alternative condition, the possible cost-benefit analysis might have raised awareness of the exposure and thus leading to higher overall uncertainty. Moreover, we have to keep in mind that the location scenario itself has lead to increased uncertainty over all the conditions which contradicts our assumption of information sensitivity. Thus, we believe, that participants were more likely to believe that we could access their location data compared to their credit card number.

Interestingly, when comparing only the three conditions i.e. control 2, alternative and reconsider, and including control variables we find that privacy perceptions are influenced by condition. Especially, people exposed to the reconsider condition report significantly higher levels of privacy compared to participants exposed to the second control condition. Model details and the carried out Type III ANOVA are shown in Table 3.26 and 3.27. The odds of being more likely to perceive privacy with regards to the chatbot increases by 35% and by 72% when being exposed to the alternative or reconsider condition compared to the second control condition. Moreover, with every one unit increase in privacy concern or privacy literacy participants are more likely to give higher ratings on perceived privacy (2.33 times for privacy concern and 1.8 times for privacy literacy). Moreover, trust in the chatbot has a large influence on privacy perception, with every

Parameter		Estimate	Std. Error	t value	p value
Scenario	Location	-0.21	0.39	-0.55	0.58
Condition	Alternative	-0.53	0.37	-1.46	0.14
	Reconsider	-0.57	0.37	-1.55	0.12
Privacy Literacy		0.46	0.24	1.94	0.05 *
Privacy Concern		2.71	0.43	6.32	2.56e-10 ***
Uncertainty Avoidance		-0.70	0.34	-2.04	0.04 *
Location:Alternative		1.23	0.56	2.17	0.03 *
Location:Reconsider		0.60	0.55	1.08	0.28

Table 3.24: Outcome of the extended ordinal logistic regression on overall uncertainty for the three conditions: control 2, alternative and reconsider. We find that one unit increase in privacy literacy and privacy concern results in a significantly positive effect on the log odds of overall uncertainty ratings while one unit increase in uncertainty avoidance results in a significantly negative effect on the log odds of overall uncertainty ratings with participants being more likely to report lower uncertainty ratings. Moreover, we find a significant interaction effect. Thus, participants exposed to the location and alternative condition are associated with higher overall uncertainty ratings as opposed to participants in the banking and control condition.

	Df	Chisq	$\Pr(> \lvert Chisq \rvert)$
Scenario	1	0.30	0.58
Condition	2	3.13	0.21
Privacy Literacy	1	3.78	0.05 *
Privacy Concern	1	42.76	6.194e-11 ***
Uncertainty Avoidance	1	4.22	0.04 *
Scenario:Condition	2	4.77	0.09

Table 3.25: Type III ANOVA of the extended ordinal logistic regression model on overall uncertainty for the three conditions: control 2, alternative and reconsider. We see that privacy literacy, privacy concern and uncertainty avoidance have a significant effect on overall uncertainty. Moreover, there is a tendency for significant interaction effects as shown in Table 3.24.

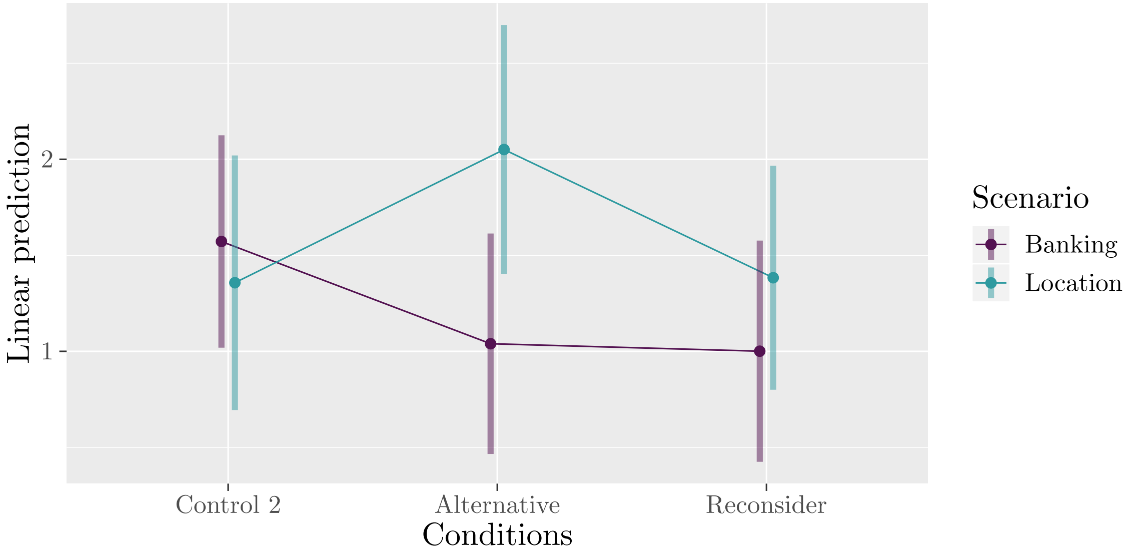


Figure 3.14: Effect plot of condition and scenario on participants' aggregated ratings on overall uncertainty. We can clearly see a crossover interaction effect between overall uncertainty ratings provided in the second control and alternative condition for the scenarios.

Parameter		Estimate	Std. Error	t value	p value
Condition	Alternative	0.3	0.27	1.1	0.27
	Reconsider	0.54	0.27	2.0	0.046 *
Privacy Literacy		0.60	0.28	2.17	0.03 *
Privacy Concern		0.85	0.35	2.4	0.02 *
Trust in the Chatbot		3.85	0.35	11.07	1.72e-28 ***

Table 3.26: Outcome of the extended ordinal logistic regression on privacy perception for the three conditions: control 2, alternative and reconsider. The reconsider condition as opposed to the second control condition is significantly associated with a higher likelihood of reporting increased privacy perception. Moreover, with one unit increase in privacy literacy, privacy concern, and trust in the chatbot the log odds of reporting higher privacy perception increase significantly.

	Df	Chisq	Pr(> $\|Chisq\|$)
Condition	2	4.02	0.13
Privacy Literacy	1	4.8	0.03 *
Privacy Concern	1	5.8	0.016 *
Trust in the Chatbot	1	131.65	< 2e-16 ***

Table 3.27: Type III ANOVA of the extended ordinal logistic regression model on privacy perception for the three conditions: control 2, alternative and reconsider. Privacy literacy, privacy concern and trust in the chatbot show a significant effect on peoples' privacy perception.

one unit increase leading to 47.6 times higher perceived privacy.

3.6.3.4 Analysis of Reaction and Completion Times

In addition to the analysis of participants' behaviour and ratings, we investigate reaction and completion times. Participants spent roughly 5.5 minutes on average on the interaction with the location chatbot and 5.2 minutes with the banking chatbot. Moreover, they spent roughly 6 minutes on average on the survey in the location scenario and 5 minutes on the survey when exposed to the banking scenario.

As in Section 3.5.2.2, we analyse reaction times according to Equation 3.5. Again some of the time stamps were not captured correctly. Nevertheless, we are left with 108 sets for the location scenario and 148 sets for the banking scenario, uniformly distributed across conditions. For those, we compute the average time taken to respond to the chatbot's questions for all interactions carried out before the condition-specific question (see Appendix D for the dialogue trees in the main study). We then compare this baseline to the time taken to respond to the condition-specific question and denote this as time difference Δt. Results are shown in Figure 3.15 for all conditions and the sets available. Positive values indicate that participants took more time to respond to the question offering further help and the questions concerning future handling of the data. Instead, negative values indicate that participants took less time to respond to those prompts compared to the average time taken before. While we see little increase in time taken in the alternative and reconsider condition, a Kruskal-Wallis test does not show any differences among groups neither in the location nor in the banking scenario. Interestingly, time differences in the first control condition span a wide range of values which indicates that some people responded rather quickly to the question on further assistance while some took more time for their decision. Contrarily, time differences of participants exposed to the second control condition, the nudging approach, indicate that participants took similarly long or were faster in their decision-making. This makes sense in light of participants' response behaviour and again suggests that we were successful in nudging participants to agree to the chatbot's suggestion.

Based on participants response behaviour and their perceptions, we further exclude the first control and the slow down condition from the analysis. We first carry out a Kruskal-Wallis test on time differences between the second control, alternative and reconsider condition for the individual scenarios. However, we do not find significant differences. Only when analysing the combined dataset of time differences in the banking and location scenario, we find significant differences between the nudging approach, consideration of alternatives and reconsideration ($\chi^2(2) = 6.11, p = 0.047$). We further investigate group differences by conducting a Wilcoxon rank-sum test for pairwise comparison between conditions. We adjust the α-level accordingly by applying the Bonferroni method to account for multiple comparison. We find that time differences differ significantly between the second control and the alternative condition (Bonferroni adjusted p-value: $p = 0.033$). Figure 3.16 shows the combined dataset of the time differences for the three conditions. We conclude that participants exposed to the second control condition, the

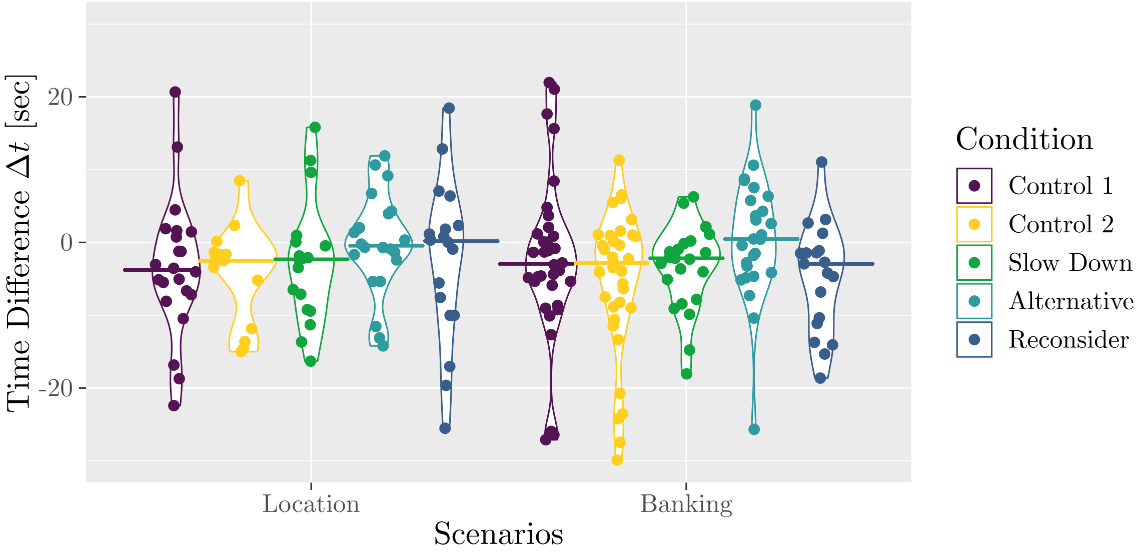


Figure 3.15: Time Differences between the time taken to respond to the condition question and the baseline provided by the average time difference over all interactions before for individual users (see Equation 3.5). 22 outliers are not visible in this Figure to allow a close-up comparison of group differences. Median values are indicated by colored lines. Kruskal-Wallis test carried out for the two scenarios individually does not show significant differences.

Nudge to save data, might use mental shortcuts to make their decision. This is supported by the observation that the majority of participants agrees to save data for future interactions, thus agreeing to the default Nudge. In addition, we observe that the response to the Nudge is on average faster than prior responses and faster than in other conditions. In contrast, the alternative condition seems to slow down decision-making. In fact, we observe a significant difference between the control 2 and alternative condition.

3.6.3.5 Analysis on Informed Decision-Making

In the medical context, the concept of informed decision-making and how to measure and assess related variables has been discussed for many years [Marteau et al., 2001, Ghanouni et al., 2016]. Studies on informed decision-making generally use different terms that encompass informed choice as noted by Marteau et al.: (1) "an informed decision is one where all the available information about the health alternatives is weighed up and used to inform the final decision; the resulting choice should be consistent with the individual's values." (2) "An evidence-informed patient choice is one in which individuals are given research-based information on two or more options and have some input into the decision-making process" (3) "an effective decision is one that is informed, consistent with the decision-maker's values and behaviourally implemented". Our strategies are not designed to give people information to consider before deciding upon

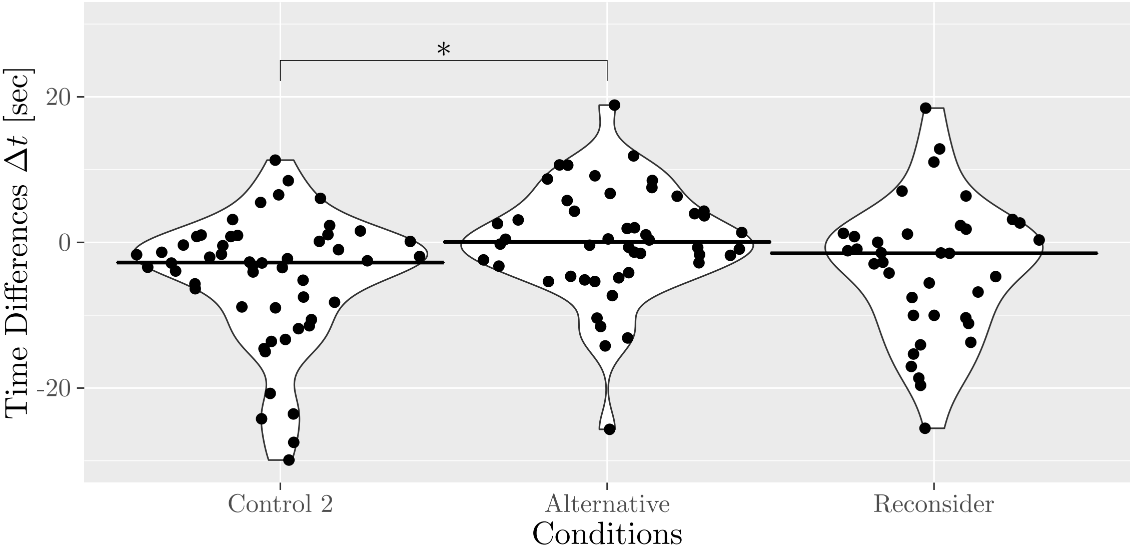


Figure 3.16: Time Differences between the time taken to respond to the condition question and the baseline provided by the average time difference over all interactions before for individual users (see Equation 3.5). We show time difference for the combined dataset of location and banking scenarios. 11 outliers are not visible in this Figure in order to allow a close-up comparison of group differences. Median values are indicated by the lines. Condition differences are highlighted with an asterisk have a p-value lower than 0.025 in the pairwise comparison of conditions as we apply Bonferroni correction to account for multiple testing.

deletion of their data. Instead, they should trigger a rational thinking process where participants use the information available to them to make their decision. Nevertheless, we expect that their choice should reflect their attitudes, i.e. people who are more concerned about their privacy should be more likely to delete their data than those who are not too much concerned about their privacy. Given the ratings on privacy concern, literacy, trust in the chatbot and usage of chabots we aim to investigate whether participants choices indeed reflect their values. We do not include uncertainty as a predictor as higher levels of uncertainty do not necessarily lead to increased intention to delete personal data. Uncertainty only serves as a measure whether we were successful in triggering a rational thinking process. Thus, given the information they have available we expect participants to perform a rational cost-benefit analysis and based on the outcome decide upon deletion or saving of the data. For the analysis, we divide the dataset into a group of people exposed to the second control condition, one group exposed to the alternative condition and one group exposed to the reconsider condition. For each of the three datasets we perform binary logistic regression with the intention to delete as the dependent variable and using scenario and control variables as predictors. We find that for people exposed to the second control condition none of the predictors shows a significant effect on peoples' deletion behaviour. For people exposed to the reconsider condition, we find that only scenario influences the deletion behaviour significantly ($\chi^2(1) = 4.58, p = 0.03$). The odds of deleting the data is 2.7 times higher

Parameter		Estimate	Std. Error	z value	PR(> $\|z\|$)
(Intercept)		-4.31	2.76	-1.6	0.12
Scenario	Location	-1.4	0.61	-2.32	0.02 *
Privacy Literacy		0.08	0.58	0.14	0.89
Privacy Concern		1.78	0.78	2.28	0.02 *
Trust in the Chatbot		-1.06	0.65	-1.63	0.1
Usage	"less than once a month"	1.8	0.99	1.8	0.07
	"2-4 times a month"	2.56	0.98	2.62	0.008 **
	"more than once a week"	2.42	1.04	2.34	0.02 *

Table 3.28: Outcome of the binary logistic regression on participants' deletion behaviour when being exposed to the alternative condition. The model resulting in a minimal Akaike Information Criterion (AIC) value includes scenario, privacy literacy, privacy concern, trust in the chatbot and usage of chatbots. We find that people in the location scenario as opposed to the banking scenario are significantly less likely to delete their data when exposed to the alternative condition. Moreover, people reporting increased privacy concerns are significantly more likely to have their data deleted. In addition, participants who are using chatbots more often as opposed to those who have never used chatbots before are significantly more likely to delete their data.

for people exposed to the location scenario than for those in the banking scenario. However, for the dataset containing only people exposed to the alternative condition, we find that deletion behaviour is significantly influenced by scenario, concern and usage of chatbots. Moreover, there is a tendency that trust in the chatbot influences the deletion behaviour. Table 3.28 and Table 3.29 show the model coefficients in detail and results of the statistical test. We find that for people exposed to the alternative condition the odds of deleting their data is 5.9 times higher for people reporting increased privacy concerns. Moreover, the odds of deleting their data decreases by $(1 - 0.24) * 100\% = 76\%$ for participants exposed to the location scenario compared to the ones in the banking scenario. While not significant, the odds of deletion also decreases by $(1 - 0.35) * 100\% = 65\%$ for participants reporting higher trust in the chatbot. Lastly, the more often people use chatbots compared to the ones not having used chatbots at all the odds of deleting data increases drastically by a factor of 6 for usage less than once a month, a factor of 13 for usage 2-4 times a month and by a factor of 11 for usage more than once a week.

Given the results presented above, we can conclude that participants were most likely to act upon their attitudes and thus perform informed decision-making based on the information available to them when being exposed to the alternative condition. Certainly, this is not the case whenever people are nudged into saving their data in the second control condition. Moreover, we found that only the scenario and thus possibly the difference in information sensitivity influences peoples' behaviour in the reconsider condition. One possible explanation could be that while we do not perceive a strong nudging towards deletion in the reconsider condition it could nevertheless outweigh participants' attitudes and values. Moreover, participants exposed to the reconsider

	Df	Chisq	Pr($>|Chisq|$)
Scenario	1	5.9	0.015 *
Privacy Literacy	1	0.02	0.89
Privacy Concern	1	5.6	0.018 *
Trust in the Chatbot	1	3.46	0.06
Usage	3	10.14	0.017 *

Table 3.29: Type III ANOVA of the binary logistic regression model on participants' deletion behaviour when being exposed to the alternative condition. We find that scenario, privacy concern and usage have significant influence on the deletion behaviour in the alternative condition. Moreover, there is a tendency for trust to have a significant impact on participants' deletion behaviour.

condition can simply answer by stating "yes" or "no". Manual analysis of the transcripts revealed that a majority of participants exposed to the alternative condition did not succeed in answering upon storage or deletion the first time they were asked. Instead of responding with the keywords "save" or "delete", most participants stated "yes" or "no" and thus the chatbot asked to clarify their response. In comparison, a majority of participants exposed to the reconsider condition did succeed in answering the question upon deletion the first time they were asked as "yes" and "no" were accepted answers. Moreover, in Section 3.6.3.4 we found that while participants exposed to the reconsider condition took slightly longer to make their decision, only the alternative condition significantly slowed down the decision-making process compared to the second control condition. Based on those observations, we conclude that the chatbot asking for clarification and significantly longer response times for participants in the alternative condition, led them to consider their attitudes before finally deciding upon storage or deletion of their data. In contrast, participants in the reconsider condition were not nudged into deletion of their data but were not significantly slower in their decision-making. Thus, they might have been more likely to agree to the chatbots' suggestion and did not fully react upon their attitudes.

Chapter 4

Discussion

4.1 Discussion on the Effect of Cognitive Forcing Strategies on Peoples' Behaviour

Our results demonstrate that cognitive forcing strategies can significantly affect user behaviour in the context of Conversational User Interface (CUI). When participants were exposed to the cognitive forcing strategies they were more likely to ask for deletion of their data compared to a nudging approach towards permanent data storage. Thereby, reconsideration and explicit decision-making between alternatives led to a significant increase in deletion requests (see Section 3.6.3 for detailed results). Instead, introducing an additional time delay to make participants think more carefully about their decision did not significantly affect behaviour. However, we hypothesized that all cognitive forcing strategies increase the likelihood of deletion (Hypothesis 1). However, we found that this is only true for the alternative and reconsider condition. Therefore, we conclude that Hypothesis 1 is only partly supported depending on the cognitive forcing strategy. Our second control condition was designed such that individuals make decisions that may favour interests of data collectors by nudging users to agree to permanent storage of their data. Similar strategies can be found in interface design. Such dark patterns are common for the design of cookie banners, where colours are chosen in a way that users are more likely to consent to the cookies rather than decline Bermejo Fernandez et al. [2021]. We found that most of the participants (over 80%) agree to store their data in our nudging control condition. One could expect that an offer to delete data might lead to opposite behaviour i.e. general deletion of data. However, our results suggest that people were not nudged into the deletion of their data. Instead when exposed to the reconsider condition, 51% to 74% of participants wanted to have their data deleted while 49% and 26% of participants wanted to have their data stored. When offering alternatives, only 24% to 48% of participants asked for deletion. Thus, we conclude that the strategies do not nudge people into strictly privacy-preserving behaviour but might rather support rational risk-benefit assessment. While we found that scenarios affect peoples' intention to delete their data, the cognitive forcing strategies showed some robustness across scenarios. Nevertheless, we did observe some differences between scenarios in decisions to delete data. Distinct information sensitivity levels of the two scenarios are not able to explain differences in behaviour satisfactorily. Previous research suggests that banking information is perceived as more sensitive than location data [Schomakers et al., 2019]. Indeed, we found that when participants were exposed to the alternative condition they were more likely to delete their data in the banking scenario compared to the location scenario. However, opposite behaviour is observable for participants exposed to the reconsider condition. Here, people were more likely to have their data saved in the banking scenario compared to the location scenario. Therefore, further research is needed to explore possible interaction effects of information sensitivity and cognitive forcing strategies. While we assumed that the banking scenario is perceived as more sensitive as the location scenario, we did not assess how people perceived the level of sensitivity regarding the information they shared. Future research could therefore include measures for the perceived level of information sensitivity in order to draw connections to peoples' intention to delete. Furthermore, we only evaluated peoples' behaviour and perceptions in two different contexts, location and banking. Thus, further research is needed to explore the impact of cognitive forcing strategies on other types of personal information which is accessed by conversational agents e.g. health data or information on purchases. Moreover, other factors than information sensitivity and context may affect peoples' decision-making when exposed to cognitive forcing strategies. When experiencing Cognitive Ease i.e. when being in a good mood or liking the chatbots' appearance or speech assistants' voice, people are more likely to be superficial in their thinking. Thus, cognitive forcing strategies might vary in their effectiveness depending on participants' cognitive state. Moreover, we believe that the shift in behaviour seen in this study is transferable to speech-based conversational agents due to the similarity between chatbots and speech assistants. Nevertheless, the effect of cognitive forcing strategies could be further explored in the context of speech assistants and embodied conversational agents.

4.2 Discussion on the Effect of Cognitive Forcing Strategies on Peoples' Perceptions

Additionally to peoples' behaviour, we investigated the effect of cognitive forcing strategies on peoples' perceptions. First, we were interested whether participants experienced a feeling of uncertainty as we believe that this could positively affect people in making rational decisions. Dual-Process Theory suggests that there are two ways of thinking: a fast, intuitive one, which is governed by heuristics and a slow, effortful one, that is governed by reason. Usually, people rely on intuitive thinking. Effortful thinking is particularly engaged in states of doubt, surprise and uncertainty. Therefore, our cognitive forcing strategies were meant to support users in transitioning from fast and intuitive to slow and effortful thinking – a transition that is accompanied by uncertainty. We used two subjective measures to assess the level of perceived uncertainty. First, we investigated feelings associated with uncertainty, in particular, the feeling of "Fear". Literature suggests that "Fear" is related to the uncertainty of future events [Smith and Ellsworth, 1985]. Hypothesis 2a stated that participants will report higher levels of "Fear" when exposed to cognitive forcing strategies than to control conditions. However, our results did not provide support for Hypothesis 2a as we did not find any effect of the cognitive forcing strategies on "Fear". Subjective assessment of emotions has been critically discussed in literature [Ciuk et al., 2015]. While surveys on emotions allow inexpensive and efficient measurement, they might not correctly capture underlying psychological processes. The responses might be rationalized or biased whenever feelings are not socially desirable [Ciuk et al., 2015]. Ciuk et al. conclude that self-reports can serve as valid indicators but may be error-prone when compared to physiological assessment of emotions. Our assessment of "Fear" is limited by the fact that participants are asked to report on feelings they had during the interaction with the chatbot retrospectively. This is problematic as present feelings at the time of filling out the survey might outweigh the perceived level of "Fear" during the interaction. Therefore, future research should assess participants' emotions at the time of the interaction e.g. by using physiological measures such as pupil dilation, brain activity of skin conductance [Martin, 2014]. Moreover, one could argue that "Fear" is a rather extreme feeling and connected to specific factors e.g. spiders, flying or deep water rather than to interactions with conversational interfaces. Instead, subjective measurements of feelings like discomfort or uneasiness could provide further insights into peoples' perceptions of cognitive forcing functions.

As a second subjective measure of uncertainty, we used a rating scale dedicated to privacy uncertainty. We found limited impact of cognitive forcing strategies on this uncertainty scale. Measures on collection and use uncertainty were only influenced by scenario with higher uncertainty values reported by people in the location scenario. This is surprising as it does not reflect the assumed differences in information sensitivity across scenarios. If the banking scenario was indeed perceived as more sensitive, we would expect to see higher uncertainty ratings for the banking scenario. Instead, we observe the opposite effect. A possible explanation would be that participants were more likely to believe that we could access their location data compared to their credit card number. Thus, they might have been more uncertain about future handling of their location data compared to their banking data. While we saw similar impact of scenario on protection uncertainty, we found that protection uncertainty was significantly influenced by our tested cognitive forcing strategies. However, we did not only find increased uncertainty ratings for the cognitive forcing strategies but also for the nudging approach when compared to the chatbot offering further assistance. This suggests that simply priming participants towards data collection can induce uncertainty. As an objective measure, reaction times can provide useful information whether participants made their decision based on System 1 or System 2 thinking. Fast decision-making implies relying on intuitive thinking while slower response times imply engagement of System 2 and therefore more conscious and controlled decision-making. Our results suggest that there is a tendency that considering alternatives and reconsideration can slow down the decision-making process when compared to the nudging approach. Considering the results on uncertainty and reaction times together, the nudging approach might not trigger uncertainty at the time of decision-making but participants might become aware of their fast, intuitive and possibly thoughtless decision at the time of filling out the survey. Instead, when exposed to cognitive forcing strategies, higher levels of perceived uncertainty might be a result of engaging in effortful thinking during the chatbot interaction. In consequence, we find that behaviour varies significantly when exposed to the suggested strategies. Thereby, our assessment of uncertainty is again limited by the fact that we measure uncertainty not in the moment of decision-making but after the interaction ended. Therefore, further research could assess uncertainty at the time of decision-making by using physiological or other objective measures such as brain activity, galvanic skin response or speech and linguistic patterns [Martin, 2014]. Moreover, cognitive load can be used as an indicator of whether people find themselves in a state of System 1 or System 2 thinking and has been frequently assessed via eye-tracking and pupil dilation [Kahneman, 2011]. Hypothesis 2b and 2 c stated that participants will report higher levels of uncertainty and will show longer reaction times when exposed to the cognitive forcing strategies compared to the control conditions. Based on our results and the discussion above, we conclude that our findings partly support Hypothesis 2b and 2c as participants reported higher levels of protection uncertainty and showed a tendency to take more time when exposed to some of the tested strategies.

Our third hypothesis stated that cognitive forcing strategies can enhance the level of perceived privacy. However, we found only little support for Hypothesis 3. Only when comparing between strategies that showed a significant shift in behaviour and controlling for other variables, do our results show that reconsideration significantly affects privacy perceptions. It might be possible that increased levels of uncertainty could outweigh peoples' perception of privacy. Moreover, cognitive forcing strategies are supposed to lead to rational risk assessment. Participants might become aware of general data collection risks that can moderate their current perception of privacy. Further research should therefore assess the impact of cognitive forcing strategies on perceived risks and benefits to better understand how they constitute to the perception of privacy. Moreover, longitudinal privacy studies are necessary to investigate the influence of the tested strategies on privacy perception over time. We believe that cognitive forcing strategies have the potential to create a long-lasting effect on privacy perceptions rather than what can be assessed right after the experiment. Because they can trigger rational and effortful thinking, they can make people aware of possible privacy risks and at the same time provide possibilities for self-protection when sharing information with CUIs. While engagement in System 2 thinking might first create unfamiliar situations as individuals are used to relying on fast and intuitive thinking, the strategies could help to increase privacy and trust over time. Research in the medical field has shown that training on cognitive biases is effective to overcome those biases in daily life [Lambe et al., 2016]. Thus, we believe that the more often people interact with CUIs that apply cognitive forcing strategies, the more likely they are to conduct a rational risk-benefit assessment in situations where cognitive forcing strategies are not available. Importantly, different to nudging people, cognitive forcing strategies allow to rationally weigh risks and benefits and thus depend on the individuals' knowledge and general attitudes rather than what is seen as favourable from a governmental, industrial or public point of view.

Previous research found that cognitive forcing strategies can impact the usability and acceptability of a system [Buçinca et al., 2021]. Similarly, we hypothesized that cognitive forcing strategies might negatively impact peoples' perception of usability (Hypothesis 4). However, we did not find any differences in usability ratings between conditions nor did we receive negative comments on the chatbot or its performance. Thus we can reject Hypothesis 4 and conclude that cognitive forcing strategies are applicable in the context of CUIs without negatively impacting the usability of the system. However, participants were paid and only exposed once to the cognitive forcing strategies. Thus, it is necessary to assess the long-term usability of cognitive forcing strategies in real-life scenarios.

4.3 Cognitive Forcing Strategies and the GDPR

Further, we aim to discuss our results in light of current legal policies, i.e. the GDPR. The GDPR aims to explicitly strengthen an individual's right to have control over his or her data. van Ooijen and Vrabec refer to individual control as "the extent to which an individual is consciously aware of the situation and has the conscious intention and the ability to start, stop or maintain a situation." In our main study, the chatbot took a proactive approach to support participants in their right to erasure following the principles of Conversational Privacy. At the same time, the cognitive forcing strategies were supposed to slow down the decision-making process and support users to make conscious and rational decisions. Thus, we believe that our tested strategies can enhance the feeling of being in control of ones' data and thus follow the principle of the data protection regulation to strengthen an individuals' right to control. Furthermore, a typical consent-based data processing pipeline can be divided into three main stages van Ooijen and Vrabec [2019]. The first stage refers to the information receiving stage. Here, data collectors are in charge of presenting users with the necessary information about data processing and consequences. The right to information builds the basis of exercising control and providing valid consent in the second stage, i.e. the approval and primary usage stage. The GDPR defines consent as "any freely given, specific, informed and unambiguous indication of the data subjects' wishes by which he or she by statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her." [European Commission, 2016]. This requires consent to be active and thus prohibits inactivity or pre-ticked boxes. Lastly, van Ooijen and Vrabec identify a third stage that concerns the secondary uses of data. In this stage, control can be exercised by the individual by exerting their right to access and their right to erasure. Conversational Privacy suggests communicating information about privacy and data processing using dialogue form. While our experiments were not designed to display information or allow participants to exert their right to information, they showed that Conversational Privacy can be used to ask for consent or allow participants to exert their right of erasure. First, the pilot study examined whether participants' behaviour and perceptions differ in regard to data collection. Here, participants could provide consent to having their data accessed by the chatbot or entering the information by themselves. Giving approval and allowing for primary usage corresponds to the second stage of the consent-based data processing pipeline. It is important to notice that we did not provide users with the information required by GDPR for valid consent. Thus, further research could build on the work of Harkous et al. and investigate how to use Conversational Privacy to effectively communicate privacy policies in CUIs. This would allow users to pass through all stages in the pipeline without the need to switch modalities. Second, in the main study, we focused on the third stage of the consent-based data processing pipeline. While we supported people in exerting their right to erasure, some of our conditions, i.e. the second control, slow down and alternative condition, explicitly asked users for their agreement to permanent storage of their data. Again, we acknowledge that this consent is not necessarily valid in the light of GDPR as we did not provide further information about data processing. However, we found that most of the participants gave consent when being exposed to the nudging approach or slow down condition. Thus, one needs to keep in mind that presenting privacy-related information in dialogue form is not free of dark patterns and nudging approaches that favor data collectors rather than data subjects. Instead, dialogue designers and developers need to consider behavioural and cognitive processes when communicating privacy-related information. However, our results show that peoples' behaviour differs whether they were asked to consent or were given explicit options to choose from. While participants were given the option to disagree to further storage in all scenarios, only the alternative condition actively mentioned both options i.e. "Do you want me to delete your data from this interaction or have it saved for future interaction?" and asked participants to explicitly decide between storage and deletion. Consent and choice can be seen as two distinct concepts with consent referring to an individuals' agreement and choice referring to providing individuals with options to choose from [Sawicki, 2012]. In the medical context, a paradigm shift is visible moving from informed consent to informed choice as informed choice allows for partnership between patients and doctors and shared decision-making instead of agreeing to a preset procedure [Weinstein, 2005, Sawicki, 2012]. Weinstein argues that informed consent should not be fully abandoned but applied when choice "may not seem relevant or appropriate". However, he states that whenever "evidence is weak or because the patient has to decide what a better quality of life means for them" informed choice is preferred. We argue that because privacy is highly subjective, cultural and context-dependent, informed choice might be more suitable than the current informed consent regulations. However, it needs to be ensured that necessary information for the decision-making process is easily accessible and understandable. Moreover, we believe that CUIs are uniquely capable of promoting informed choice regarding data processing and moving towards shared decision-making similar to what can be seen in a doctor-patient relationship. Because of their natural and human-like conversational style they can provide guidance and support users with relevant information in their decision-making process. Moreover, informed choice would allow users to become more literate about privacy and data processing and at the same time allow them to base their choice on their own values and preferences. Our results suggest that when the chatbot provided participants with alternatives i.e. asking whether they want to delete or save their data, participants were more likely to react regarding their attitudes and values. For example, participants reporting increased privacy concerns were more likely to have their data deleted. Therefore, we believe that informed choice is applicable to data protection and privacy in the context of CUIs and should be further explored in the future. While we gave only little information by stating that the data will be saved for future interactions, future research could investigate the impact of information provision on informed choice. Moreover, research is needed to explore the applicability of CUIs to provide guidance and incorporate a shared decision-making process. Although informed choice will not be suitable for every user and context, we believe that it can increase privacy and transparency in contexts with sensitive data or whenever third parties are involved.

As already stated above, informed consent is defined under GDPR as "any freely given, specific, informed and unambiguous indication of the data subjects' wishes by which he or she by statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her." [European Commission, 2016]. However, there is an ongoing discussion around the concept of informed consent, informed choice and informed decision-making and their measurements in the medical field [Marteau et al., 2001, Ghanouni et al., 2016]. In 1990, Merz and Fischhoff published an article on "Informed consent does not mean rational consent: Cognitive limitations on decision-making". They argue that "if information is intended to allow patients to make informed, knowledgable, rational decisions in their own best interest, then under the current legal definition of consent, the notion of consent as informed is a legal fiction." By that, they make it clear that providing information is not enough to ensure informed and rational decision-making. First, it is unclear how detailed information needs to be to make sure that patients can perform informed decision-making. This might include whether to use other forms to display information e.g. pictures or other media [Merz and Fischhoff, 1990]. Second, as discussed in Sections 2.4.2 and 2.5, cognitive processes can limit peoples' decision-making skills making the decision subject to heuristics and biases. Generally, the rational decision-making model consists of multiple steps to reach a decision (1) Identify the problem (2) establish decision criteria (3) weigh decision criteria (4) generate alternatives (5) evaluate the alternatives (6) choose the best alternative [Lumen Learning, 2022]. We can easily transfer that to the decision participants had to make in our experiment. First, the chatbot identified the problem by asking whether participants wanted to have their data stored for future interactions and more generally giving them a chance to exert control over their data. To make a decision based on rationality, one had to establish decision criteria including personal interests, values and preferences into the process e.g. interest in using the service a second time or privacy concerns. In the next step, participants would need to weigh the identified decision criteria according to their importance. Further, alternatives had to be created and evaluated. This could have been asking for deletion of data or terminating the process. Among the alternative options, the best would have been selected and adopted. While rational decision-making is not always favourable or available to humans in everyday life due to cognitive limitations and time constraints, we argue that with the increasing number of devices accessing personal data and threatening peoples' privacy, decisions around those topics are important and worth engaging in more rational decision-making. Cognitive forcing strategies are one way to support users in making more rational decisions. Our results showed that when providing users with alternatives, they were more likely to act upon their values and preferences. Moreover, we believe that CUIs can support users in each step of the rational decision-making model by considering alternatives, overcoming cognitive biases and speeding up the process. First, they can make users aware of problems arising around data protection and thus initiate the decision-making process as demonstrated in our experiment. Second, they can support users in establishing decision criteria by using human-like and natural conversations and learning user interests, values and preferences regarding their personal information over time. While research has shown that generating alternatives is most problematic and difficult for humans [Nutt, 2004], CUIs might be capable of generating alternatives in a fast, effective and encompassing way and evaluating them according to users' decision criteria. This can make it easy for users to choose among two or three alternatives that have been found to match the decision criteria best by the CUI.

Chapter 5

Conclusion

In this thesis, we explored debiasing strategies known from other research disciplines in the context of privacy self-management in Conversational User Interface (CUI). We based our approach on the principle of Conversational Privacy where privacy-related information is communicated to the user in dialogue form. In particular, we investigated the effect of cognitive forcing strategies on peoples' behaviour and perceptions when disclosing personal information to a chatbot. Our cognitive forcing strategies were designed to support people to exert their right to erasure after having disclosed personal information to the chatbot while at the same time promoting rational decision-making. Our results show that confronting people with alternatives of either having their data saved or deleted or a simple offer to delete their data significantly changes peoples' behaviour. Participants were more likely to ask for deletion and tended to report higher privacy values. Moreover, we found support for the thesis that cognitive forcing strategies can help to transition between a default fast and intuitive thinking state to a more effortful and rational thinking state in interactions with chatbots. Thus, people were more likely to act upon their attitudes when exposed to the cognitive forcing strategies. Therefore, cognitive forcing strategies might be applied to address the discrepancy between users attitudes and behaviour known as the Privacy Paradox. Previous studies have shown that although people report privacy concerns, they are not likely to engage in privacy preserving behaviour. Instead when exposed to cognitive forcing strategies and given the options to save or delete their data, participants who reported increased general privacy concerns where more likely to ask for deletion. This suggests that cognitive forcing strategies can help users to overcome cognitive biases and support them in their decision-making process while considering individual attitudes rather than intuition. Moreover, our strategies are easily accessible, deployable and scalable in the context of CUI. Thus, they can support users in their privacy self-management without impacts on usability. We believe that the investigated strategies are a first step towards easily accessible privacy protection strategies that promote rational thinking in the context of CUIs. Privacy self-management strategies are not sufficient on their own and collaboration between all the actors in the field is needed to abandon privacy-intrusive practices and support people in exerting their rights. However, we believe that future research on debiasing strategies can be highly beneficial for the privacy research field and its applications.

Appendix A

Dialogue Trees Pilotstudy

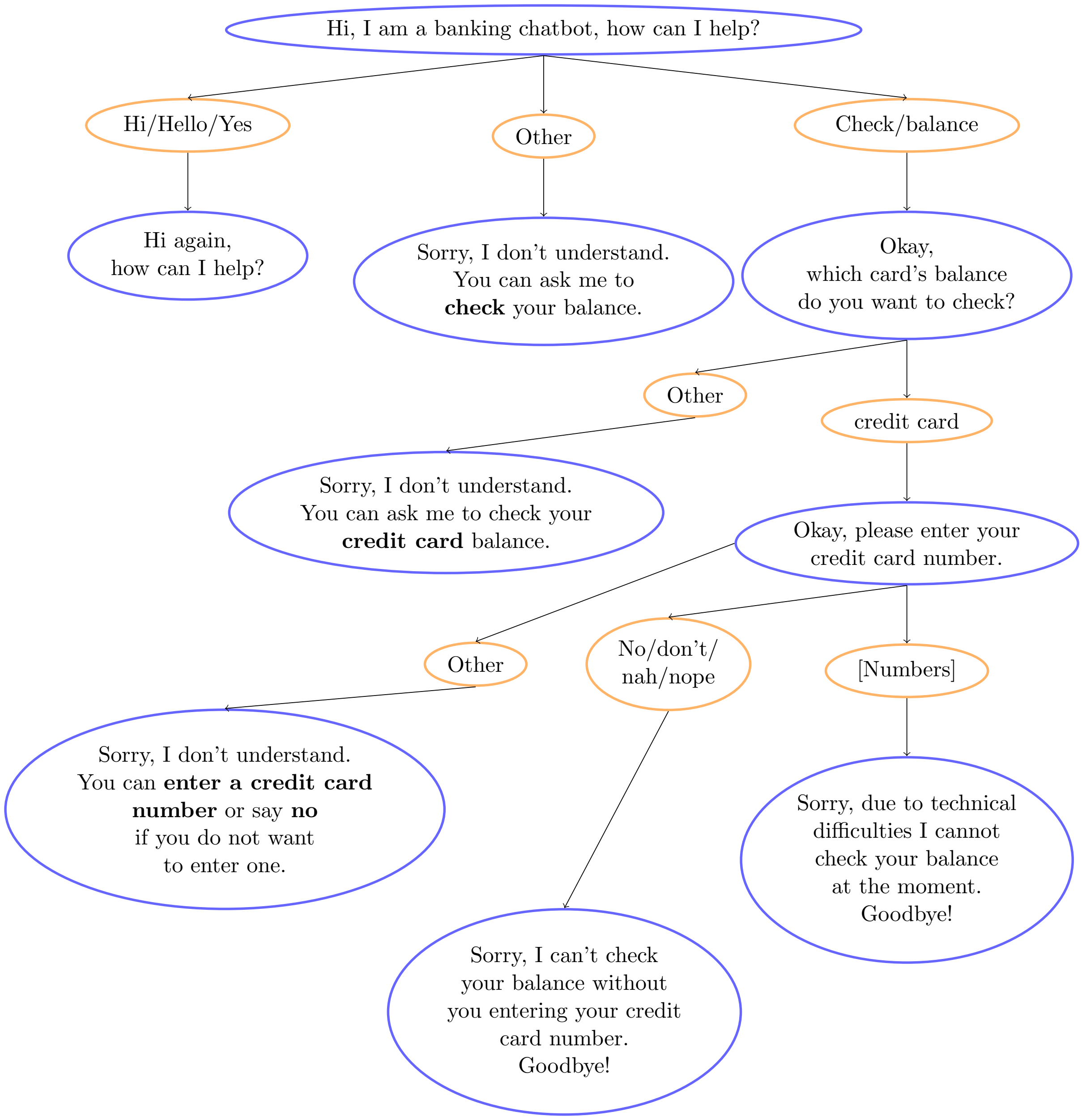

Hi, I am a banking chatbot, how can I help?
Hi/Hello/Yes
Other
Check/balance
Hi again, how can I help?
Sorry, I don't understand. You can ask me to **check** your balance.
Okay, which card's balance do you want to check?
Other
credit card
Sorry, I don't understand. You can ask me to check your **credit card** balance.
Okay, please enter your credit card number.
Other
No/don't/nah/nope
[Numbers]
Sorry, I don't understand. You can **enter a credit card number** or say **no** if you do not want to enter one.
Sorry, due to technical difficulties I cannot check your balance at the moment. Goodbye!
Sorry, I can't check your balance without you entering your credit card number. Goodbye!

Figure A.1: Dialogue Tree for the banking scenario in the enter condition, blue circle show the chatbot, orange circles show the possible inputs for the user.

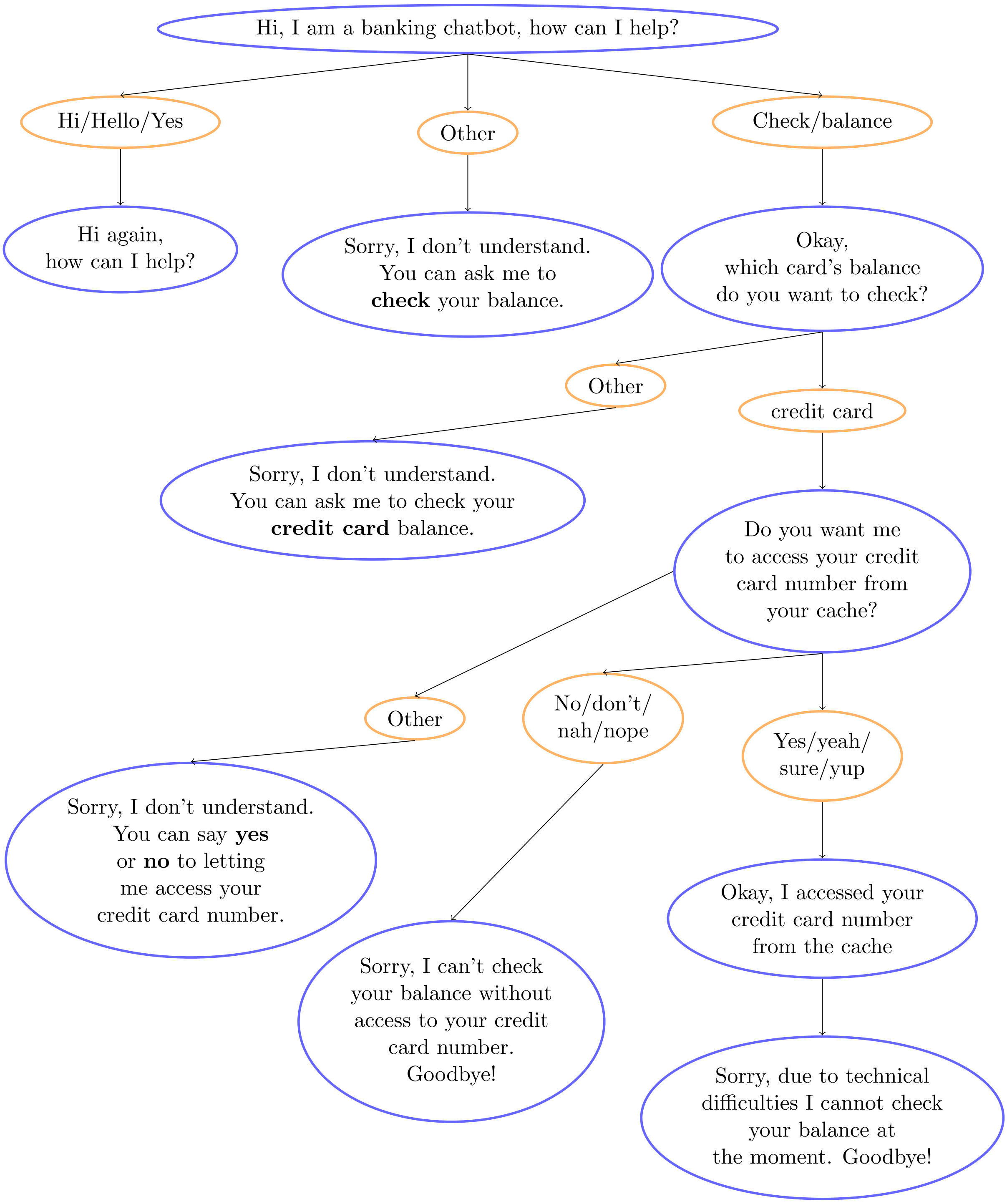

Hi, I am a banking chatbot, how can I help?
Hi/Hello/Yes
Other
Check/balance
Hi again, how can I help?
Sorry, I don't understand. You can ask me to **check** your balance.
Okay, which card's balance do you want to check?
Other
credit card
Sorry, I don't understand. You can ask me to check your **credit card** balance.
Do you want me to access your credit card number from your cache?
Other
No/don't/ nah/nope
Yes/yeah/ sure/yup
Sorry, I don't understand. You can say **yes** or **no** to letting me access your credit card number.
Okay, I accessed your credit card number from the cache
Sorry, I can't check your balance without access to your credit card number. Goodbye!
Sorry, due to technical difficulties I cannot check your balance at the moment. Goodbye!

Figure A.2: Dialogue Tree for the banking scenario in the access condition, blue circle show the chatbot, orange circles show the possible inputs for the user.

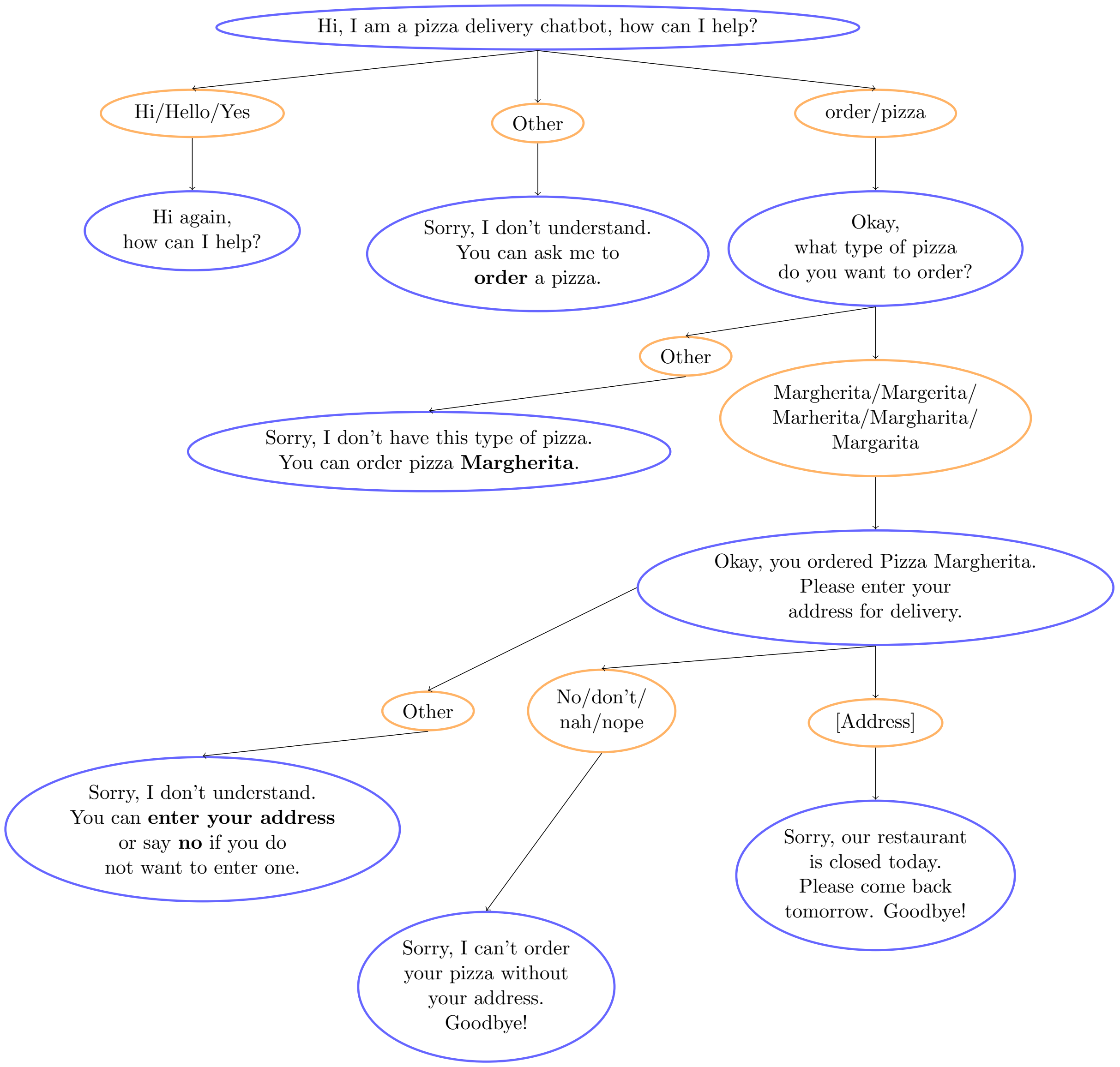


Figure A.3: Dialogue Tree for the location scenario in the enter condition, blue circle show the chatbot, orange circles show the possible inputs for the user.

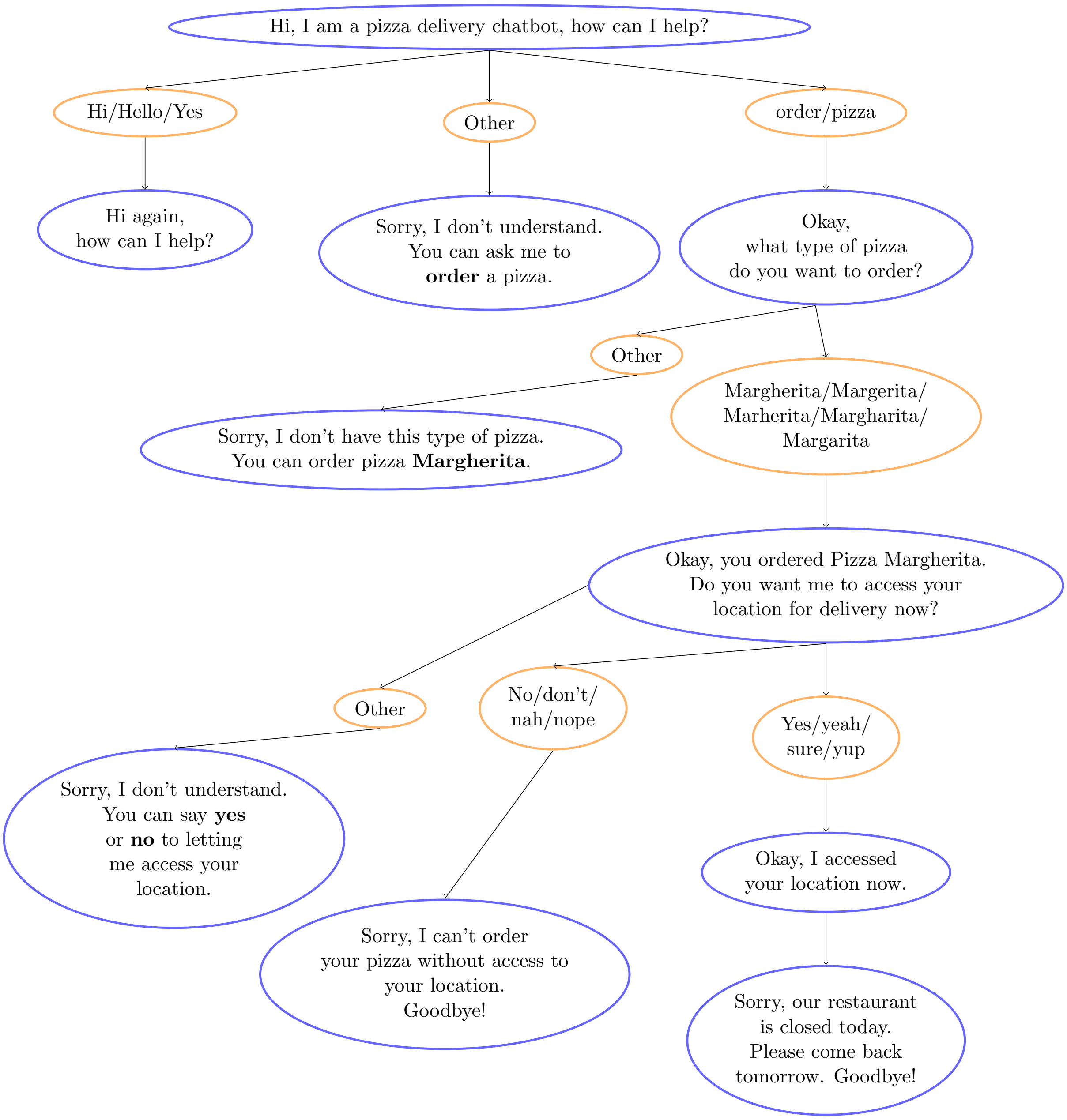

Hi, I am a pizza delivery chatbot, how can I help?
Hi/Hello/Yes
Other
order/pizza
Hi again, how can I help?
Sorry, I don't understand. You can ask me to **order** a pizza.
Okay, what type of pizza do you want to order?
Other
Margherita/Margerita/ Marherita/Margharita/ Margarita
Sorry, I don't have this type of pizza. You can order pizza **Margherita**.
Okay, you ordered Pizza Margherita. Do you want me to access your location for delivery now?
Other
No/don't/ nah/nope
Yes/yeah/ sure/yup
Sorry, I don't understand. You can say **yes** or **no** to letting me access your location.
Okay, I accessed your location now.
Sorry, I can't order your pizza without access to your location. Goodbye!
Sorry, our restaurant is closed today. Please come back tomorrow. Goodbye!

Figure A.4: Dialogue Tree for the location scenario in the access condition, blue circle show the chatbot, orange circles show the possible inputs for the user.

Appendix B

Survey Pilotstudy

Appendix B shows the questionnaire items used in the pilotstudy. We show the items together with the corresponding construct they were supposed to measure. Those were not presented to the participants as it might have influenced their responses. The three screening questions which are here shown after one another were distributed over the questionnaire.

Screening Questions

1a. It is important that you pay attention to the statements. Please agree by choosing 'strongly agree' from the options.

strongly disagree ○—○—○—○—○ strongly agree

1b. To ensure that you are paying attention, please select 'strongly disagree' from the options

strongly disagree ○—○—○—○—○ strongly agree

1c. I recognize the importance of paying attention to the questions in the questionnaire. Please select 'strongly agree' to confirm your agreement.

strongly disagree ○—○—○—○—○ strongly agree

PANAS-X (Fear, Serenity, Surprise + Frustration)

Indicate to what extent you have felt this way while interacting with the chatbot

2a. frustrated very slightly or not at all ○—○—○—○—○ extremely

2b. afraid very slightly or not at all ○—○—○—○—○ extremely

2c. calm very slightly or not at all ○—○—○—○—○ extremely

2d. frightened very slightly or not at all ○—○—○—○—○ extremely

2e. surprised very slightly or not at all ○—○—○—○—○ extremely

2f. nervous very slightly or not at all ○—○—○—○—○ extremely

2g. relaxed very slightly or not at all ○—○—○—○—○ extremely

2h. jittery very slightly or not at all ○—○—○—○—○ extremely

2i. amazed very slightly or not at all ○—○—○—○—○ extremely

2j. scared very slightly or not at all ○—○—○—○—○ extremely

2k. at ease very slightly or not at all ○—○—○—○—○ extremely

2l. astonished very slightly or not at all ○—○—○—○—○ extremely

2m. shaky very slightly or not at all ○—○—○—○—○ extremely

Physicians' Reaction to Uncertainty

Indicate to what extent you agree with the following statements. While I was interacting with the chatbot...

3a. I felt anxious when sharing personal data with the chatbot

strongly disagree ○—○—○—○—○ strongly agree

3b. I found the uncertainty involved in the chatbot interaction disconcerting

strongly disagree ○—○—○—○—○ strongly agree

3c. Uncertainty in the chatbot interaction makes me uneasy

strongly disagree ○—○—○—○—○ strongly agree

3d. I was quite comfortable with the uncertainty in the chatbot interaction

strongly disagree ○—○—○—○—○ strongly agree

3e. The uncertainty during the chatbot interaction troubled me

strongly disagree ○—○—○—○—○ strongly agree

3f. When I was uncertain of the data sharing process I imagined all sorts of bad scenarios - unallowed data collection, data misuse, unsafe data protection

strongly disagree ○—○—○—○—○ strongly agree

3g. I fear privacy breaches due to sharing personal information

strongly disagree ○—○—○—○—○ strongly agree

3h. I worry about privacy breaches when I do not know how the chatbot handles my personal information

strongly disagree ○—○—○—○—○ strongly agree

Collection, Use, Protection and Overall Uncertainty

4a. I was uncertain about what information will be collected

strongly disagree ○—○—○—○—○ strongly agree

4b. I was concerned about the amount of information that was collected by the chatbot

strongly disagree ○—○—○—○—○ strongly agree

4c. I was afraid the chatbot would collect more information than I was initially told

strongly disagree ○—○—○—○—○ strongly agree

4d. I was concerned that I will have to provide more information than I originally thought

strongly disagree ○—○—○—○—○ strongly agree

4e. I was concerned about how the chatbot provider would use the information that was recorded by the chatbot

strongly disagree ○—○—○—○—○ strongly agree

4f. I was uncertain about who would have access to the information that was

recorded

strongly disagree ○—○—○—○—○ strongly agree

4g. **I was worried that the information that was recorded will be shared with others**

strongly disagree ○—○—○—○—○ strongly agree

4h. **I was unsure if the information that was recorded might be misused**

strongly disagree ○—○—○—○—○ strongly agree

4i. **I was afraid that if given the chance the chatbot provider might profit by selling the information to someone else**

strongly disagree ○—○—○—○—○ strongly agree

4j. **I was concerned that the information that was collected will not be protected**

strongly disagree ○—○—○—○—○ strongly agree

4k. **I was uncertain about what the chatbot provider would do to ensure that the information collected was secure**

strongly disagree ○—○—○—○—○ strongly agree

4l. **I was unsure if the chatbot provider would effectively safeguard the information that was collected**

strongly disagree ○—○—○—○—○ strongly agree

4m. **Overall, I was unsure if the chatbot provider would safeguard my privacy**

strongly disagree ○—○—○—○—○ strongly agree

4n. **Overall, I was uncertain if the chatbot provider would be good at managing my private information**

strongly disagree ○—○—○—○—○ strongly agree

4o. **Overall, I was worried if my information would be safe with the chatbot provider**

strongly disagree ○—○—○—○—○ strongly agree

4p. **Overall, I was concerned that the chatbot provider might breach formal and**

informal privacy agreements

strongly disagree ○—○—○—○—○ strongly agree

Privacy Perception

Indicate to what extent you agree with the following statements.

5a. I think this chatbot shows concern for the privacy of its users

strongly disagree ○—○—○—○—○ strongly agree

5b. I feel safe when I send personal information to this chatbot

strongly disagree ○—○—○—○—○ strongly agree

5c. I think this chatbot abides by personal data protection laws

strongly disagree ○—○—○—○—○ strongly agree

5d. I think this chatbot only collects user personal data that are necessary for its activity

strongly disagree ○—○—○—○—○ strongly agree

5e. I think this chatbot respects the user's rights when obtaining personal information

strongly disagree ○—○—○—○—○ strongly agree

5f. I think that this chatbot will not provide my personal information to other companies without my consent

strongly disagree ○—○—○—○—○ strongly agree

Usability

6a. With this chatbot everything is easy to understand

strongly disagree ○—○—○—○—○ strongly agree

6b. This chatbot is simple to use, even when using it for the first time

strongly disagree ○—○—○—○—○ strongly agree

6c. It is easy to find the information I need from this chatbot

strongly disagree ○—○—○—○—○ strongly agree

6d. The structure and contents of this chatbot are easy to understand

strongly disagree ○—○—○—○—○ strongly agree

6e. It is easy to move within this chatbot

strongly disagree ○—○—○—○—○ strongly agree

6f. When I am using the chatbot I feel I am in control of what I can do

strongly disagree ○—○—○—○—○ strongly agree

6g. I would like to use the chatbot frequently

strongly disagree ○—○—○—○—○ strongly agree

Demand

7a. I found this interaction difficult

strongly disagree ○—○—○—○—○ strongly agree

7b. The chatbot was complex

strongly disagree ○—○—○—○—○ strongly agree

Trust in Chatbot and Chatbot Provider

8a. The chatbot will be trustworthy in handling my personal information

strongly disagree ○—○—○—○—○ strongly agree

8b. The chatbot will tell the truth and fulfil promises related to the information provided by me

strongly disagree ○—○—○—○—○ strongly agree

8c. I trust that the chatbot will keep my interests in mind when dealing with the information

strongly disagree ○—○—○—○—○ strongly agree

8d. Chatbots are in general predictable and consistent regarding the usage of the information

strongly disagree ○—○—○—○—○ strongly agree

8e. Chatbots are always honest with users when it comes to using the information that I would provide

strongly disagree ○—○—○—○—○ strongly agree

8f. The chatbot provider will be trustworthy in handling my personal information

strongly disagree ○—○—○—○—○ strongly agree

8g. The chatbot provider will tell the truth and fulfil promises related to the information provided by me

strongly disagree ○—○—○—○—○ strongly agree

8h. I trust that the chatbot provider will keep my interests in mind when dealing with the information

strongly disagree ○—○—○—○—○ strongly agree

8i. Chatbot providers are in general predictable and consistent regarding the usage of the information

strongly disagree ○—○—○—○—○ strongly agree

8j. Chatbot providers are always honest with users when it comes to using the information that I would provide

strongly disagree ○—○—○—○—○ strongly agree

Internet Users' Information Privacy Concerns (IUIPC)

Indicate to what extent you agree with the following statements.

9a. Consumer online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used and shared

strongly disagree ○—○—○—○—○ strongly agree

9b. Consumer control of personal information lies at the heart of consumer privacy

strongly disagree ○—○—○—○—○ strongly agree

9c. Companies seeking information online should disclose the way the data are collected, processed and used

strongly disagree ○—○—○—○—○ strongly agree

9d. A good consumer online privacy policy should have a clear and conspicuous disclosure

strongly disagree ○—○—○—○—○ strongly agree

9e. It usually bothers me when online companies ask me for personal information

strongly disagree ○—○—○—○—○ strongly agree

9f. When online companies ask me for personal information, I sometimes think twice before providing it

strongly disagree ○—○—○—○—○ strongly agree

9g. It bothers me to give personal information to so many online companies

strongly disagree ○—○—○—○—○ strongly agree

9h. I'm concerned that online companies are collecting too much personal information about me

strongly disagree ○—○—○—○—○ strongly agree

Privacy Literacy

10a. I am able to understand and evaluate questions about data protection and privacy on the Internet

strongly disagree ○—○—○—○—○ strongly agree

10b. I feel like I understand the most important things related to data protection

strongly disagree ○—○—○—○—○ strongly agree

10c. I know a lot about data protection and online privacy

strongly disagree ○—○—○—○—○ strongly agree

10d. I feel like I know more about data protection and privacy on the Internet than

most other people

strongly disagree ○—○—○—○—○ strongly agree

Uncertainty Avoidance

11a. Standardized work procedures are helpful

strongly disagree ○—○—○—○—○ strongly agree

11b. It is important to have instructions spelled out in detail so that I always know what I'm expected to do

very unimportant ○—○—○—○—○ very important

11c. It is important to closely follow instructions and procedures

very unimportant ○—○—○—○—○ very important

11d. Rules and regulations are important because they inform me of what is expected of me

very unimportant ○—○—○—○—○ very important

11e. Instructions for operations are important

very unimportant ○—○—○—○—○ very important

Questions about you:

12. Gender

○ Male

○ Female

○ Diverse

○ I prefer not to say

13. Age: ______________________________

14. Are you a native English Speaker?

○ Yes

○ No

15. **How often do you use chatbots on average?**

    ○ Not at all

    ○ Less than once a month

    ○ 2-4 times a month

    ○ more than once a week

16. **You can leave comments here (optional):** ______________________________

Thank you for your participation in our research! If you did not enter personal information during the interaction, no personal data of yours was accessed.

Appendix C

Explanatory Factor Analysis of the Scales used in the Pilot Study

	Factor1	Factor 2	Factor 3
afraid	0.772	0.104	0.132
frightened	0.766		-0.239
nervous	0.892	0.126	0.352
jittery	0.766	0.202	
scared	0.863		
shaky	0.801	0.220	
calm		0.567	
relaxed		0.567	
atease	0.224	0.482	0.242
surprised	0.307	0.550	-0.115
amazed	0.160	0.657	
astonished	0.658	0.293	
	Factor 1	Factor 2	Factor 3
SS loadings	4.561	1.821	0.294
Proportion Var	0.380	0.152	0.025
Cumulative Var	0.380	0.532	0.556

Table C.1: Explanatory Factor Analysis on PANAS-X scale using orthogonal rotation: $\chi^2(33) = 54.04, p = 0.0119$

	Factor 1	Factor 2
I felt anxious when sharing personal data with the chatbot	0.612	0.206
I found the uncertainty involved in the chatbot interaction disconcerting	0.614	0.168
Uncertainty in the chatbot interaction makes me uneasy	0.776	
I was quite comfortable with the uncertainty in the chatbot interaction	-0.144	0.440
When I was uncertain of the data sharing process I imagined all sorts of bad scenarios – unallowed data collection, data misuse, unsafe data protection	0.337	0.550
I fear privacy breaches due to sharing personal information	0.154	0.625
I worry about privacy breaches when I do not know how the chatbot handles my personal information	0.134	0.651
	Factor 1	Factor 2
SS loadings	1.530	1.384
Proportion Var	0.219	0.198
Cumulative Var	0.219	0.416

Table C.2: Explanatory Factor Analysis of the PRU scale (missing the item on “The uncertainty during the chatbot interaction troubled me” as it was missing from the location chatbot experiment) using oblique rotation: $\chi^2(8) = 8.53, p = 0.383$

	Factor 1	Factor 2	Factor 3
I was uncertain about what information will be collected			1.024
I was concerned about the amount of information that was collected by the chatbot	0.392	0.371	
I was afraid the chatbot would collect more information than I was initially told	0.748		
I was concerned that I will have to provide more information than I originally thought	0.455	0.244	
I was concerned about how the chatbot provider would use the information that was recorded by the chatbot	0.571	0.207	
I was uncertain about who would have access to the information that was recorded	0.442	0.113	0.220
I was worried that the information that was recorded will be shared with others	0.792	-0.187	0.129
I was unsure if the information that was recorded might be misused	0.275	0.374	0.154
I was afraid that if given the chance the chatbot provider might profit by selling the information to someone else	0.811		
I was concerned that the information that was collected will not be protected	0.382	0.423	
I was uncertain about what the chatbot provider would do to ensure that the information collected was secure	0.747		
I was unsure if the chatbot provider would effectively safeguard the information that was collected	0.400	0.498	-0.127
Overall, I was unsure if the chatbot provider would safeguard my privacy		0.433	0.317
Overall, I was uncertain if the chatbot provider would be good at managing my private information	-0.107	0.805	
Overall, I was worried if my information would be safe with the chatbot provider	0.414	0.405	
Overall, I was concerned that the chatbot provider might breach formal and informal privacy agreements		0.899	
	Factor 1	Factor 2	Factor 3
SS loadings	3.858	2.676	1.282
Proportion Var	0.241	0.167	0.080
Cumulative Var	0.241	0.408	0.489

Table C.3: Explanatory Factor Analysis of collection, use, protection and overall uncertainty scale using oblique rotation: $\chi^2(75) = 95.76, p = 0.0533$

	Factor 1
I think this chatbot shows concern for the privacy of its users	0.608
I feel safe when I send personal information to this chatbot	0.593
I think this chatbot abides by personal data protection laws	0.705
I think this chatbot only collects user personal data that are necessary for its activity	0.622
I think this chatbot respects the user's rights when obtaining personal information	0.657
I think that this chatbot will not provide my personal information to other companies without my consent	0.588
	Factor 1
SS loadings	2.382
Proportion Var	0.397

Table C.4: Explanatory Factor Analysis on privacy perception scale using oblique rotation: $\chi^2(9) = 15.95, p = 0.068$

	Factor 1
With this chatbot everything is easy to understand	0.545
This chatbot is simple to use, even when using it for the first time	0.622
It is easy to find the information I need from this chatbot	0.467
The structure and contents of this chatbot are easy to understand	0.435
It is easy to move within this chatbot	0.663
When I am using the chatbot I feel I am in control of what I can do	0.556
I would like to use the chatbot frequently	0.581
	Factor 1
SS loadings	2.177
Proportion Var	0.311

Table C.5: Explanatory Factor Analysis of usability scale using oblique rotation: $\chi^2(14) = 12.63, p = 0.556$

	Factor 1	Factor 2	Factor 3	Factor 4
The chatbot will be trustworthy in handling my personal information	-0.122	0.631	0.338	
The chatbot will tell the truth and fulfil promises related to the information provided by me			0.107	0.874
I trust that the chatbot will keep my interests in mind when dealing with the information	0.296	0.354	0.138	-0.122
Chatbots are in general predictable and consistent regarding the usage of the information	0.508	0.170		-0.156
Chatbots are always honest with users when it comes to using the information that I would provide	0.662	-0.240	0.253	
The chatbot provider will be trustworthy in handling my personal information	0.624			0.204
The chatbot provider will tell the truth and fulfil promises related to the information provided by me		0.811	-0.203	
I trust that the chatbot provider will keep my interests in mind when dealing with the information	0.220	0.189	0.230	
Chatbot providers are in general predictable and consistent regarding the usage of the information	-0.101		0.927	
Chatbot providers are always honest with users when it comes to using the information that I would provide	0.846	0.134	-0.234	
	Factor 1	Factor 2	Factor 3	Factor 4
SS loadings	1.971	1.328	1.217	0.878
Proportion Var	0.197	0.133	0.122	0.088
Cumulative Var	0.197	0.330	0.452	0.539

Table C.6: Explanatory Factor Analysis of trust in chatbot and trust in chatbot provider scale using oblique rotation: $\chi^2(11) = 17.74, p = 0.0878$

	Factor 1	Factor 2	Factor 3
Consumer online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used and shared	0.371	0.117	0.184
Consumer control of personal information lies at the heart of consumer privacy	0.778		-0.280
Companies seeking information online should disclose the way the data is collected, processed and used		1.055	
A good consumer online privacy policy should have a clear and conspicuous disclosure	0.275	0.138	0.106
It usually bothers me when online companies ask me for personal information	-0.150	-0.125	0.511
When online companies ask me for personal information, I sometimes think twice before providing it	0.526		
It bothers me to give personal information to so many online companies	0.258		0.357
I'm concerned that online companies are collecting too much personal information about me	0.442		0.465
	Factor 1	Factor 2	Factor 3
SS loadings	1.388	1.184	0.738
Proportion Var	0.173	0.148	0.092
Cumulative Var	0.173	0.321	0.414

Table C.7: Explanatory Factor Analysis of IUIPC scale using oblique rotation: $\chi^2(7) = 8.11, p = 0.323$

	Factor 1
I am able to understand and evaluate questions about data protection and privacy on the Internet	0.365
I feel like I understand the most important things related to data protection	0.412
I know a lot about data protection and online privacy	0.407
I feel like I know more about data protection and privacy on the Internet than most other people	0.566
	Factor 1
SS loadings	0.789
Proportion Var	0.197

Table C.8: Explanatory Factor Analysis of privacy literacy scale using oblique rotation: $\chi^2(2) = 7.4, p = 0.0247$

	Factor 1
Standardized work procedures are helpful	0.595
It is important to have instructions spelled out in detail so that I always know what I'm expected to do	0.493
It is important to closely follow instructions and procedures	0.492
Rules and regulations are important because they inform me of what is expected of me	0.499
Instructions for operations are important	0.435
	Factor 1
SS loadings	1.278
Proportion Var	0.256

Table C.9: Explanatory Factor Analysis of uncertainty avoidance scale using oblique rotation: $\chi^2(5) = 8.25, p = 0.143$

Appendix D

Dialogue Trees Mainstudy

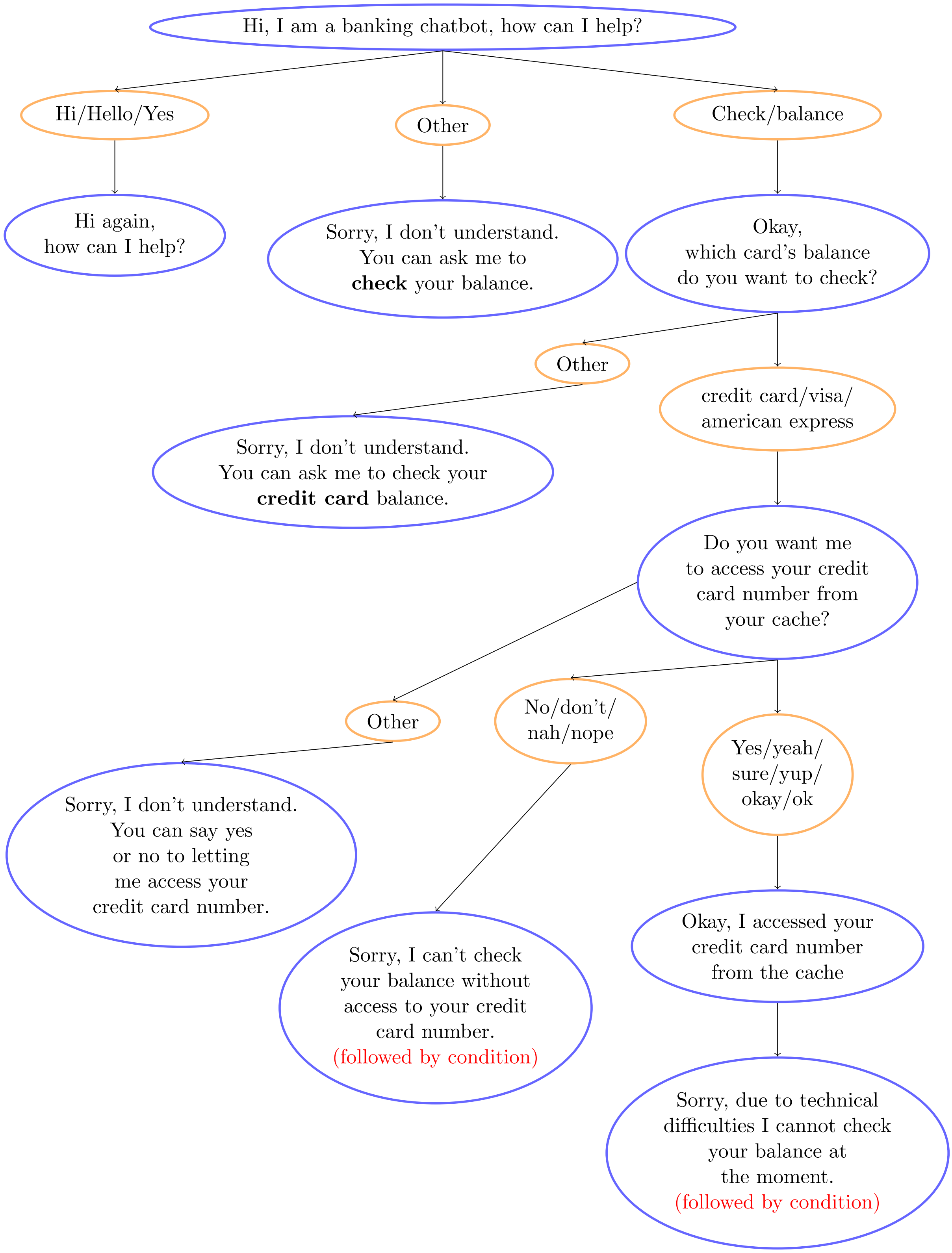

Hi, I am a banking chatbot, how can I help?
Hi/Hello/Yes
Other
Check/balance
Hi again, how can I help?
Sorry, I don't understand. You can ask me to **check** your balance.
Okay, which card's balance do you want to check?
Other
credit card/visa/ american express
Sorry, I don't understand. You can ask me to check your **credit card** balance.
Do you want me to access your credit card number from your cache?
Other
No/don't/ nah/nope
Yes/yeah/ sure/yup/ okay/ok
Sorry, I don't understand. You can say yes or no to letting me access your credit card number.
Sorry, I can't check your balance without access to your credit card number. (followed by condition)
Okay, I accessed your credit card number from the cache
Sorry, due to technical difficulties I cannot check your balance at the moment. (followed by condition)

Figure D.1: Dialogue Tree for the banking scenario for the main study, blue circles show the chatbot, orange circles show the possible inputs for the user.

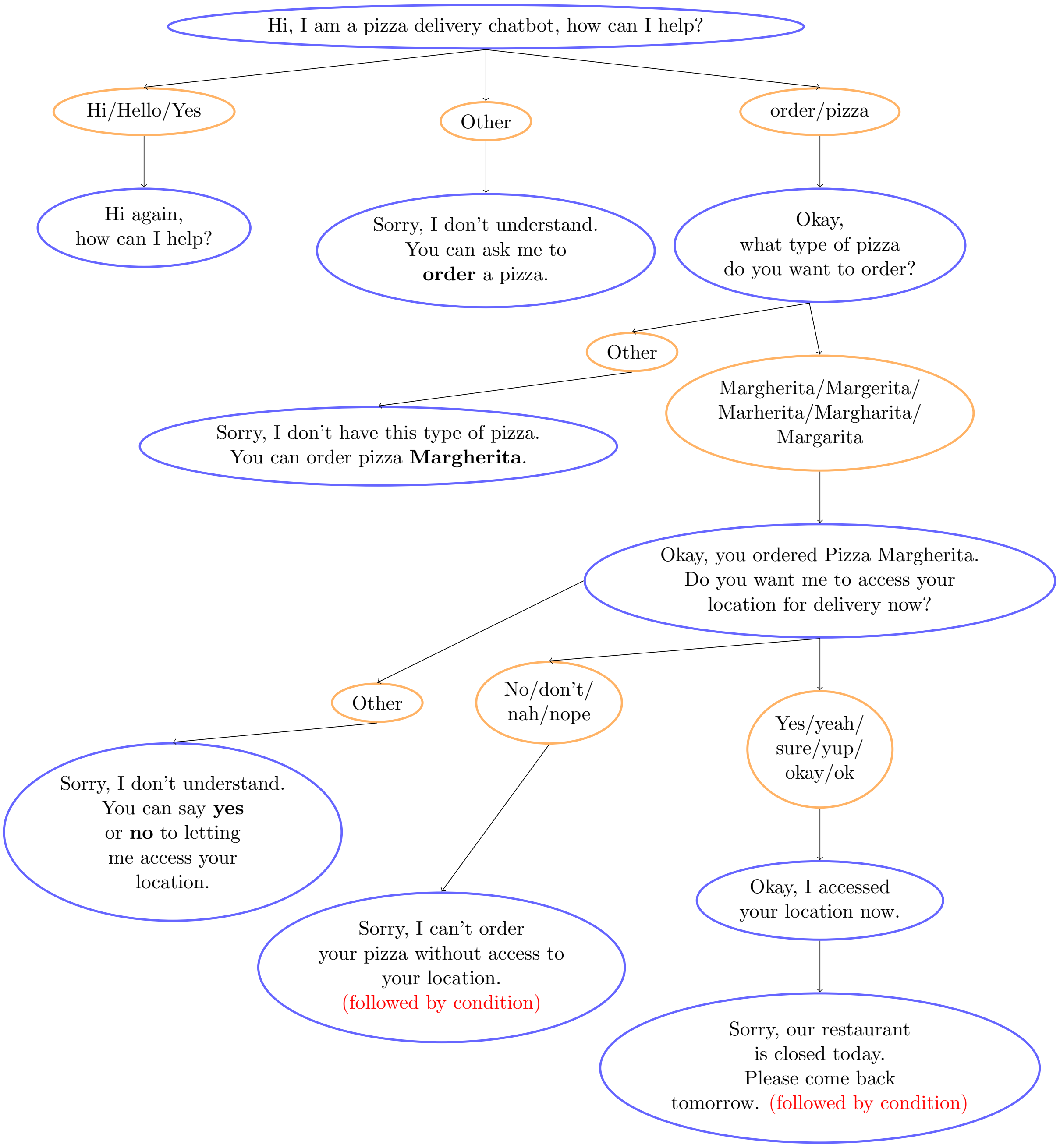

Hi, I am a pizza delivery chatbot, how can I help?
Hi/Hello/Yes
Other
order/pizza
Hi again, how can I help?
Sorry, I don't understand. You can ask me to **order** a pizza.
Okay, what type of pizza do you want to order?
Other
Margherita/Margerita/ Marherita/Margharita/ Margarita
Sorry, I don't have this type of pizza. You can order pizza **Margherita**.
Okay, you ordered Pizza Margherita. Do you want me to access your location for delivery now?
Other
No/don't/ nah/nope
Yes/yeah/ sure/yup/ okay/ok
Sorry, I don't understand. You can say **yes** or **no** to letting me access your location.
Okay, I accessed your location now.
Sorry, I can't order your pizza without access to your location. (followed by condition)
Sorry, our restaurant is closed today. Please come back tomorrow. (followed by condition)

Figure D.2: Dialogue Tree for the location scenario for the main study, blue circles show the chatbot, orange circles show the possible inputs for the user.

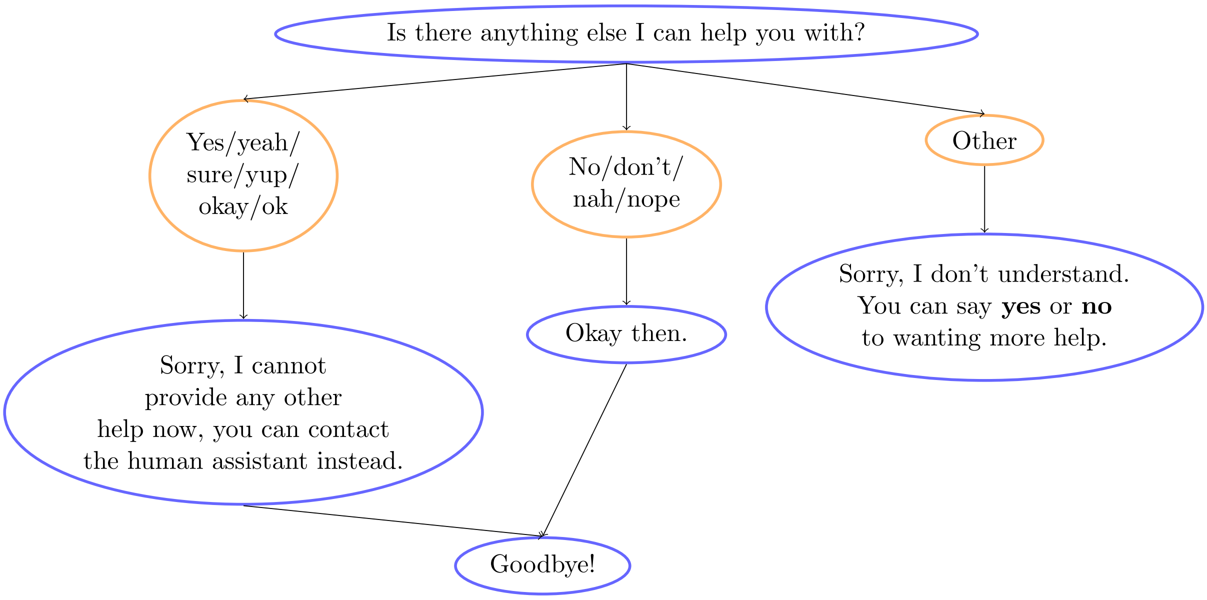

Is there anything else I can help you with?
Yes/yeah/ sure/yup/ okay/ok
No/don't/ nah/nope
Other
Sorry, I cannot provide any other help now, you can contact the human assistant instead.
Okay then.
Sorry, I don't understand. You can say **yes** or **no** to wanting more help.
Goodbye!

Figure D.3: Dialogue Tree for Control Condition 1 in the main study, blue circles show the chatbot, orange circles show the possible inputs for the user.

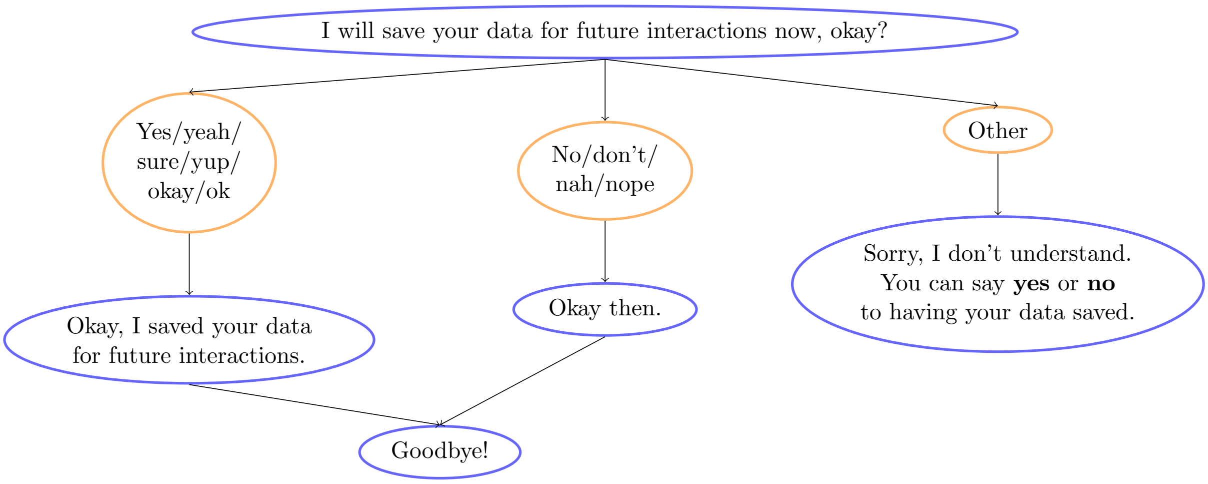

I will save your data for future interactions now, okay?
Yes/yeah/ sure/yup/ okay/ok
No/don't/ nah/nope
Other
Okay, I saved your data for future interactions.
Okay then.
Sorry, I don't understand. You can say **yes** or **no** to having your data saved.
Goodbye!

Figure D.4: Dialogue Tree for Control Condition 2 in the main study, blue circles show the chatbot, orange circles show the possible inputs for the user.

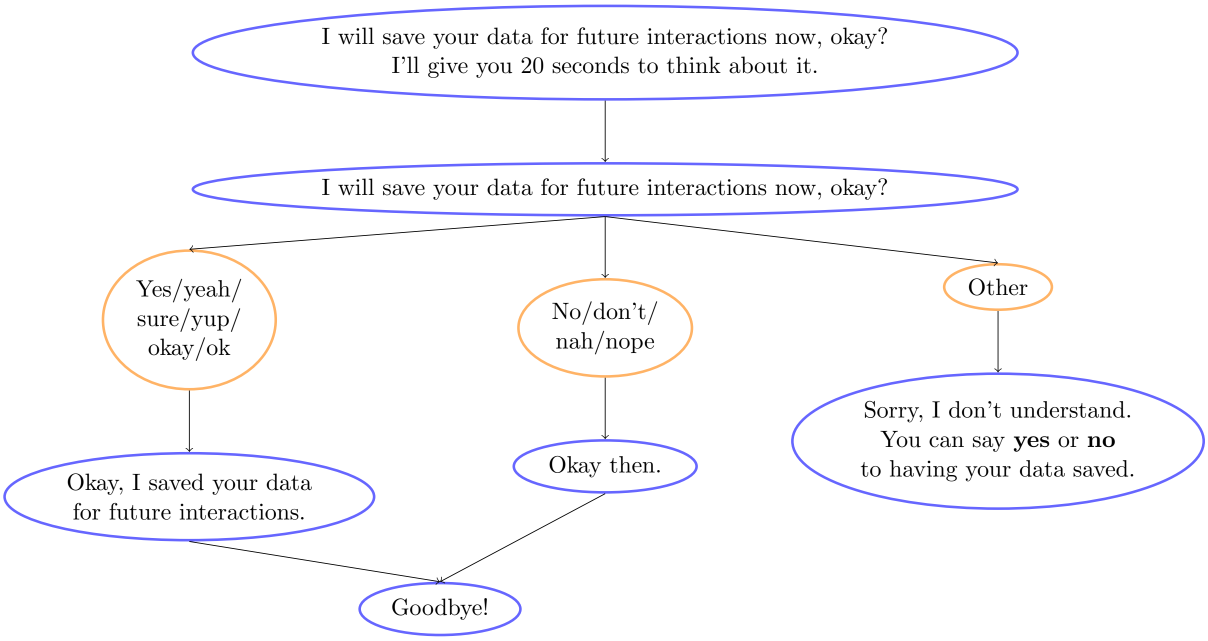

I will save your data for future interactions now, okay?
I'll give you 20 seconds to think about it.
I will save your data for future interactions now, okay?
Yes/yeah/ sure/yup/ okay/ok
No/don't/ nah/nope
Other
Okay, I saved your data for future interactions.
Okay then.
Sorry, I don't understand. You can say **yes** or **no** to having your data saved.
Goodbye!

Figure D.5: Dialogue Tree for Slow Down Condtion in the main study, blue circles show the chatbot, orange circles show the possible inputs for the user.

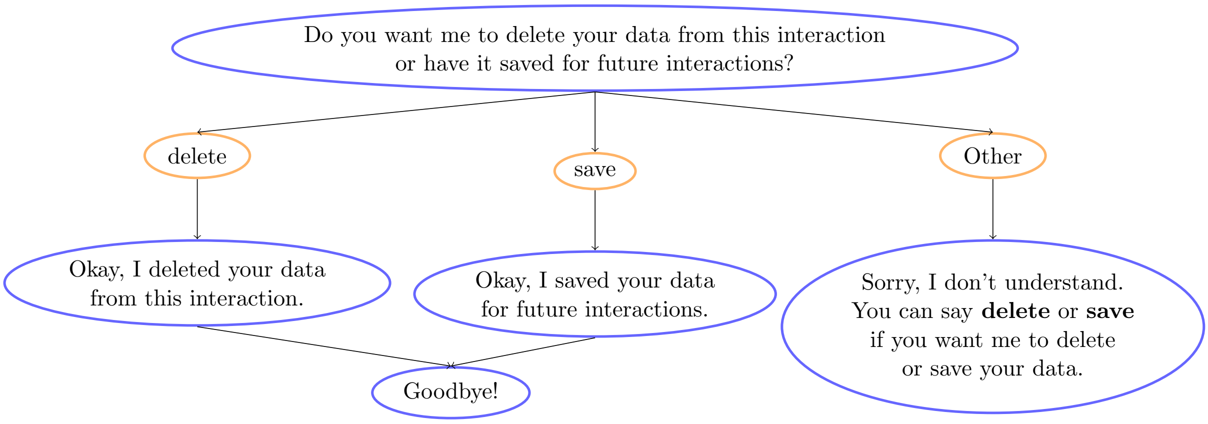

Do you want me to delete your data from this interaction or have it saved for future interactions?
delete
save
Other
Okay, I deleted your data from this interaction.
Okay, I saved your data for future interactions.
Sorry, I don't understand. You can say **delete** or **save** if you want me to delete or save your data.
Goodbye!

Figure D.6: Dialogue Tree for Alternative Condition in the main study, blue circles show the chatbot, orange circles show the possible inputs for the user.

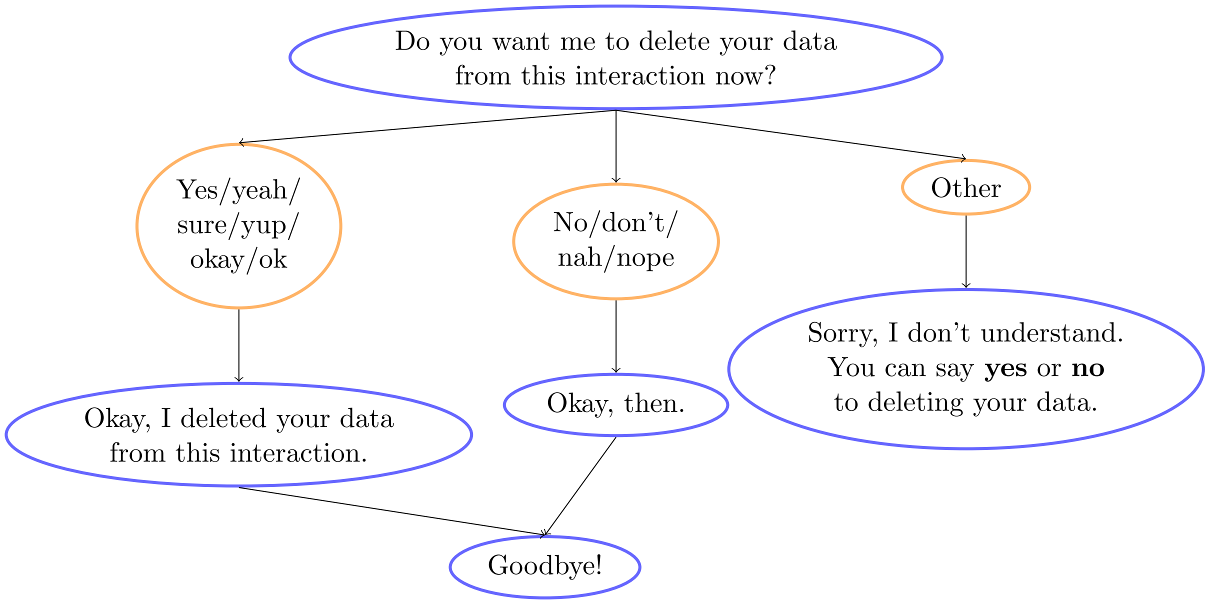

Do you want me to delete your data
from this interaction now?
Yes/yeah/
sure/yup/
okay/ok
No/don't/
nah/nope
Other
Okay, I deleted your data
from this interaction.
Okay, then.
Sorry, I don't understand.
You can say **yes** or **no**
to deleting your data.
Goodbye!

Figure D.7: Dialogue Tree for Reconsider Condition in the main study, blue circles show the chatbot, orange circles show the possible inputs for the user.

Appendix E

Survey Mainstudy

Appendix E shows the questionnaire items used in the mainstudy. We show the items together with the corresponding construct they were supposed to measure. Those were not presented to the participants as it might have influenced their responses. The three screening questions which are here shown after one another were distributed over the questionnaire.

Screening Questions

1a. It is important that you pay attention to the statements. Please agree by choosing 'strongly agree' from the options.

strongly disagree ○—○—○—○—○ strongly agree

1b. To ensure that you are paying attention, please select 'strongly disagree' from the options

strongly disagree ○—○—○—○—○ strongly agree

1c. I recognize the importance of paying attention to the questions in the questionnaire. Please select 'strongly agree' to confirm your agreement.

strongly disagree ○—○—○—○—○ strongly agree

PANAS-X (Fear + Frustration)

Indicate to what extent you have felt this way while interacting with the chatbot

2a. frustrated very slightly or not at all ○—○—○—○—○ extremely

2b. afraid very slightly or not at all ○—○—○—○—○ extremely

2c. frightened very slightly or not at all ○—○—○—○—○ extremely

2d. nervous very slightly or not at all ○—○—○—○—○ extremely

2e. jittery very slightly or not at all ○—○—○—○—○ extremely

2f. scared very slightly or not at all ○—○—○—○—○ extremely

2g. shaky very slightly or not at all ○—○—○—○—○ extremely

Collection, Use, Protection and Overall Uncertainty

Indicate to what extent you agree with the following statements. While I was interacting with the chatbot...

3a. I was uncertain about what information will be collected

strongly disagree ○—○—○—○—○ strongly agree

3b. I was concerned about the amount of information that was collected by the chatbot

strongly disagree ○—○—○—○—○ strongly agree

3c. I was afraid the chatbot would collect more information than I was initially told

strongly disagree ○—○—○—○—○ strongly agree

3d. I was concerned that I will have to provide more information than I originally thought

strongly disagree ○—○—○—○—○ strongly agree

3e. I was concerned about how the chatbot provider would use the information that was recorded by the chatbot

strongly disagree ○—○—○—○—○ strongly agree

3f. I was uncertain about who would have access to the information that was recorded

strongly disagree ○—○—○—○—○ strongly agree

3g. I was worried that the information that was recorded will be shared with others

strongly disagree ○—○—○—○—○ strongly agree

3h. I was unsure if the information that was recorded might be misused

strongly disagree ○—○—○—○—○ strongly agree

3i. I was afraid that if given the chance the chatbot provider might profit by selling the information to someone else

strongly disagree ○—○—○—○—○ strongly agree

3j. I was concerned that the information that was collected will not be protected

strongly disagree ○—○—○—○—○ strongly agree

3k. I was uncertain about what the chatbot provider would do to ensure that the information collected was secure

strongly disagree ○—○—○—○—○ strongly agree

3l. I was unsure if the chatbot provider would effectively safeguard the information that was collected

strongly disagree ○—○—○—○—○ strongly agree

3m. Overall, I was unsure if the chatbot provider would safeguard my privacy

strongly disagree ○—○—○—○—○ strongly agree

3n. Overall, I was uncertain if the chatbot provider would be good at managing my private information

strongly disagree ○—○—○—○—○ strongly agree

3o. Overall, I was worried if my information would be safe with the chatbot provider

strongly disagree ○—○—○—○—○ strongly agree

3p. Overall, I was concerned that the chatbot provider might breach formal and informal privacy agreements

strongly disagree ○—○—○—○—○ strongly agree

Privacy Perception

Indicate to what extent you agree with the following statements.

4a. I think this chatbot shows concern for the privacy of its users

strongly disagree ○—○—○—○—○ strongly agree

4b. I feel safe when I send personal information to this chatbot

strongly disagree ○—○—○—○—○ strongly agree

4c. I think this chatbot abides by personal data protection laws

strongly disagree ○—○—○—○—○ strongly agree

4d. I think this chatbot only collects user personal data that are necessary for its activity

strongly disagree ○—○—○—○—○ strongly agree

4e. I think this chatbot respects the user's rights when obtaining personal information

strongly disagree ○—○—○—○—○ strongly agree

4f. I think that this chatbot will not provide my personal information to other companies without my consent

strongly disagree ○—○—○—○—○ strongly agree

Usability

5a. With this chatbot everything is easy to understand

strongly disagree ○—○—○—○—○ strongly agree

5b. This chatbot is simple to use, even when using it for the first time

strongly disagree ○—○—○—○—○ strongly agree

5c. It is easy to find the information I need from this chatbot

strongly disagree ○—○—○—○—○ strongly agree

5d. The structure and contents of this chatbot are easy to understand

strongly disagree ○—○—○—○—○ strongly agree

5e. It is easy to move within this chatbot

strongly disagree ○—○—○—○—○ strongly agree

5f. When I am using the chatbot I feel I am in control of what I can do

strongly disagree ○—○—○—○—○ strongly agree

5g. I would like to use the chatbot frequently

strongly disagree ○—○—○—○—○ strongly agree

Trust in the Chatbot

6a. The chatbot will be trustworthy in handling my personal information

strongly disagree ○—○—○—○—○ strongly agree

6b. The chatbot will tell the truth and fulfil promises related to the information provided by me

strongly disagree ○—○—○—○—○ strongly agree

6c. I trust that the chatbot will keep my interests in mind when dealing with the information

strongly disagree ○—○—○—○—○ strongly agree

6d. Chatbots are in general predictable and consistent regarding the usage of the information

strongly disagree ○—○—○—○—○ strongly agree

6e. Chatbots are always honest with users when it comes to using the information that I would provide

strongly disagree ○—○—○—○—○ strongly agree

Internet Users' Information Privacy Concerns (IUIPC)

Indicate to what extent you agree with the following statements.

7a. Consumer online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used and shared

strongly disagree ○—○—○—○—○ strongly agree

7b. Consumer control of personal information lies at the heart of consumer privacy

strongly disagree ○—○—○—○—○ strongly agree

7c. Companies seeking information online should disclose the way the data are collected, processed and used

strongly disagree ○—○—○—○—○ strongly agree

7d. A good consumer online privacy policy should have a clear and conspicuous disclosure

strongly disagree ○—○—○—○—○ strongly agree

7e. When online companies ask me for personal information, I sometimes think twice before providing it

strongly disagree ○—○—○—○—○ strongly agree

7f. It bothers me to give personal information to so many online companies

strongly disagree ○—○—○—○—○ strongly agree

7g. I'm concerned that online companies are collecting too much personal information about me

strongly disagree ○—○—○—○—○ strongly agree

Privacy Literacy

8a. I am able to understand and evaluate questions about data protection and privacy on the Internet

strongly disagree ○—○—○—○—○ strongly agree

8b. I feel like I understand the most important things related to data protection

strongly disagree ○—○—○—○—○ strongly agree

8c. I know a lot about data protection and online privacy

strongly disagree ○—○—○—○—○ strongly agree

8d. I feel like I know more about data protection and privacy on the Internet than most other people

strongly disagree ○—○—○—○—○ strongly agree

Uncertainty Avoidance

9a. Standardized work procedures are helpful

strongly disagree ○—○—○—○—○ strongly agree

9b. It is important to have instructions spelled out in detail so that I always know what I'm expected to do

very unimportant ○—○—○—○—○ very important

9c. It is important to closely follow instructions and procedures

very unimportant ○—○—○—○—○ very important

9d. Rules and regulations are important because they inform me of what is expected of me

very unimportant ○—○—○—○—○ very important

9e. Instructions for operations are important

very unimportant ○—○—○—○—○ very important

Questions about you:

10. Gender

○ Male

○ Female

○ Diverse

○ I prefer not to say

11. Age: ________________________

12. Are you a native English Speaker?

○ Yes

○ No

13. How often do you use chatbots on average?

○ Not at all

○ Less than once a month

○ 2-4 times a month

○ more than once a week

14. You can leave comments here (optional): ________________________

Thank you for your participation in our research! If you did not enter personal information during the interaction, no personal data of yours was accessed.

Appendix F

Explanatory Factor Analysis of the Scales used in the Main Study

	Factor1	Factor 2
afraid		0.799
frightened	0.5	0.362
nervous	0.825	
jittery	0.652	0.206
scared	0.255	0.655
shaky	0.370	0.543
	Factor 1	Factor 2
SS loadings	1.561	1.537
Proportion Var	0.260	0.256
Cumulative Var	0.260	0.516

Table F.1: Explanatory Factor Analysis on PANAS-X scale using oblique rotation: $\chi^2(4) = 2.42, p = 0.66$

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
I was uncertain about what information will be collected	0.335		0.385		0.164
I was concerned about the amount of information that was collected by the chatbot	-0.156	0.357			0.528
I was afraid the chatbot would collect more information than I was initially told	0.820				
I was concerned that I will have to provide more information than I originally thought		0.752			-0.147
I was concerned about how the chatbot provider would use the information that was recorded by the chatbot	-0.129	0.670			-0.147
I was uncertain about who would have access to the information that was recorded	0.745		-0.184		0.239
I was worried that the information that was recorded will be shared with others	0.683				
I was unsure if the information that was recorded might be misused	0.731				
I was afraid that if given the chance the chatbot provider might profit by selling the information to someone else	0.113			0.887	
I was concerned that the information that was collected will not be protected	0.203	0.421			0.168
I was uncertain about what the chatbot provider would do to ensure that the information collected was secure	0.640	0.211			-0.217

I was unsure if the chatbot provider would effectively safeguard the information that was collected	0.751	0.138			
Overall, I was unsure if the chatbot provider would safeguard my privacy			0.998		
Overall, I was uncertain if the chatbot provider would be good at managing my private information	0.307	-0.121			0.555
Overall, I was worried if my information would be safe with the chatbot provider	0.599				
Overall, I was concerned that the chatbot provider might breach formal and informal privacy agreements	0.285	0.472			
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
SS loadings	3.955	1.637	1.205	0.836	0.797
Proportion Var	0.247	0.102	0.075	0.052	0.050
Cumulative Var	0.247	0.350	0.425	0.477	0.527

Table F.2: Explanatory Factor Analysis of collection, use, protection and overall uncertainty scale using oblique rotation: $\chi^2(50) = 59.81, p = 0.161$

	Factor 1	Factor 2
I think this chatbot shows concern for the privacy of its users		0.949
I feel safe when I send personal information to this chatbot	0.514	0.144
I think this chatbot abides by personal data protection laws	0.517	
I think this chatbot only collects user personal data that are necessary for its activity	0.746	-0.153
I think this chatbot respects the user's rights when obtaining personal information	0.557	0.102
I think that this chatbot will not provide my personal information to other companies without my consent	0.648	
	Factor 1	Factor 2
SS loadings	1.820	0.962
Proportion Var	0.303	0.160
Cumulative Var	0.303	0.464

Table F.3: Explanatory Factor Analysis on privacy perception scale using oblique rotation: $\chi^2(4) = 6.45, p = 0.168$

	Factor 1	Factor 2	Factor 3
With this chatbot everything is easy to understand	0.460		0.161
This chatbot is simple to use, even when using it for the first time		1.032	
It is easy to find the information I need from this chatbot	0.719	-0.109	
The structure and contents of this chatbot are easy to understand			0.405
It is easy to move within this chatbot			0.772
When I am using the chatbot I feel I am in control of what I can do	0.526	0.207	
I would like to use the chatbot frequently	0.688		
	Factor 1	Factor 2	Factor 3
SS loadings	1.491	1.139	0.793
Proportion Var	0.213	0.163	0.113
Cumulative Var	0.213	0.376	0.489

Table F.4: Explanatory Factor Analysis of usability scale using oblique rotation: $\chi^2(3) = 3.03, p = 0.387$

	Factor 1
The chatbot will be trustworthy in handling my personal information	0.669
The chatbot will tell the truth and fulfil promises related to the information provided by me	0.565
I trust that the chatbot will keep my interests in mind when dealing with the information	0.615
Chatbots are in general predictable and consistent regarding the usage of the information	0.473
Chatbots are always honest with users when it comes to using the information that I would provide	0.690
	Factor 1
SS loadings	1.844
Proportion Var	0.369

Table F.5: Explanatory Factor Analysis of trust in the chatbot scale using oblique rotation: $\chi^2(5) = 7.65, p = 0.177$

	Factor 1	Factor 2	Factor 3
Consumer online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used and shared	0.619		
Consumer control of personal information lies at the heart of consumer privacy	0.358	0.194	
Companies seeking information online should disclose the way the data is collected, processed and used	0.504	-0.116	
A good consumer online privacy policy should have a clear and conspicuous disclosure			0.998
When online companies ask me for personal information, I sometimes think twice before providing it	0.528		
It bothers me to give personal information to so many online companies		0.997	
I'm concerned that online companies are collecting too much personal information about me	0.528		-0.102
	Factor 1	Factor 2	Factor 3
SS loadings	1.322	1.049	1.023
Proportion Var	0.189	0.150	0.146
Cumulative Var	0.189	0.339	0.485

Table F.6: Explanatory Factor Analysis of IUIPC scale using oblique rotation: $\chi^2(3) = 5.74, p = 0.125$

	Factor 1
I am able to understand and evaluate questions about data protection and privacy on the Internet	0.550
I feel like I understand the most important things related to data protection	0.539
I know a lot about data protection and online privacy	0.602
I feel like I know more about data protection and privacy on the Internet than most other people	0.524
	Factor 1
SS loadings	1.230
Proportion Var	0.307

Table F.7: Explanatory Factor Analysis of privacy literacy scale using oblique rotation: $\chi^2(2) = 13.36, p = 0.00126$

	Factor 1	Factor 2
Standardized work procedures are helpful	0.769	-0.124
It is important to have instructions spelled out in detail so that I always know what I'm expected to do	0.392	0.165
It is important to closely follow instructions and procedures	0.382	0.209
Rules and regulations are important because they inform me of what is expected of me	0.793	
Instructions for operations are important	-0.118	1.053
	Factor 1	Factor 2
SS loadings	1.534	1.199
Proportion Var	0.307	0.240
Cumulative Var	0.307	0.547

Table F.8: Explanatory Factor Analysis of uncertainty avoidance scale using oblique rotation: $\chi^2(1) = 0.58, p = 0.445$

Bibliography

Alessandro Acquisti. Nudging privacy: The behavioral economics of personal information. *IEEE Security Privacy*, 7(6):82–85, 2009. doi: 10.1109/MSP.2009.163.

Alessandro Acquisti and Jens Grossklags. Uncertainty, ambiguity and privacy. In *Fourth Workshop on the Economics of Information Security (WEIS05) 2005*, pages 2–3, 2005.

Eleni Adamopoulou and Lefteris Moussiades. An overview of chatbot technology. In Ilias Maglogiannis, Lazaros Iliadis, and Elias Pimenidis, editors, *Artificial Intelligence Applications and Innovations*, pages 373–383, Cham, 2020. Springer International Publishing. ISBN 978-3-030-49186-4.

AJ Burt. Conversational ui: it's not just chat bots and voice assistants – a ux case study. `https://uxdesign.cc/conversational-ui-its-not-just-chat-bots-and-voice-assistants-case-study-cb1865da306a`, 2022. [Online; accessed 2022-03-03].

Hirotogu Akaike. *Information Theory and an Extension of the Maximum Likelihood Principle*, pages 199–213. Springer New York, New York, NY, 1998. ISBN 978-1-4612-1694-0. doi: 10.1007/978-1-4612-1694-0_15.

Sameh Al-Natour, Hasan Cavusoglu, Izak Benbasat, and Usman Aleem. An Empirical Investigation of the Antecedents and Consequences of Privacy Uncertainty in the Context of Mobile Apps. *Information Systems Research*, 31(4):1037–1063, December 2020. ISSN 1047-7047, 1526-5536. doi: 10.1287/isre.2020.0931.

Allied Market Research. Conversational ai market by component (platform and services [support and maintenance, training and consulting, and system integration]), deployment (cloud and on-premises), type (iva and chatbots), and technology (machine learning, deep learning, nlp, and automated speech recognition), and end user (bfsi, retail & e-commerce, healthcare & life science, telecom, media & entertainment, and others): Global opportunity analysis and industry forecast, 2021-2030. `https://www.alliedmarketresearch.com/conversational-ai-market-A13682`, 2021. [Online; accessed 2022-03-03].

Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. Your location has been shared 5,398 times! a field study on mobile app privacy nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 787–796, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450331456. doi: 10.1145/2702123.2702210.

Rebecca Balebako and Lorrie Cranor. Improving app privacy: Nudging app developers to protect user privacy. *IEEE Security Privacy*, 12(4):55–58, 2014. doi: 10.1109/MSP.2014.70.

John A. Bargh, M. Chen, and Lj Burrows. Automaticity of social behavior: direct effects of trait construct and stereotype-activation on action. *Journal of personality and social psychology*, 71 2:230–44, 1996.

Susan B. Barnes. A privacy paradox: Social networking in the United States. *First Monday*, September 2006. ISSN 1396-0466. doi: 10.5210/fm.v11i9.1394.

Susanne Barth and Menno D.T. de Jong. The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics*, 34(7):1038–1058, November 2017. ISSN 07365853. doi: 10.1016/j.tele.2017.04.013.

Alastair Beresford, Andrew Rice, Nicholas Skehin, and Ripduman Sohan. Mockdroid: Trading privacy for application functionality on smartphones. *HotMobile 2011: The 12th Workshop on Mobile Computing Systems and Applications*, 03 2011. doi: 10.1145/2184489.2184500.

Carlos Bermejo Fernandez, Dimitris Chatzopoulos, Dimitrios Papadopoulos, and Pan Hui. This Website Uses Nudging: MTurk Workers' Behaviour on Cookie Consent Notices. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–22, October 2021. ISSN 2573-0142. doi: 10.1145/3476087.

Brain Code for Equity. Chatbot trends report 2021. `https://chatbotslife.com/chatbot-trends-report-2021-4181eef67dcc`, 2021. [Online; accessed 2022-03-03].

Birgit Brüggemeier and Philip Lalone. Wos - open source wizard of oz for speech systems. In *IUI Workshops*, 2019.

Birgit Brüggemeier and Philip Lalone. ChatBot Language and Experiment Framework, July 2021.

Birgit Brüggemeier and Philip Lalone. Perceptions and reactions to conversational privacy initiated by a conversational user interface. *Computer Speech & Language*, 71:101269, 2022. ISSN 0885-2308.

Michael Buhrmester, Tracy Kwang, and Samuel D. Gosling. Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5, 2011. doi: 10.1177/1745691610393980.

Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–21, April 2021. ISSN 2573-0142. doi: 10.1145/3449287.

Edward Carmines and Richard Zeller. *Reliability and Validity Assessment*. SAGE Publications, Inc, Thousand Oaks, California, January 1979. doi: 10.4135/9781412985642.

Eun Kyoung Choe, Jaeyeon Jung, Bongshin Lee, and Kristie Fisher. Nudging People Away from Privacy-Invasive Mobile Apps through Visual Framing. In Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2013*, pages 74–91, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-40477-1.

David Ciuk, Allison Troy, and Markera Jones. Measuring emotion: Self-reports vs. physiological indicators. *SSRN Electronic Journal*, 01 2015. doi: 10.2139/ssrn.2595359.

Cognigy. Conversational ai glossary. https://www.cognigy.com/resources/conversational-artificial-intelligence-glossary, 2022. [Online; accessed 2022-03-03].

Council of Europe. Convention for the protection of human rights and fundamental freedoms, 1950.

Council of Europe. Convention for the protection of individuals with regard to automatic processing of personal data, 1981.

Council of Europe. Chart of signatures and ratification of treaty 108. https://www.coe.int/en/web/conventions/full-list, 2021. [Online; accessed 2021-12-13].

Pat Croskerry. Cognitive forcing strategies in clinical decisionmaking. *Annals of Emergency Medicine*, 41 (1):110–120, 2003. ISSN 0196-0644. doi: https://doi.org/10.1067/mem.2003.22.

Mary J. Culnan and Pamela K. Armstrong. Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation. *Organization Science*, 10(1):104–115, February 1999. ISSN 1047-7039. doi: 10.1287/orsc.10.1.104.

W. Phillips Davison. The Third-Person Effect in Communication. *Public Opinion Quarterly*, 47(1):1–15, 01 1983. ISSN 0033-362X. doi: 10.1086/268763.

Kenan Degirmenci, Nadine Guhr, and Michael Breitner. Mobile applications and access to personal information: A discussion of users' privacy concerns. *International Conference on Information Systems (ICIS 2013): Reshaping Society Through Information Systems Design*, 3:2570–2590, 12 2013.

van der Sloot and A. De Groot. *The Handbook of Privacy Studies: An Interdisciplinary Introduction.* Amsterdam University Press, 2018. ISBN 9789462988095.

Angelika Dimoka, Yili Hong, and Paul A. Pavlou. On product uncertainty in online markets: Theory and evidence. *MIS Quarterly*, 36(2):395–426, 2012. ISSN 02767783.

John W. Ely, Mark L. Graber, and Pat Croskerry. Checklists to Reduce Diagnostic Errors. *Academic Medicine*, 86(3):307–313, March 2011. ISSN 1040-2446. doi: 10.1097/ACM.0b013e31820824cd.

European Commission. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), 2016.

European Data Protection Board (EDPB). Guidelines on virtual voice assistants, 2021.

Jonathan St B. T. Evans. Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59:255–278, 2008. ISSN 0066-4308. doi: 10.1146/annurev.psych.59.103006.093629.

Gartner Research. Making sense of the chatbot and conversational ai platform market. https://www.gartner.com/en/documents/3993709/making-sense-of-the-chatbot-and-conversational-ai-platfo, 2020. [Online; accessed 2022-03-03].

Martha S. Gerrity, Kinnard P. White, Robert F. DeVellis, and Robert S. Dittus. Physicians' Reactions to Uncertainty: Refining the constructs and scales. *Motivation and Emotion*, 19(3):175–191, September 1995. ISSN 0146-7239, 1573-6644. doi: 10.1007/BF02250510.

Alex Ghanouni, Cristina Renzi, Susanne F Meisel, and Jo Waller. Common methods of measuring 'informed choice' in screening participation: Challenges and future directions. *Preventive Medicine Reports*, 4:601–607, October 2016. ISSN 2211-3355. doi: 10.1016/j.pmedr.2016.10.017.

Charulata Ghosh and Matthew S. Eastin. Understanding Users' Relationship with Voice Assistants and How It Affects Privacy Concerns and Information Disclosure Behavior. In Abbas Moallem, editor, *HCI for Cybersecurity, Privacy and Trust*, Lecture Notes in Computer Science, pages 381–392, Cham, 2020. Springer International Publishing. ISBN 978-3-030-50309-3. doi: 10.1007/978-3-030-50309-3_25.

Thomas Groß. Validity and reliability of the scale internet users' information privacy concern (iuipc) [extended version], 2020.

Hamza Harkous, Kassem Fawaz, Kang G. Shin, and Karl Aberer. PriBots: Conversational privacy with chatbots. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, Denver, CO, June 2016. USENIX Association.

S.C. Hayes and L.J. Hayes. Verbal relations and the evolution of behavior analysis. *American Psychologist*, 47:1383–1395, 1992.

Tianzhi He, Farrokh Jazizadeh, and Laura Arpan. Ai-powered virtual assistants nudging occupants for energy saving: proactive smart speakers for hvac control. *Building Research & Information*, 0(0):1–16, 2021. doi: 10.1080/09613218.2021.2012119.

M. Hof. Questionnaire evaluation with factor analysis and cronbach ' s alpha an example -, 2012.

Arghavan Hosseinzadeh., Andreas Eitel., and Christian Jung. A systematic approach toward extracting technically enforceable policies from data usage control requirements. In *Proceedings of the 6th International Conference on Information Systems Security and Privacy - ICISSP,*, pages 397–405. INSTICC, SciTePress, 2020. ISBN 978-989-758-399-5. doi: 10.5220/0008936003970405.

Interaction Design Foundation. Forcing Functions. `https://www.interaction-design.org/literature/book/the-glossary-of-human-computer-interaction/forcing-functions`, 2021. [Online; accessed 2021-12-21].

Carolin Ischen, Theo Araujo, Hilde Voorveld, Guda van Noort, and Edith Smit. Privacy Concerns in Chatbot Interactions. In Asbjørn Følstad, Theo Araujo, Symeon Papadopoulos, Effie Lai-Chong Law, Ole-Christoffer Granmo, Ewa Luger, and Petter Bae Brandtzaeg, editors, *Chatbot Research and Design*, pages 34–48, Cham, 2020. Springer International Publishing. ISBN 978-3-030-39540-7.

ITU-T P.808. Subjective evaluation of speech quality with a crowdsourcing approach, june 2021.

Yousra Javed, Shashank Sethi, and Akshay Jadoun. Alexa's Voice Recording Behavior: A Survey of User Understanding and Awareness. In *Proceedings of the 14th International Conference on Availability, Reliability and Security*, pages 1–10, Canterbury CA United Kingdom, August 2019. ACM. ISBN 978-1-4503-7164-3. doi: 10.1145/3339252.3340330.

Christopher Johns and Christopher Johns, editors. *Guided reflection: a narrative approach to advancing professional practice.* Blackwell Pub, Chichester, West Sussex ; Ames, Iowa, 2nd ed edition, 2010. ISBN 978-1-4051-8568-4.

Juniper Research. Hey siri, how will you make money?, 2020.

Daniel Kahneman. *Thinking, fast and slow.* Farrar, Straus and Giroux, New York, 2011. ISBN 9780374275631 0374275637.

Hyunjin Kang and Jeeyun Oh. Communication privacy management for smart speaker use: Integrating the role of privacy self-efficacy and the multidimensional view. *New Media & Society*, June 2021. ISSN 1461-4448, 1461-7315. doi: 10.1177/14614448211026611.

Flavius Kehr, Tobias Kowatsch, Daniel Wentzel, and Elgar Fleisch. Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus: Privacy calculus: dispositions and affect. *Information Systems Journal*, 25(6):607–635, November 2015. ISSN 13501917. doi: 10.1111/isj.12062.

Dongyeon Kim, Kyuhong Park, Yongjin Park, and Jae-Hyeon Ahn. Willingness to provide personal information: Perspective of privacy calculus in IoT services. *Computers in Human Behavior*, 92:273–281, March 2019. ISSN 0747-5632. doi: 10.1016/j.chb.2018.11.022.

Rafal Kocielnik, Daniel Avrahami, Jennifer Marlow, Di Lu, and Gary Hsieh. Designing for workplace reflection: A chat and voice-based conversational agent. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, page 881–894, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450351980. doi: 10.1145/3196709.3196784.

Spyros Kokolakis. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64:122–134, January 2017. ISSN 01674048. doi: 10.1016/j.cose.2015.07.002.

Marina Konrad, Sabine Koch-Sonneborn, and Christopher Lentzsch. The Right to Privacy in Socio-Technical Smart Home Settings: Privacy Risks in Multi-Stakeholder Environments. In Constantine Stephanidis and Margherita Antona, editors, *HCI International 2020 - Posters*, volume 1226, pages 549–557. Springer International Publishing, Cham, 2020. ISBN 978-3-030-50731-2 978-3-030-50732-9. doi: 10.1007/978-3-030-50732-9_71. Series Title: Communications in Computer and Information Science.

Kathryn Ann Lambe, Gary O'Reilly, Brendan D Kelly, and Sarah Curristan. Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. *BMJ Quality & Safety*, 25(10): 808–820, October 2016. ISSN 2044-5415, 2044-5423. doi: 10.1136/bmjqs-2015-004417.

Lehigh University. Conducting research using crowdsourcing platforms: Best practices. `https://research.cc.lehigh.edu/crowdsourcing`, 2022. [Online; accessed 2022-01-27].

Russell V. Lenth. *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2022. R package version 1.7.2.

Christopher Lentzsch, Sheel Jayesh Shah, Benjamin Andow, Martin Degeling, Anupam Das, and William Enck. Hey alexa, is this skill safe?: Taking a closer look at the alexa skill ecosystem. *Network and Distributed Systems Security (NDSS) Symposium2021*, 2021. doi: 10.14722/ndss.2021.23111.

Anna Leschanowsky, Birgit Brüggemeier, and Nils Peters. Design implications for human-machine interactions from a qualitative pilot study on privacy. In *Proc. 2021 ISCA Symposium on Security and Privacy in Speech Communication*, pages 76–79, 11 2021. doi: 10.21437/SPSC.2021-16.

Haiko Luepsen. Anova with binary variables: the f-test and some alternatives. *Communications in Statistics - Simulation and Computation*, 0(0):1–25, 2021. doi: 10.1080/03610918.2020.1869983.

Lumen Learning. Rational Decision Making vs. Other Types of Decision Making | Principles of Management. `https://courses.lumenlearning.com/wm-principlesofmanagement/chapter/rational-decision-making-vs-other-types-of-decision-making/`, 2022. [Online; accessed 2022-03-06].

Christoph Lutz and Gemma Newlands. Privacy and smart speakers: A multi-dimensional approach. *The Information Society*, 37(3):147–162, May 2021. ISSN 0197-2243. doi: 10.1080/01972243.2021.1897914.

Naresh K. Malhotra, Sung S. Kim, and James Agarwal. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, 15(4):336–355, December 2004. ISSN 1047-7047, 1526-5536. doi: 10.1287/isre.1040.0032.

Silvia Mamede, Henk G. Schmidt, and Júlio César Penaforte. Effects of reflective practice on the accuracy of medical diagnoses. *Medical Education*, 42(5):468–475, May 2008. ISSN 1365-2923. doi: 10.1111/j.1365-2923.2008.03030.x.

T. M. Marteau, E. Dormandy, and S. Michie. A measure of informed choice. *Health Expectations: An International Journal of Public Participation in Health Care and Health Policy*, 4(2):99–108, June 2001. ISSN 1369-6513. doi: 10.1046/j.1369-6513.2001.00140.x.

Stewart Martin. Measuring cognitive load and cognition: metrics for technology-enhanced learning. *Educational Research and Evaluation*, 20(7-8):592–621, 2014. doi: 10.1080/13803611.2014.997140.

J. Marton-Williams. *Questionnaire design, Consumer Market Research Handbook*. McGraw-Hill Book Company, London, 1986.

Philipp K. Masur. *Situational Privacy and Self-Disclosure*. Springer International Publishing, Cham, 2019. ISBN 978-3-319-78883-8 978-3-319-78884-5. doi: 10.1007/978-3-319-78884-5.

Daniel McNeish. Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3):412–433, September 2018. ISSN 1939-1463, 1082-989X. doi: 10.1037/met0000144.

Jon Merz and Baruch Fischhoff. Informed consent does not mean rational consent. cognitive limitations on decision-making. *The Journal of legal medicine*, 11:321–50, 10 1990. doi: 10.1080/01947649009510831.

Arthur R. Miller. *The assault on privacy : computers, data banks, and dossiers / Arthur R. Miller*. University of Michigan Press Ann Arbor, 1971. ISBN 0472655000.

Christoph Molnar. *Interpretable Machine Learning.* Christoph Molnar c/o Mucbook Clubhouse, 2019.

Andreas Nautsch, Catherine Jasserand, Els Kindt, Massimiliano Todisco, Isabel Trancoso, and Nicholas Evans. The GDPR & Speech Data: Reflections of Legal and Technology Communities, First Steps Towards a Common Understanding. In *Interspeech 2019*, pages 3695–3699. ISCA, September 2019. doi: 10.21437/Interspeech.2019-2647.

Oliver Neumann. Does misfit loom larger than fit? experimental evidence on motivational person-job fit, public service motivation, and prospect theory. *International Journal of Manpower*, 37:822–839, 07 2016. doi: 10.1108/IJM-12-2014-0268.

Helen Nissenbaum. Privacy as Contextual Integrity. *Washington Law Review*, 79:41, 2004.

Helen Fay Nissenbaum. *Privacy in context: technology, policy, and the integrity of social life.* Stanford Law Books, Stanford, Calif, 2010. ISBN 978-0-8047-5236-7 978-0-8047-5237-4.

Patricia A. Norberg, Daniel R. Horne, and David A. Horne. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs*, 41(1):100–126, June 2007. ISSN 00220078. doi: 10.1111/j.1745-6606.2006.00070.x.

Donald A. Norman. *The design of everyday things.* Basic Books, [New York], 2002. ISBN 0465067107 9780465067107.

Paul C. Nutt. Expanding the Search for Alternatives during Strategic Decision-Making. *The Academy of Management Executive (1993-2005)*, 18(4):13–28, 2004. ISSN 1079-5545. Publisher: Academy of Management.

Judith S. Olson and Wendy A. Kellogg. *Ways of Knowing in HCI.* Springer Publishing Company, Incorporated, 2014. ISBN 1493903772.

Ottomatias Peura. What are voice user interfaces. `https://www.speechly.com/blog/what-is-voice-user-interface`, 2020. [Online; accessed 2022-03-03].

Paul A. Pavlou, Huigang Liang, and Yajiong Xue. Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. *MIS Quarterly*, 31(1):105–136, 2007. ISSN 02767783.

Iryna Pentina, Lixuan Zhang, Hatem Bata, and Ying Chen. Exploring privacy paradox in information-sensitive mobile app adoption: A cross-cultural comparison. *Computers in Human Behavior*, 65:409–419, December 2016. ISSN 0747-5632. doi: 10.1016/j.chb.2016.09.005.

William Revelle. Using R and the psych package to find ω, 2017.

William Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research.* Northwestern University, Evanston, Illinois, 2021. R package version 2.1.9.

Judy Robertson and Maurits Kaptein. *Modern Statistical Methods for HCI.* Springer Publishing Company, Incorporated, 2018. ISBN 3319799843.

Rajwant Sandhu and Benjamin J. Dyson. Re-evaluating visual and auditory dominance through modality switching costs and congruency analyses. *Acta Psychologica*, 140(2):111–118, June 2012. ISSN 00016918. doi: 10.1016/j.actpsy.2012.04.003.

Robert Sawicki. Informed Consent or Informed Choice? | OSF HealthCare, October 2012.

Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A Design Space for Effective Privacy Notices*. In Evan Selinger, Jules Polonetsky, and Omer Tene, editors, *The Cambridge Handbook of Consumer Privacy*, pages 365–393. Cambridge University Press, 1 edition, March 2018. ISBN 978-1-316-83196-0 978-1-107-18110-6. doi: 10.1017/9781316831960.021.

Eva-Maria Schomakers, Chantal Lidynia, Dirk Müllmann, and Martina Ziefle. Internet users' perceptions of information sensitivity – insights from Germany. *International Journal of Information Management*, 46:142–150, June 2019. ISSN 02684012. doi: 10.1016/j.ijinfomgt.2018.11.018.

Marcia Y Shade, Kyle Rector, Rasila Soumana, and Kevin Kupzyk. Voice assistant reminders for pain self-management tasks in aging adults. *Journal of gerontological nursing*, pages 1–7, 2020.

Herbert A. Simon. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1): 99–118, February 1955. ISSN 0033-5533. doi: 10.2307/1884852.

Herbert A. Simon. *Models of Bounded Rationality: Empirically Grounded Economic Reason*, volume 3. MIT Press, Cambridge, MA, USA, July 1997. ISBN 978-0-262-19372-6.

Craig Smith and Phoebe Ellsworth. Patterns of cognitive appraisal in emotion. *Journal of personality and social psychology*, 48:813–38, 05 1985. doi: 10.1037//0022-3514.48.4.813.

softengi.com. Voice Controlled Checklist App. `https://softengi.com/projects/voice-controlled-checklist-app/`, 2022. [Online; accessed 2022-02-24].

Daniel J Solove. The Myth of the Privacy Paradox. *The George Washington Law Review*, 89:52, 2020.

Sonrat Priyanka. Intelligent virtual assistants – the next-gen of chatbots. `https://medium.com/kevit-technologies/intelligent-virtual-assistants-the-next-gen-of-chatbots-b92cd7b55e22`, 2020. [Online; accessed 2022-03-07].

Speaker Identifitcation Integrated Project (SIIP). D2.8 final report on pbd (privacy by design) and operations guidelines for integrated audio and voice recognition, August 2018.

Keith E. Stanovich and Richard F. West. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(5):645–665, October 2000. ISSN 1469-1825, 0140-525X. doi: 10.1017/S0140525X00003435. Publisher: Cambridge University Press.

Herman T. Tavani. Philosophical theories of privacy: Implications for an adequate online privacy policy. *Metaphilosophy*, 38(1):1–22, January 2007. ISSN 0026-1068, 1467-9973. doi: 10.1111/j.1467-9973.2006.00474.x.

Katie Teague. Amazon, Apple and Google are always listening: How to opt out and delete your voice recordings - CNET. `https://www.cnet.com/home/smart-home/`

alexa-delete-what-i-just-said-heres-how-to-prevent-amazon-from-listening-in/, 2021. [Online; accessed 2021-12-13].

Richard H. Thaler and Cass R. Sunstein. *Nudge*. Yale University Press, New Haven, CT and London, 2008. ISBN 978-0-300-12223-7.

Sabine Trepte, Leonard Reinecke, Nicole B. Ellison, Oliver Quiring, Mike Z. Yao, and Marc Ziegele. A Cross-Cultural Perspective on the Privacy Calculus. *Social Media + Society*, 3(1), January 2017. ISSN 2056-3051, 2056-3051. doi: 10.1177/2056305116688035.

Matina Tsavli, Pavlos Efraimidis, Vasilios Katos, and Lilian Mitrou. Reengineering the user: Privacy concerns about personal data on smartphones. *Information & Computer Security*, 23:394–405, 01 2015.

United Nations. *Universal Declaration of Human Rights*. UN General Assembly, December 1948.

United Nations (General Assembly). International covenant on civil and political rights. *Treaty Series*, 999:171, December 1966.

United Nations (General Assembly). Chart of signatures and ratification of the international covenant on civil and political rights. https://indicators.ohchr.org/, 2022. [Online; accessed 2022-01-17].

Hans van der Heijden. Priming System 1 Influences User Acceptance. *SIGHCI 2013 Proceedings*, page 25, 2013.

I. van Ooijen and Helena U. Vrabec. Does the GDPR Enhance Consumers' Control over Personal Data? An Analysis from a Behavioural Perspective. *Journal of Consumer Policy*, 42(1):91–107, March 2019. ISSN 1573-0700. doi: 10.1007/s10603-018-9399-7.

J. Vitak. Feature creep or just plain creepy? How advances in "smart" technologies affect attitudes toward data privacy. *Annals of the International Communication Association*, January 2020.

Vixen Labs Limited in partnership with Open Voice Network. Voice Consumer Index June 2021, 2021.

Yang Wang, Pedro Giovanni Leon, Kevin Scott, Xiaoxuan Chen, Alessandro Acquisti, and Lorrie Faith Cranor. Privacy nudges for social media: An exploratory facebook study. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion, page 763–770, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450320382. doi: 10.1145/2487788.2488038.

Samuel D. Warren and Louis D. Brandeis. The right to privacy. *Harvard Law Review*, 4(5):193–220, 1890.

Marley W. Watkins. The reliability of multidimensional neuropsychological measures: from alpha to omega. *The Clinical Neuropsychologist*, 31(6-7):1113–1126, October 2017. ISSN 1385-4046. doi: 10.1080/13854046.2017.1317364. Publisher: Routledge _eprint: https://doi.org/10.1080/13854046.2017.1317364.

David B Watson and Lee Anna Clark. The panas-x manual for the positive and negative affect schedule, 1994.

James N. Weinstein. Partnership: Doctor and Patient: Advocacy for Informed Choice: Vs.: Informed Consent. *Spine*, 30(3):269–272, February 2005. ISSN 0362-2436. doi: 10.1097/01.brs.0000155479.88200.32.

Alan F. Westin. *Privacy and Freedom.* Atheneum, New York, 1967.

Wikipedia. Checklist. `https://en.wikipedia.org/w/index.php?title=Checklist&oldid=1057709965`, November 2021. [Online; accessed 2021-12-21].

Meredydd Williams, Jason R. C. Nurse, and Sadie Creese. Privacy is the Boring Bit: User Perceptions and Behaviour in the Internet-of-Things. In *2017 15th Annual Conference on Privacy, Security and Trust (PST)*, pages 181–18109, Calgary, AB, August 2017. IEEE. ISBN 978-1-5386-2487-6. doi: 10.1109/PST.2017.00029.

Boonghee Yoo, Naveen Donthu, and Tomasz Lenartowicz. Measuring hofstede's five dimensions of cultural values at the individual level: Development and validation of cvscale. *Journal of International Consumer Marketing*, 23(3-4):193–210, 2011. doi: 10.1080/08961530.2011.578059.

Richard E. Zinbarg, William Revelle, Iftah Yovel, and Wen Li. Cronbach's, α revelle's β and mcdonald's ω h: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1):123–133, March 2005. ISSN 0033-3123. doi: 10.1007/s11336-003-0974-7.

Glossary

Affect Heuristic A mental shortcut that allows to make decisions quickly and efficiently but influenced by current emotions. 22

Big Five Personality A taxonomy for personality traits. The Big Five personality traits are extraversion, agreeableness, openness, conscientiousness and neuroticism. 18

Bounded Rationality The idea that rationality in human decision-making is limited and determined by a satisfactory decision rather than by an optimal decision. 19

Cognitive Ease The ease with which our brain processes information. 21, 36, 94

Cohen's d An effect size used to indicate the standardised difference between two means. 47

Construct Validity The extent to which a test or measure accurately assess what it is supposed to. 52–54, 72

Contextual Integrity A theory of privacy developed by Helen Nissenbaum. The framework assumes that privacy is associated with and regulated by a context-dependent flow of information based on norms. 11, 35

Convention 108 The Convention for the protection of Individuals with regard to Automatic Processing of Personal Data, the first legally binding international instrument in the data protection field. 12

Conversational Privacy In the context of CUI it refers to conversational agents that express privacy-related information in dialogue form. 6, 9, 14, 32, 35, 40, 59, 61, 97, 98, 101

Council of Europe An international organisation to uphold human rights, democracy and the rule of law in Europe. 11, 12

Cronbach's Alpha The most common text score reliability coefficient. 50–52, 58, 72, 73, 158

Data Protection by Design and by Default Article 25 of the GDPR which states that data protection needs to be integrated into processing activities and business practices from the design stage through the lifecycle. 12

Dual-Process Theory A cognitive psychology theory that divides the processing of information in two pathways, a fast, automatic and unconscious process (System 1) and a slow, controlled and conscious process (System 2). 6, 19, 20, 24, 31, 94, 159

Guidelines on Virtual Voice Assistants Guidelines by the EDPB on Virtual Voice Assistants that identify most relevant compliance challenges concerning the GDPR and e-Privacy Directive and provide recommendations. 12, 13

Information Asymmetry A condition under which one party possesses more information than the other party they are dealing with. 39

Informational Privacy A concept that refers to privacy in relation with information and communication technology and an individual's related data. 10, 11

Likert Scale A psychometric scale commonly used to scaling responses in survey research. 36, 38, 40–42, 47

Need for Cognition The extent to which individuals are inclined towards effortful cognitive activities. 27

Nudge as defined by Thaler and Sunstein, a nudge is any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. 22–24, 35, 61, 64, 66, 67, 88

Omega A reliability coefficient that is based on factor analysis and overcomes the deficiencies of Cronbach's Alpha. 50–52, 72, 73

PANAS-X An expanded form of the PANAS scale that consists of 60 items and incorporated the original PANAS as well as measures of 11 lower order emotional states. 37, 52, 53, 58, 119, 141

Priming or Priming Effect occurs when an individual's exposure to a certain stimulus influences their response to a subsequent stimulus, without being aware of the connection. 21, 22, 28, 41

Privacy Calculus A theory that states that individuals rationally weigh potential benefits and potential risks of decisions regarding disclosure of personal information. 9, 17, 18

Privacy Paradox The discrepancy between an individuals' attitudes on privacy and their actual behaviour. 6, 9, 14–19, 101

Prospect Theory A theory of behavioral economics that describes how individuals assess their loss and gain perspectives in an asymmetric manner. 20

Quantum Theory An alternative probabilistic framework for modelling decision-making compared with classical probability theory. It applies mathematical formalism of quantum physics to model cognitive phenomena. 20

Rational Choice Theory of Human Behaviour A theory that states that individuals rely on rational calculations to make rational choices that result in outcomes aligned with their own best interests.. 17

Right to Erasure ("Right to be Forgotten") Article 17 of the GDPR that gives data subjects the right to obtain from the controller the erasure of personal data converning him or her without undue delay. 12

System 1 One pathway of the Dual-Process Theory where processing of information happens fast and automatically, with little or no effort and no sense of voluntary control. 20–22, 28, 32, 48, 57, 96, 157

System 2 One pathway of the Dual-Process Theory where processing of information happens slow and controlled, also known as the rational or rule-based system. 20–22, 28, 31, 32, 36, 60, 61, 72, 82, 96, 157

Third-Person Effect Theory A theory that states that an individual believes that mass communication has different and greater effects on others than on themselves. 19

Wizard of Oz A research experiment in which subjects interact with a computer system that subjects believe to be autonomous but which is actually being operated or partially operated by an unseen human being. 33