Uncertain yet rational - Uncertainty as an Evaluation Measure of Rational Privacy Decision-Making in Conversational AI

Anna Leschanowsky[1 [0000−0003−2994−2336]], Birgit Popp[1 [0000−0003−2193−9861]], and Nils Peters[2 [0000−0001−6758−4491]]

[1] Fraunhofer IIS, Germany
[2] International Audio Laboratories Erlangen, Germany*
{anna.leschanowsky, birgit.popp}@iis.fraunhofer.de
nils.peters@fau.de

Abstract. In today's connected world, privacy decision-making is crucial for people to maintain control over their personal information and effectively manage their privacy. However, people's decisions on privacy are likely to be subject to bias and can lead to frustration and regret. Privacy strategies in Conversational AI can aim at debiasing peoples' choices by drawing from dual-process theory and triggering a more rational thinking process. Previous research on evaluation measures for such strategies has focused on minimizing regret or aligning user behaviour with their attitudes. In this paper, we propose a subjective measure of uncertainty to evaluate the effectiveness of debiasing strategies in a Conversational AI privacy scenario. We investigate two different scales of uncertainty - an adapted privacy uncertainty scale consisting of four subscales and the PANAS-X scale on the affective state of fear. We find that only one of the adapted subscales and the scale on fear showed sufficient reliability and validity results. Moreover, we did not find differences in uncertainty between our tested strategies. Finally, we propose alternative measures to investigate uncertainty and evaluate privacy strategies that promote rational thinking in the future.

Keywords: Conversational AI · System 1 and 2 · Debiasing · Uncertainty · Privacy.

1 Introduction

With numerous connected devices deployed in peoples' homes, cars and public spaces, privacy decision-making becomes increasingly frequent and inevitable. Previous research on privacy decision-making has shown that people often rely on heuristics and biases when deciding whether to disclose personal information or adjust privacy settings [1]. Consequently, judgements can be suboptimal

* The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg and Fraunhofer IIS.

and lead to frustration and regret [1, 22]. Debiasing strategies can be a means of supporting people in making choices that are better aligned with their attitudes [23]. They are theoretically grounded in the dual-process theory that distinguishes between System 1, fast and intuitive thinking, and System 2, slow and analytical thinking [17]. As biased judgements usually originate from System 1, debiasing strategies that promote transitioning towards System 2 are one way to promote more rational judgements. Such strategies can aim at producing competing intuitions, e.g. by presenting alternatives, which can induce a feeling of uncertainty during decision-making. Therefore, based on theoretical and experimental research on the dual-process theory and debiasing strategies, we argue that a subjective measure of uncertainty could provide insights into System 2 activity.

We evaluate the level of perceived uncertainty in the context of privacy decision-making in Conversational AI (CAI). While various debiasing strategies have been investigated for traditional user interfaces to support users in making more optimal choices [1, 4, 34], similar controls for CAI systems have not yet been analyzed. Because traditional privacy controls in CAI require unfavorable modality switching and were found insufficient and cumbersome to use [21], an increasing stream of research argues for conversational privacy [6, 16, 28]. Our conversational privacy strategies are based on previous research on debiasing strategies and are designed to induce a controlled level of uncertainty. They aim at triggering System 2 activity and at supporting people in overcoming their biases and making better judgements. Previous research on privacy nudges proposed evaluation measures that focus on minimizing regret or aligning user behaviour with their attitudes [1]. However, such measures do not necessarily allow drawing conclusions about peoples' underlying thinking process. Therefore, we show how a subjective measure of uncertainty can provide insights on System 2 activity in Section 2. We describe our experimental setup and the conversational privacy strategies used to induce uncertainty in a Conversational AI privacy scenario in Section 3. Lastly, we show that only two subjective scales have proven reliable and propose alternative measures for assessing uncertainty and evaluating conversational privacy strategies that promote analytical thinking in Sections 4 and 5.

2 Uncertainty as an Evaluation Measure

Situations of privacy decision-making are influenced by uncertainty and risks [2]. While uncertainty and risk are closely related, economic research describes them as two distinct constructs [19]. The distinction is based on whether probabilities of possible outcomes are known. In situations under uncertainty, probabilities are unknown, while they are knowable in situations under risks [19]. Individuals are often subject to uncertainty when making judgments about their privacy. First, information asymmetry, the fact that the user has less information available compared to the provider of a service, is a critical driver for uncertainty [2, 3]. Moreover, individuals face difficulties in predicting outcomes and consequences

of their actions not only because of insufficient information but because technological developments may be unknowable and difficult to predict [2].

Considering uncertainty as a factor in privacy decision-making can be informed by research in the behavioural economics field. Here, characteristics of judgments under uncertainty and their underlying theory have been studied for almost five decades [32]. In uncertain scenarios, people are likely to rely on mental shortcuts, i.e. heuristics, to reduce complexity and the need to correctly assess probabilities of possible outcomes. While these techniques can be useful in our daily lives, they can lead to privacy choices that are biased and unaligned with peoples' attitudes. Heuristics and biases that have been identified as relevant hurdles for privacy choices include the availability heuristic, representativeness heuristic and optimism bias and overconfidence [1]. Explanations about judgments under uncertainty are primarily based on the dual-process theory or System 1 and 2 reasoning [17]. While decision-making under System 1 is fast, automatic and effortless, choices under System 2 are slower, more controlled and effortful. It is assumed that impressions that are created by System 1 are mostly adopted by System 2 without further intervention. Only in situations that are surprising or violate the mental model, System 2 and with this, analytical reasoning, is activated [17]. This activation can be accompanied by uncertainty.

Debiasing strategies have been extensively researched in the medical field to reduce diagnostic errors resulting from cognitive biases [20]. Similarly, they can be applied in privacy decision-making to support people in overcoming their biases [1, 23]. Thereby, debiasing strategies can induce a controlled level of uncertainty often by producing an internal conflict or competing intuitions. Consequently, System 2 activation is likely to be triggered. For example, in a study on the effectiveness of debiasing strategies, it was shown that the Socratic procedure, i.e. posing thought-provoking questions, and the Devil's advocate approach, i.e. encouraging people to consider an opposing point of view, led to an increase of subjective uncertainty [8]. In particular, the authors focused on overconfidence bias and availability heuristics, both of them also prevalent in privacy decision-making. It is important to note that we are not interested in people's perceptions of general uncertainty in privacy decision-making settings. Instead, we focus on the uncertainty that was induced by debiasing strategies and can be seen as a possible result of an internal conflict. This distinction will be critical when defining our experimental setup.

While the dual-process model has been subject to criticism and has been adapted and modified over the years, uncertainty has remained a crucial factor. For example, criticism regarding the dichotomy of the two systems has led to further development of the dual process model that relies on a component to monitor uncertainty [27]. Thereby, competing intuitions are monitored for their similarity. Uncertainty increases the more similar competing intuitions are until a certain threshold is reached and System 2 is activated. In the medical field, Croskerry [10] developed a universal model of diagnostic reasoning based on pattern recognition and the dual-process theory. Again, uncertainty plays a crucial role in the activation of System 2. If a patient's symptoms do not

match previous situations or cannot be directly assigned to a certain illness, uncertainty increases and calls for analytical thinking. In addition, behavioural, electrophysiological and neuroimaging studies have shown evidence for a relationship between uncertainty and cognitive control [26]. In particular, there is evidence that uncertain environments lead to increased monitoring activity of one's behavior [26].

Given this theoretical basis and evidence from experimental research, we hypothesize that uncertainty can be used as an evaluation measure for analytical decision-making in the privacy context. Therefore, we construct a privacy scenario in which uncertainty can be induced and varied using debiasing strategies. Moreover, while a majority of previous studies have relied on behavioural and physiological measures to assess uncertainty, we aim to investigate subjective measures of uncertainty. A subjective measure could preferably complement already established objective measures and is less time-consuming and burdensome for participants.

3 Experimental Design

3.1 CAI System, Scenarios and Experimental Conditions

We will investigate the perceived level of uncertainty in two different chatbot scenarios and five varying conditions. We use a text-based CAI system and will refer to the implementation as chatbots which use natural language to interact with a human via text [29]. We use Chatbot Language (CBL) [29] to implement the chatbot on Amazon Mechanical Turk (Mturk). As information sensitivity can impact perceived uncertainty, we investigate two chatbot scenarios – a banking chatbot asking for permission to access users' credit card information and a location chatbot asking for permission to access users' location. These two scenarios were chosen based on previous studies where credit card numbers and location data were perceived significantly different with respect to information sensitivity [30].

Our privacy strategies are designed to provide transparency in interactions with CAI based on the principle of conversational privacy [6, 16]. Moreover, we aim to make the "right to deletion" required by the GDPR easily accessible by proactively presenting users with an offer to delete their data [13]. Table 1 gives an overview of control and privacy conditions. The first control condition resembles a common interaction with a CAI system nowadays, is unrelated to data privacy and thus, serves as a baseline for people's perceived level of uncertainty. As mentioned above, we are interested in the level of uncertainty induced by the debiasing strategies rather than the general uncertainty experienced in situations of privacy decision-making.

Our second control condition gives people the opportunity to actively control their privacy while at the same time nudging them into disclosing behaviour. Similar strategies, called "dark patterns" are used in interface design, e.g., when designing cookie banners [5]. Here, interfaces are designed such that individuals

Table 1. Overview of conditions and their questions asked by the CAI system, including two control conditions and three privacy strategies.

Condition	Question
Control 1	Is there anything else I can help you with?
Control 2	I will save your data for future interactions now, okay?
Slow Down	I will save your data for future interactions now, okay? I'll give you 20 seconds to think about it.
Alternative	Do you want me to delete your data from this interaction or have it saved for future interactions?
Deletion	Do you want me to delete your data from this interaction now?

make decisions that favour data collectors rather than themselves [5]. Despite the fact that the second control condition is related to data privacy and might lead to an increased level of general uncertainty, we expect most people to choose intuitively and therefore, rely on System 1 activity. Therefore, their perceived level of uncertainty should remain relatively low compared to the one experienced by participants who are exposed to the debiasing strategies.

We implement three different privacy strategies based on the idea of debiasing, in particular on cognitive forcing [7, 23]. The privacy strategies are applied at the time of decision-making to disrupt heuristic reasoning. They are designed to induce a controlled level of uncertainty, make users engage in System 2 thinking and thereby, support the process of rational cost-benefit analysis. Drawing from previous studies on privacy nudges and cognitive forcing, we implement 1) a slow-down condition to give users time to reflect and possibly reconsider their decision [7, 34], 2) an alternative condition that requires an active choice [20] and 3) an option to delete their data from the interaction [6]. The slow-down and alternative options aim at producing an internal conflict and may consequently lead to an increased level of perceived uncertainty. The option to delete data allows people to reconsider their previous disclosure to the chatbot. Similarly to the other privacy strategies, reconsideration could produce competing intuitions. On the other hand, this offer might come surprising to participants as it is not frequently used in real-life scenarios. Based on the theory, surprise is likely to trigger the activation of System 2 [17]. Moreover, surprise has been shown to be accompanied by the feeling of uncertainty and thus, could be assessed via the subjective uncertainty scale [31].

After granting or denying access to their data, participants were exposed to one of the three privacy strategies or to one of the two control conditions. Our experiment follows a between-subject design and conditions were randomly assigned to participants. We used CBL to inform participants about data protection regulations, to provide a task description and to display a survey on the chatbot interaction after the experiment.

3.2 Survey Design

We make use of subjective scales from previous research to investigate whether uncertainty can serve as an evaluation measure for our privacy strategies. We measure uncertainty retrospectively after the interaction was carried out. To the best of our knowledge, subjective measures of uncertainty have not yet been used in the context of CAI. Therefore, we adopt a privacy uncertainty scale that was used in the context of mobile applications [3]. Their study showed that privacy uncertainty was positively influenced by uncertainty regarding the collection, use and protection of users' data. Relying on this distinction allows us to validate the established scale in the context of CAI and to get a more detailed picture of possible subdimensions that induce uncertainty and may be involved in the System 2 activation process. While the original scale was used to measure privacy uncertainty before and after purchasing a mobile app, we only rely on the post-purchase scale 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 these items seem not relevant [3]. Lastly, we rephrased the items to match the context of the chatbot interaction (see Table 6 in the Appendix for the rephrased items).

Our second scale is based on the relationship between affective states and uncertainty. As fear was found to be significantly influenced by uncertainty [31], we assess participants' affective state on fear using the PANAS-X scale [35].

3.3 Ethical Considerations

In the following, we discuss the ethical considerations of our experimental design and describe the measures taken to ensure that participants were treated ethically. First, participants were not told prior to participating that we evaluated data-saving practices as this might have affected their behaviour and perceptions. However, our task description clearly stated that participants will be asked personal questions by the chatbot system and are free to what extent they respond truthfully. Moreover, the two scenarios were designed to ask only for information that was required to fulfil the task and thus follow current best practices of privacy design [13, 22].

While our experiment made participants believe that we could access their data, our system was not able to access any personal information other than the text users shared during the interaction. This deceptive design choice was based on lessons learned from a previous study [6]. There, participants were provided with an artificial credit card number and asked to check the corresponding balance. Checking the balance for an artificial bank account did limit the interpretability of the results as it does not represent a real-life scenario in which users enter personal data. Thus, users' perceptions and behaviour may differ as they may be more concerned when disclosing real personal information. We fully disclosed our practices by debriefing participants after the study and highlighting that no personal data was accessed if they had not entered personal information during the interaction. Finally, we paid participants 2$ for their participation

Table 2. Summary of demographic and experimental data for the banking and location scenario

Demographic and experimental data	Banking	Location
# conditions	5	5
# participants	315	330
# excluded participants	33	53
# accepted participants in the different conditions (Control 1/ Control 2/ Slow Down/ Alternative/ Reconsider)	58/56/51/56/61	63/55/53/51/55
# accepted participants' disclosure behaviour (Granting Access/ Denying Access)	228/54 (81%/19%)	245/32 (88%/12%)
Mean (SD) age of workers in years	34 (10)	35 (10)
# Gender (female/male/diverse/not provided)	151/131/0/0	114/163/0/0

which calculates to an average hourly wage of 17$ for the banking scenario and 20$ for the location scenario.

4 Results

We show experimental and demographic data in Table 2. We excluded participants who failed at least one out of three screening questions from our analysis. Based on the results of a power analysis, we ensured that each of the groups yielded more than 50 accepted participants. The disclosure behaviour of participants was similar between scenarios with more than 80% granting access to their personal information. This is an essential prerequisite for our conditions as they rely on users sharing information in the first place.

4.1 Evaluation of the Adapted Privacy Uncertainty Scale

Evaluation using Structural Equation Modelling Evaluation of the privacy uncertainty scale is based on covariance-based structural equation modelling (CB-SEM). This is different to the original study where privacy uncertainty was part of a larger structural model evaluated via partial least squares SEM [3]. We do not aim for comparability but for an evaluation of the uncertainty scale in the context of CAI and rational privacy decision-making. We rely on CB-SEM as we include only reflective constructs and like to assess global goodness-of-fit measures [15]. Our structural model is based on the assumption that collection, use and protection uncertainty positively influence overall privacy uncertainty [3]. Therefore, we assume that these three constructs, i.e. collection, use and protection uncertainty, load on overall privacy uncertainty, while all four constructs are treated as a cause of their corresponding indicators. We evaluate the CB-SEM using R and the package *lavaan*. As privacy uncertainty was measured on an ordinal 5-point Likert-Scale, we use the robust estimator Weighted

Least Squares with Adjustments for Means and Variances (WLSMV) [18]. The model is built on participants' data who passed the screening questions (N=559).

In SEM one distinguishes between the measurement model and the path analysis [11]. First, the measurement model is tested on reliability and validity by performing a confirmatory factor analysis (CFA) on the latent variables, i.e. collection, use, protection and overall uncertainty [11]. Second, the structural relationship between latent variables is evaluated by performing a path analysis [11]. To evaluate the CB-SEM measurement model, we constrain the loading of the first indicator on each latent factor to unity. We assess item reliability by investigating the standardized loadings of the individual items on their constructs (see Table 6 in the Appendix). Standardized factor loadings vary between 0.5 and 0.8 for the individual uncertainty items. All factor loadings were above the generally recommended lower limit of 0.4 with most of them exceeding factor loadings of 0.7. Thus, indicating sufficient item reliability.

Further, we analyze internal consistency reliability, convergent validity and discriminant validity of the CB-SEM measurement model. The results are shown in Table 3. A scale is generally considered reliable for Cronbach's $\alpha \geq 0.7$ and composite reliability ≥ 0.7 [15]. Moreover, we check whether convergent validity could be established with an average variance extracted (AVE) of > 0.50. Lastly, we investigate discriminant validity using the heterotrait-monotrait correlation ratio (HTMT) which has been recommended over the Fornell-Larcker criterion [15]. As we are working with conceptually similar concepts, we apply a more relaxed cut-off value of 0.9 for discriminant validity to be present [15]. While the use and overall uncertainty scale showed sufficient reliability and convergence validity scores, Cronbach's α, composite reliability and AVE were equal to or below the recommended cut-off thresholds for the collection and protection uncertainty scale. Lastly, discriminant validity could not be established for any of the scales as the HTMT yielded values ≥ 0.95.

In addition to these weaknesses in reliability and validity, the model showed a poor global model fit (see Table 4). We analyze the global model fit by using the chi-square test, Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Tucker-Lewis-Index (TLI) and Standardized Root Mean Square Residual (SRMR) in their robust versions. We rely on commonly applied cut-off thresholds, i.e. close or below 0.08 for RMSEA, close or above 0.95 for CFI and TLI and close or below 0.08 for SRMR [18]. The chi-square test shows significant results. However, it is known to be sensitive to sample size and should therefore be assessed together with other global model fit indices. In addition, we find that the robust RMSEA is well above the recommended cut-off threshold indicating a poor fit of the model. Similarly, CFI and TLI do not show satisfactory global fit values.

Evaluation using CFA for Individual One-Factor Models Given that the CB-SEM measurement model showed weaknesses in reliability and validity, we could not confirm the hypothesized model structure. However, we are interested in whether people's perception of uncertainty varies among conditions

Table 3. Reliability and convergent validity results based on Cronbach's α, Composite Reliability (CR) and Averaged Variance Extracted (AVE). Common cut-off thresholds are above or equal 0.7 for Cronbach's α and CR and above or equal 0.5 for AVE.

Construct	Cronbach's α	CR	AVE
Adapted Privacy Uncertainty - CB-SEM			
Collection Uncertainty	0.70	0.69	0.41
Use Uncertainty	0.81	0.82	0.50
Protection Uncertainty	0.70	0.71	0.49
Overall Uncertainty	0.81	0.79	0.54
Adapted Privacy Uncertainty - Individual One-Factor Models			
Collection Uncertainty	0.70	0.72	0.42
Use Uncertainty	0.81	0.81	0.50
Protection Uncertainty	0.70	0.71	0.51
Overall Uncertainty	0.81	0.79	0.54
PANAS-X Fear - One-Factor Model			
Fear	0.95	0.94	0.79

and scenarios. Therefore, we evaluate the individual scales on their reliability and validity by conducting CFA on the four one-factor models. Again we investigate item reliability based on the factor loadings (see Table 7). Similarly to the CB-SEM measurement model, factor loadings of the individually fitted one-factor models showed sufficient reliability. Moreover, Cronbach's α values, composite reliability and convergent validity scores for the four individual scales were similar to the CB-SEM measurement model with low convergent validity for the collection uncertainty scale (see Table 3). Lastly, we investigate robust global model fit measures for the four individual one-factor models (see Table 4). For the collection uncertainty, the one-factor model shows an overall poor model fit. While the one-factor model for the use uncertainty scale showed satisfactory values for CFI, TLI and SRMR, RMSEA is above 0.05 indicating a good but not close fit. Model fit indices for the protection uncertainty one-factor model are not provided as the model is based on only three items and thus, just-identified with zero degrees of freedom. Lastly, the overall uncertainty one-factor model shows a non-significant chi-square test, a close fit based on RMSEA and satisfactory results for CFI, TLI and SRMR.

Based on this analysis, we recommend the usage of the overall uncertainty scale as it has proven reliable and valid in the context of CAI. We used ordinal logistic regression to investigate differences between conditions and scenarios for overall uncertainty ratings. However, our analysis did not show any differences in people's perceived levels of overall uncertainty.

4.2 Evaluation of the PANAS-X Fear Scale

We evaluate the PANAS-X scale related to fear on its reliability and validity by conducting a CFA. Factor loadings were generally high with values between

Table 4. Robust measurements of model fit for the adapted privacy uncertainty scale and the PANAS-X Fear scale.

Model	χ^2 / df (p-value)	RMSEA [90% CI]	CFI	TLI	SRMR
Adapted Privacy Uncertainty Scale					
CB-SEM	1068.32 / 98 (0.00)	0.12 [0.11-0.12]	0.86	0.83	0.07
Collection Uncertainty					
One-Factor Model	80.9 / 2 (0.00)	0.27 [0.22-0.32]	0.84	0.53	0.08
Use Uncertainty					
One-Factor Model	17.4 / 5 (0.00)	0.06 [0.03-0.10]	0.99	0.98	0.02
Protection Uncertainty - just-identified model					
Overall Uncertainty					
One-Factor Model	4.55 / 2 (0.1)	0.05 [0.00-0.1]	0.997	0.99	0.01
PANAS-X Fear					
One-Factor Model	34.99 / 9 (0.00)	0.09 [0.06-0.12]	0.99	0.98	0.02

0.86 and 0.90. Further, reliability and convergent validity showed satisfactory results (see Table 3). Lastly, the robust model fit indices suggest an overall good fit in accordance with commonly considered values (see Table 4). Nevertheless, the close-fit model criterion was not fulfilled as the robust RMSEA was greater than 0.05.

While the scale showed good reliability and validity values in the context of CAI and privacy decision-making, participants' ratings on this scale showed high variability, i.e. high standard variations across conditions. While subjective assessments of emotions allow inexpensive and efficient measurement, they might not correctly capture underlying psychological processes and have been discussed critically in the literature [9]. Even though the questionnaire was presented right after the interaction with the chatbot, present feelings at the time of filling out the survey might have outweighed feelings experienced during the interaction and led to inconsistent ratings. Moreover, some participants might have connected fear to specific factors, e.g., spiders or flying, while others did not. This might result in inconsistent usage of the scale. Instead of assessing a rather extreme feeling as fear, a future study could investigate feelings like discomfort or uneasiness to make assumptions about peoples' underlying thinking process.

5 Alternative Measures

5.1 Alternative Measures of Uncertainty

While we did not find differences in overall uncertainty and fear ratings in the context of Conversational AI, previous research successfully applied a subjective measure of uncertainty to prove the effectiveness of debiasing strategies [8]. Therefore, other subjective scales to measure the level of uncertainty might be better suited and could be tested in future studies. On the other hand, objective measures can be used to investigate perceived uncertainty. Objective measures,

Table 5. This table provides an overview of alternative measures that can be used to assess uncertainty. It also shows alternative measures for the evaluation of privacy strategies that aim at supporting rational decision-making.

Method	Alternative Uncertainty Measures	Alternative Evaluation Measures
Subjective	Self-assessed uncertainty based on scales not used in this study	Self-Assessed Mental Demand
		Privacy Regret
Objective	Reaction Time	Pupil Dilation
	Fixation Time	Galvanic Skin Response
	Neuroimaging	
Mixed		Attitude-Behavior Alignment

e.g. physiological measures, are beneficial as they can assess uncertainty at the time of decision-making. When using subjective measures, we asked participants to report their perceived uncertainty retrospectively after the interaction with the chatbot. However, uncertainty experienced during the interaction might be difficult to recall and present feelings might outweigh the previously experienced ones. Therefore, objective measurements can provide more reliable insights. We provide an overview of objective measures that have been used to assess uncertainty in Table 5.

A study on the detection of uncertainty researched physiological as well as behavioural measurements to sense uncertainty in interactive systems [14]. They identified keyboard behaviour, in particular the time of typing and the time looking at a question, as reliable indicators for uncertainty, whereas heart rate measurements did not provide useful information. Moreover, they suggested the usage of combined measurements to enhance reliability. Other studies have used functional magnetic resonance imaging (fMRI) to research brain activity patterns associated with judgements under uncertainty. Mushtaq et al. [26] provides an extensive overview of brain areas that were shown to be activated when uncertainty was involved in decision-making. The reviewed studies used a variety of tasks to manipulate the level of uncertainty, e.g. by varying the accuracy of predictors or by changing task rules. Thereby, studies include game-like tasks based on cards or checkerboards as well as more realistic tasks such as decision-making in a flight or driving simulator. Future studies could use neuroimaging to assess the level of uncertainty in a Conversational AI privacy scenario and compare the activation of brain areas with the ones previously identified.

5.2 Additional Considerations to Measuring Rational Decision-Making

We tested uncertainty as an evaluation measure for privacy strategies that are based on the idea of debiasing and aim for System 2 activation. In addition, there might be alternative evaluation measures to gain insights into participants' underlying thinking processes (see Table 5 for an overview of alternative evaluation

measures). In the medical field, the effectiveness of debiasing strategies has been assessed by evaluating error rates in diagnostic reasoning [20]. However, error rates do not seem to be a viable measure for debiasing strategies in the privacy context. As privacy decision-making is highly subjective and depends on the participants' attitudes, the "correct" outcome of a privacy decision remains unknown to the examiner. However, previous research on social media networks has shown that people are likely to regret the disclosure of private information as a result of intuitive thinking [12, 33]. Therefore, a subjective measure of privacy regret could be used for evaluating privacy strategies [1]. Moreover, privacy strategies that support people in overcoming their biases should lead to decisions that are aligned with people's attitudes. Consequently, the alignment of attitudes and behaviour could serve as an evaluation measure for such strategies [1].

In addition to uncertainty, the activation process can be accompanied by other factors, e.g. cognitive load, that can be measured either subjectively or objectively [17]. Various methods have been researched to assess cognitive load, both subjective and objective. These include self-reports on mental effort, eye-tracking and pupil dilation, or galvanic skin response [24]. Particularly pupil dilation has been investigated in various contexts to assess cognitive load with larger sizes indicating the usage of more cognitive resources [24, 25]. Therefore, future research could assess pupil dilation to evaluate the effectiveness of debiasing strategies.

While future studies can consider alternative measures to evaluate debiasing strategies, they can also make changes to the experimental design. First, our dialogue was designed so that no service was provided to the users due to apparent technical difficulties or the closure of the restaurant. This was 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. 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 – a feeling that can be accompanied by uncertainty [31]. This means that our negative outcome scenario could also lead to uncertainty which is not related to the effectiveness of the privacy strategies but to the outcome of the scenario. While we assessed frustration in the survey to account for it, our experimental choice might have overridden small differences between groups. Therefore, future research could avoid experimental setups where users experience frustration, e.g. by providing real-life services.

Second, future research could try to increase the expected effect of uncertainty by multiple or longer exposures to debiasing strategies in a dialogue. For example, previous research has shown that people report higher levels of uncertainty when being exposed to the Socratic procedure or the Devil's advocate approach [8]. Here, people are presented with multiple thought-provoking questions or opposing points of view. Such an experimental setup ensures that people experience competing intuitions which can lead to increased levels of uncertainty.

Similarly, guided reflection – a debiasing strategy known from the medical field – could be applied to practice more critical thinking in CAI privacy scenarios and could increase the expected effect of uncertainty [23]. Thereby, CAI could function as a guide or mentor and instruct users on what to consider in their privacy decision-making. Finally, when changing the experimental setup, multiple measures to assess the effectiveness of debiasing strategies should be considered to improve reliability.

6 Conclusion

We investigated perceived uncertainty as an evaluation measure for privacy strategies in Conversational AI. Our approach is theoretically grounded on the dual-process theory and previous research on the evaluation of debiasing strategies using a subjective measure of uncertainty. Our privacy strategies aim at supporting people in their privacy decision-making based on the idea of debiasing and conversational privacy. Thereby, they were designed to induce a controlled level of uncertainty and to trigger more analytical thinking. We used two subjective scales to investigate perceived uncertainty - an adapted privacy uncertainty scale and the PANAS-X scale on the affective state of fear. Only one subscale of the adapted privacy uncertainty scale, i.e. the overall uncertainty scale, and the scale on fear showed satisfactory reliability and validity results and can be recommended for future research in the context of CAI. As we did not find differences in peoples' perceived level of uncertainty on these two scales, we propose alternative measures to investigate uncertainty and evaluate privacy strategies that promote rational decision-making in the future.

Acknowledgments

Our work is partially funded by the German Federal Ministry for Economic Affairs and Energy as part of their AI innovation initiative (funding code 01MK20011A).

Bibliography

[1] Acquisti, A., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L.F., Komanduri, S., Leon, P.G., Sadeh, N., Schaub, F., Sleeper, M., Wang, Y., Wilson, S.: Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. ACM Computing Surveys **50**(3), 1–41 (May 2018), https://doi.org/10.1145/3054926

[2] Acquisti, A., Grossklags, J.: Uncertainty, ambiguity and privacy. In: Workshop on the Economics of Information Security (2005)

[3] Al-Natour, S., Cavusoglu, H., Benbasat, I., Aleem, U.: An Empirical Investigation of the Antecedents and Consequences of Privacy Uncertainty in the Context of Mobile Apps. Information Systems Research **31**(4), 1037–1063 (Dec 2020), https://doi.org/10.1287/isre.2020.0931

[4] Almuhimedi, H., Schaub, F., Sadeh, N., Adjerid, I., Acquisti, A., Gluck, J., Cranor, L.F., Agarwal, Y.: 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, p. 787–796, CHI '15, Association for Computing Machinery, New York, NY, USA (2015)

[5] Bermejo Fernandez, C., Chatzopoulos, D., Papadopoulos, D., Hui, P.: This Website Uses Nudging: MTurk Workers' Behaviour on Cookie Consent Notices. Proceedings of the ACM on Human-Computer Interaction **5**(CSCW2), 1–22 (Oct 2021)

[6] Brüggemeier, B., Lalone, P.: Perceptions and reactions to conversational privacy initiated by a conversational user interface. Computer Speech & Language **71**, 101269 (2022)

[7] Buçinca, Z., Malaya, M.B., Gajos, K.Z.: 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 (Apr 2021)

[8] Büyükkurt, B.K., Büyükkurt, M.D.: An Experimental Study of the Effectiveness of Three Debiasing Techniques*. Decision Sciences **22**(1), 60–73 (1991), https://doi.org/10.1111/j.1540-5915.1991.tb01262.x

[9] Ciuk, D., Troy, A., Jones, M.: Measuring emotion: Self-reports vs. physiological indicators. SSRN Electronic Journal (Jan 2015), https://doi.org/10.2139/ssrn.2595359

[10] Croskerry, P.: A Universal Model of Diagnostic Reasoning:. Academic Medicine **84**(8), 1022–1028 (Aug 2009), https://doi.org/10.1097/ACM.0b013e3181ace703

[11] Dash, G., Paul, J.: CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting. Technological Forecasting and Social Change **173**, 121092 (Dec 2021), https://doi.org/10.1016/j.techfore.2021.121092

[12] Díaz Ferreyra, N.E., Meis, R., Heisel, M.: Learning from Online Regrets: From Deleted Posts to Risk Awareness in Social Network Sites. In: Ad-

junct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, pp. 117–125, ACM, Larnaca Cyprus (Jun 2019), https://doi.org/10.1145/3314183.3323849
[13] 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)
[14] Greis, M., Karolus, J., Schuff, H., Woźniak, P.W., Henze, N.: Detecting uncertain input using physiological sensing and behavioral measurements. In: Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia, pp. 299–304, ACM, Stuttgart Germany (Nov 2017), https://doi.org/10.1145/3152832.3152859
[15] Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P., Ray, S.: Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook. Classroom Companion: Business, Springer International Publishing, Cham (2021), https://doi.org/10.1007/978-3-030-80519-7
[16] Harkous, H., Fawaz, K., Shin, K.G., Aberer, K.: PriBots: Conversational privacy with chatbots. In: Twelfth Symposium on Usable Privacy and Security (SOUPS 2016), USENIX Association, Denver, CO (Jun 2016)
[17] Kahneman, D.: Thinking, fast and slow. Farrar, Straus and Giroux, New York (2011)
[18] Kline, R.B.: Principles and practice of structural equation modeling. Guilford publications (2015)
[19] Knight, F.H.: Risk, Uncertainty and Profit (1921)
[20] Lambe, K.A., O'Reilly, G., Kelly, B.D., Curristan, S.: Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. BMJ Quality & Safety **25**(10), 808–820 (Oct 2016)
[21] Lau, J., Zimmerman, B., Schaub, F.: Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. Proc. ACM Hum.-Comput. Interact. **2**(CSCW) (Nov 2018)
[22] Leschanowsky, A., Brüggemeier, B., Peters, N.: 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, pp. 76–79 (11 2021), https://doi.org/10.21437/SPSC.2021-16
[23] Leschanowsky, A., Popp, B., Peters, N.: Adapting debiasing strategies for conversational ai. Zagreb, Croatia p. 74 (2022)
[24] Martin, S.: Measuring cognitive load and cognition: metrics for technology-enhanced learning. Educational Research and Evaluation **20**(7-8), 592–621 (2014), https://doi.org/10.1080/13803611.2014.997140
[25] Mirhoseini, M., Early, S., Hassanein, K.: All eyes on misinformation and social media consumption: A pupil dilation study. In: Information Systems and Neuroscience: NeuroIS Retreat 2022, pp. 73–80, Springer (2022)
[26] Mushtaq, F., Bland, A.R., Schaefer, A.: Uncertainty and Cognitive Control. Frontiers in Psychology **2**, 249 (Oct 2011), https://doi.org/10.3389/fpsyg.2011.00249

[27] Neys, W.D.: Advancing theorizing about fast-and-slow thinking. Behavioral and Brain Sciences pp. 1–68 (Sep 2022), https://doi.org/10.1017/S0140525X2200142X, publisher: Cambridge University Press

[28] Pearman, S., Young, E., Cranor, L.F.: User-friendly yet rarely read: A case study on the redesign of an online HIPAA authorization. Proceedings on Privacy Enhancing Technologies **2022**(3), 558–581 (Jul 2022), https://doi.org/10.56553/popets-2022-0086

[29] Popp, B., Lalone, P., Leschanowsky, A.: Chatbot language – crowdsource perceptions and reactions to dialogue systems to inform dialogue design decisions. Journal for Behavior Research Methods (2022)

[30] Schomakers, E.M., Lidynia, C., Müllmann, D., Ziefle, M.: Internet users' perceptions of information sensitivity – insights from Germany. International Journal of Information Management **46**, 142–150 (Jun 2019)

[31] Smith, C., Ellsworth, P.: Patterns of cognitive appraisal in emotion. Journal of personality and social psychology **48**, 813–38 (May 1985), https://doi.org/10.1037//0022-3514.48.4.813

[32] Tversky, A., Kahneman, D.: Judgment under uncertainty: Heuristics and biases. Science **185**(4157), 1124–1131 (1974), https://doi.org/10.1126/science.185.4157.1124

[33] Wang, Y., Leon, P.G., Acquisti, A., Cranor, L.F., Forget, A., Sadeh, N.: A field trial of privacy nudges for facebook. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2367–2376, ACM, Toronto Ontario Canada (Apr 2014), https://doi.org/10.1145/2556288.2557413

[34] Wang, Y., Leon, P.G., Scott, K., Chen, X., Acquisti, A., Cranor, L.F.: Privacy nudges for social media: An exploratory facebook study. In: Proceedings of the 22nd International Conference on World Wide Web, p. 763–770, WWW '13 Companion, Association for Computing Machinery, New York, NY, USA (2013)

[35] Watson, D.B., Clark, L.A.: The PANAS-X manual for the positive and negative affect schedule (1994)

Appendix

Table 6. Rephrased items for the four uncertainty subscales and parameter estimates for the CB-SEM measurement model. Parameters were estimated using WLSMV and first indicator factor loadings were set to unity. Robust standard errors are computed.

Abbr.	Construct	Estimate	Std. Err	z-value	P(> |z|)	Standardized Loadings
Coll1	I was uncertain about what information will be collected	1.00				0.75
Coll2	I was concerned about the amount of information that was collected by the chatbot	0.77	0.04	20.54	0.00	0.58
Coll3	I was afraid the chatbot would collect more information than I was initially told	0.95	0.03	34.05	0.00	0.72
Coll4	I was concerned that I will have to provide more information than I originally thought	0.67	0.04	16.29	0.00	0.50
Use1	I was concerned about how the chatbot provider would use the information that was recorded by the chatbot	1.00				0.57
Use2	I was uncertain about who would have access to the information tha was recorded	1.28	0.06	20.54	0.00	0.73
Use3	I was worried that the information that was recorded will be shared with others	1.35	0.06	20.97	0.00	0.76
Use4	I was unsure if the information that was recorded might be misused	1.27	0.06	20.56	0.00	0.72
Use5	I was afraid that if given the chance the chatbot provider might profit by selling the information to someone else	1.33	0.06	21.73	0.00	0.76
Prot1	I was concerned that the information that was collected will not be protected	1.00				0.58
Prot2	I was uncertain about what the chatbot provider would to to ensure that the information collected was secure	1.28	0.06	23.47	0.00	0.74
Prot3	I was unsure if the chatbot provider would effectively safeguard the information that was collected	1.34	0.05	24.95	0.00	0.77
All1	Overall, I was unsure if the chatbot provider would safeguard my privacy	1.00				0.80
All2	Overall, I was uncertain if the chatbot provider would be good at managing my private information	0.92	0.03	35.12	0.00	0.73
All3	Overall, I was worried if my information would be safe with the chatbot provider	0.97	0.22	44.09	0.00	0.77
All4	Overall, I was concerned that the chatbot provider might breach formal and informal privacy agreements	0.79	0.03	24.68	0.00	0.63

Table 7. Rephrased items for the four uncertainty subscales and parameter estimates for the individual One-Factor CFA models. Parameters were estimated using WLSMV and first indicator factor loadings were set to unity. Robust standard errors are computed.

Abbr.	Construct	Estimate	Std. Err	z-value	P($>\|z\|$)	Standardized Loadings
Collection Uncertainty						
Coll1	I was uncertain about what information will be collected	1.00				0.66
Coll2	I was concerned about the amount of information that was collected by the chatbot	0.96	0.06	15.02	0.00	0.63
Coll3	I was afraid the chatbot would collect more information than I was initially told	1.06	0.07	14.56	0.00	0.70
Coll4	I was concerned that I will have to provide more information than I originally thought	0.89	0.06	13.93	0.00	0.59
Use Uncertainty						
Use1	I was concerned about how the chatbot provider would use the information that was recorded by the chatbot	1.00				0.44
Use2	I was uncertain about who would have access to the information tha was recorded	1.53	0.12	12.43	0.00	0.68
Use3	I was worried that the information that was recorded will be shared with others	1.82	0.14	12.93	0.00	0.81
Use4	I was unsure if the information that was recorded might be misused	1.72	0.13	12.83	0.00	0.76
Use5	I was afraid that if given the chance the chatbot provider might profit by selling the information to someone else	1.78	0.14	12.84	0.00	0.79
Protection Uncertainty						
Prot1	I was concerned that the information that was collected will not be protected	1.00				0.50
Prot2	I was uncertain about what the chatbot provider would to to ensure that the information collected was secure	1.37	0.10	13.73	0.00	0.68
Prot3	I was unsure if the chatbot provider would effectively safeguard the information that was collected	1.81	0.17	10.48	0.00	0.90
Overall Uncertainty						
All1	Overall, I was unsure if the chatbot provider would safeguard my privacy	1.00				0.80
All2	Overall, I was uncertain if the chatbot provider would be good at managing my private information	0.94	0.04	23.09	0.00	0.75
All3	Overall, I was worried if my information would be safe with the chatbot provider	1.00	0.05	22.10	0.00	0.80
All4	Overall, I was concerned that the chatbot provider might breach formal and informal privacy agreements	0.74	0.05	16.15	0.00	0.59