Enabling Users to Specify Correct Privacy Requirements

Manuel Rudolph, Svenja Polst and Joerg Doerr

Fraunhofer IESE, Kaiserslautern, Germany
{first.last}@iese.fraunhofer.de

Abstract. Privacy becomes more and more important for users of digital services. Recent studies show that users are concerned about having too little control over their personal data. However, if users get more possibilities for self-determining the privacy effecting their personal data, it must be guaranteed that the resulting privacy requirements are correct. This means, they reflect the user's actual privacy demands. There exist multiple approaches for specifying privacy requirements as an end user, which we call specification paradigms. We assume that a matching of specification paradigms to users based on empirical data can positively influence the objective and perceived correctness. We use the user type model by Dupree, which categorizes users by their motivation and knowledge. We experimentally determined the best match of user types and paradigms. We show that participants with less knowledge and motivation make more mistakes and that a strong limitation of selection options increases objective and perceived correctness of the specified privacy requirements.

Keywords: Privacy Requirements Specification, User Types, Specification Interfaces, Objective Correctness, Perceived Correctness.

1 Introduction

Since the dawn of the Internet age, users have been increasingly sending (personal) data to services that process and analyze data. At the same time, users become increasingly aware and partially afraid of data misuse and their need for a better privacy protection raises [1,2]. Even if the need arises, many users do not configure their privacy settings for Internet services. One major cause is that users have problems in adequately specifying their own privacy requirements, which we showed in a previous study [3]. Users rate the specification as too complicated and time consuming. In practice, services provide different specification interfaces, which offer the user a variety of options, specification processes and guidance during the specification of privacy requirements. We refer to those different types of interfaces as specification paradigms. In order to achieve ideal results, we need to provide users a specification paradigm that matches their needs and capabilities best. We assume that the appropriate selection of the specification paradigm for a user can have a positive effect on the acceptance of the tool itself, and can increase its effectiveness. Thus, we investigated the effectiveness of the privacy requirement specification (objective and perceived correctness of the specified requirements), efficiency (necessary time span for specification) and user satisfaction (how much users like the paradigm). Our results regarding user satisfaction and efficiency were published in [20]. In this article, we focus on the effectiveness. The susceptibility to mistakes should always be of particular interest. Users are very different with respect to their capabilities (e.g., knowledge, available time and cognitive capacity) and preferences (interaction processes they like). Thus, there will probably not be a specification paradigm that delivers the best results for all user types. We use the model of Dupree for clustering users into user types [5]. Whether a paradigm fits a user depends on his specific characteristics. The lack of work on matching specification paradigms to user types motivated this work. Our main contributions in this article are observations and recommendations for best suitable specification paradigms for specific user types regarding effectiveness. They are derived from an experiment in which we asked users representing different personas to solve tasks with four specification paradigms. We measured mistakes produced by the users and the users' perception of correctness.

In this paper, we present the used specification paradigms and their derivation from literature and practice in Section 2. In Section 3, we discuss available user type model in literature and discuss the selection of the Dupree model. Next, we explain the design and execution of our experiment in Section 4. We present and discuss the results in Section 5. Finally, we conclude and discuss future work in Section 6.

2 THE VARIETY OF PRIVACY SPECIFICATION INTERFACES

Users specify their privacy requirements as policies in different systems using specification interfaces. Depending on the system, different types of specification interfaces are offered, which we call specification paradigms. These differ in following aspects:

- Specification process: With which interactions do users set their privacy requirements in the interface?
- Number of decisions: How many decisions do users have to take in the specification?
- Degree of guidance: How much support is given to users during specification?

In the following, we identified relevant privacy specification approaches and interfaces in the state of the art and practice and derived appropriate specification paradigms.

2.1 Related Work regarding End-user Privacy Specification Interfaces

In the state of the art, a lot of work was performed in the area of specifying privacy requirements in form of machine-understandable policies by experts. Even if the focus of our work is to enable non-experts to specify privacy requirements in natural language, the interface concepts for machine-understandable policies can be transferred to natural language interfaces for privacy policy specification.

PERMIS [13] is a generic RBAC-based (Role-Based Access Control) authorization infrastructure. PERMIS policies are created, for example, via a “Policy Wizard”. This tool uses a step-by-step specification wizard as the policy specification paradigm. It asks supportive questions to guide the user through the specification process. KAoS [15] is a policy and domain service framework. It contains the KAoS Policy Administration Tool (KPAT) that is based on natural English sentences using hypertext templates. Policy templates are specified in an ontology and specified policies are automatically transformed into machine-readable equivalents. Johnson et al. [14] describe a method and a tool named SPARCLE for eliciting concrete security requirements of users with varying background knowledge. The tool allows the user to enter his security requirement in natural language or in a structured natural language-based format. SPARCLE can transform the structured format into machine-understandable policies. P3P (Platform for Privacy Preference Project) is a protocol that allows websites to declare their intended use of information they collect from users [18]. In addition, APPEL (A P3P Preference Exchange Language) was developed for users to describe collections of privacy preferences [19]. Fang and LeFevre [17] propose an active learning wizard that enables users to set their own privacy and security policies by making regular, brief decisions on whether or not to share a particular data item with an entity.

Besides the academic approaches, many domain specific policy authoring tools exist in practice. The Local Group Policy Editor of Windows systems (e.g., Windows 7) mainly targets system administrators and offers a variety of settings (e.g., firewall settings, password policies, startup/shutdown scripts) for Windows environments. Facebook allows its users to specify their privacy requirements in a very fine-grained manner. Even if studies revealed that users expected in some cases a different behavior from the specified privacy policies [16], they are in general empowered to specify them at all. Both tools, the Windows editor as well as the Facebook privacy settings, provide a lot of specification support, such as explanations or examples. They use template based specification and small wizards for specific security and privacy settings. All modern browsers contain privacy and security settings. Google Chrome (Version 64), Microsoft Edge (Version 41) and Mozilla Firefox (Version 52) allow their users to enable and disable pre-defined default privacy and security policies. The Microsoft Internet Explorer (Version 11) uses a security level approach for setting the coarse-grained security settings. If required, users can customize these security levels by selecting from predefined default options. For the privacy requirements in online accounts, Google has introduced a privacy check wizard that guides the user through multiple pages to configure the use of personal information by Google services and third parties.

2.2 Selected Specification Paradigms

We found that all specification paradigms from literature and practice differ in their configurability (how many decisions they request) and their guidance (how much help does the user receive during the specification). We rated all specification paradigms accordingly and selected paradigms (All Screens displayed to the subjects and further supplementary material such as sample solution and access to primary data can be found in [22]) with all four combinations of high and low configurability (C) and guidance (G):

1. Template instantiation (high-C, low-G): The user can instantiate desired privacy requirements by adjusting selection options in a template-based interface. The templates offer multiple decisions and thus allow a fine-grained specification of own privacy requirements. The user can choose the order of specification.
2. Default Policies (low-C, low-G): The user can chose from multiple predefined privacy policies per topic. The number of decisions in the specification is limited.
3. Wizard (high-C, high-G): The user can instantiate privacy requirements based on a template-based interface, which is split in several small steps. The user cannot decide on the specification order. The specification process is well guided in each step.
4. Security levels (low-C, high-G): The user can select a level of privacy that contains a predefined set of default privacy requirements without having customization possibilities per requirement.

3 THE DIFFERENT TYPES OF USERS

Each user has different characteristics, capabilities and resources. This leads us to the assumption that different paradigms are likely to fit differently well to a certain user. To explore the relationship between suitable specification paradigms and users, we first explored related work regarding user type models and then selected a model for clustering users according to relevant characteristics.

3.1 Related Work regarding User Type Models

There are several ways to cluster users into categories that explain their character traits and behavior. Some clustering methods describe human traits and behavior in general, i.e., they are not bound to a particular situation or domain. Examples are the Big Five personality traits [6], Keirsey's Temperaments [7] and the Myers-Briggs Type Indicators [8]. Besides the generic clustering approaches, other work relates to the use of computers and the character traits relevant for security and privacy decisions. For example, Westin’s classification is based on users’ privacy concerns. In most of his 30 privacy surveys, he clusters the users into three categories: Fundamentalist (high concern), Pragmatist (medium concern), and Unconcerned (low concern). Westin’s approach is controversially discussed in the literature. For example, Urban and Hoofnagel [9] argue that Westin’s work is neglecting the importance of knowledge or available information about privacy practices, domain specific business processes. Smith’s approach “Concern for Information Privacy (CFIP)” [10] measures the privacy concern of a person as a numerical value based on a calculation on fifteen statements about privacy. The scenarios of CFIP are kept quite abstract and do not directly relate to online services that collect and process user data. Malhotra et al. improved and extended previous work (e.g., CFIP) in their approach called Internet Users' Information Privacy Concerns (IUIPC) [12]. They reflect the concerns of internet users about information privacy with a special focus on the individuals’ perception of fairness in the context of data privacy. Morton’s Information Seeking Preferences [11] are an approach to cluster users into five groups based on the ranking of 40 privacy related statements. The groups are: Information controllers, security concerned, benefit seekers, crowd followers and organizational assurance seekers. Considering the criticism on Westin's privacy indexes, Dupree proposed her privacy personas [5]. Those five personas can be differentiated on two attributes of the user: the user's knowledge about security and privacy as well as the user's motivation to spend effort to protect privacy and security. The personas also describe the handling of personal data in the internet age and the general need for security in the IT sector.

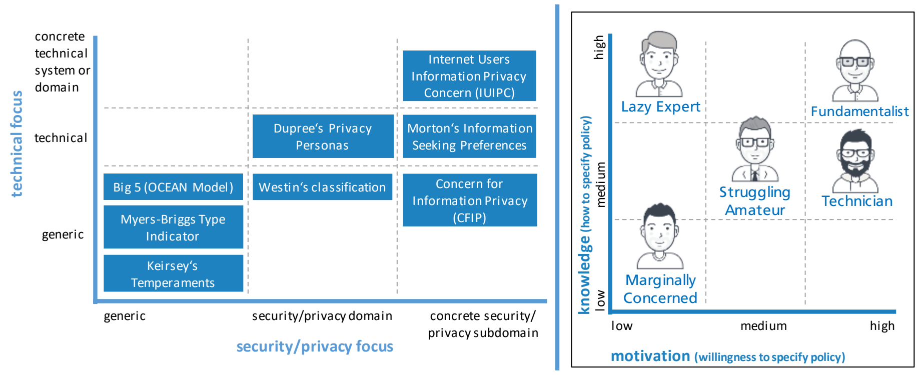
technical focus
concrete technical system or domain
technical
generic
Internet Users Information Privacy Concern (IUIPC)
Dupree's Privacy Personas
Morton's Information Seeking Preferences
Big 5 (OCEAN Model)
Westin's classification
Concern for Information Privacy (CFIP)
Myers-Briggs Type Indicator
Keirsey's Temperaments
generic
security/privacy domain
concrete security/ privacy subdomain
security/privacy focus
knowledge (how to specify policy)
high
medium
low
Lazy Expert
Fundamentalist
Struggling Amateur
Technician
Marginally Concerned
low
medium
high
motivation (willingness to specify policy)

Fig. 1. Left: Classification of different user type models | Right: Dupree' Persona Matrix

3.2 Selection of the User Type Model

When searching for the appropriate model, we found that all available models can be characterized by two properties: focus on IT security and privacy and focus on technical systems (see **Fig. 1** left). In both cases, there are very special models developed for a specific subdomain or system as well as generic approaches. We chose the Dupree model [5] as a suitable middle way. This model mainly distinguishes users by their motivation and their knowledge to specify privacy requirements (see **Fig. 1** right). Dupree has derived the five personas from personal interviews with 32 university related digital natives, who had an average age of 26.3 (SD = 5.9). The personas are:

- Marginally Concerned: Low knowledge and low motivation
- Amateur: Medium knowledge and medium motivation
- Technician: Medium knowledge and high motivation
- Lazy Expert: High knowledge and low motivation
- Fundamentalist: High knowledge and high motivation

4 EXPERIMENT DESIGN AND EXECUTION

4.1 Research Questions

The experiment objective was to identify which paradigms are suitable for a specific persona with regard to objective and perceived correctness. Each paradigm requires the user to make a certain number of decisions during the specification of privacy requirements. If the decision taken differs from the sample solution, we regard this deviation as a mistake. We consider a paradigm to be suitable if the ratio of mistakes to all decisions is low (high objective correctness). Moreover, we aimed at finding the best matching paradigm for a precise self-estimation with respect to the objective correctness (Can people estimate that they made mistakes?). We defined following research questions:

- RQ1: Which paradigm best suits a particular type of person (represented by a persona) in terms of objective correctness?
- RQ2: Which paradigm is best suited to a particular type of person (represented by a persona) in terms of correctly estimated perceived correctness (confidence regarding objective correctness)?

4.2 Scenario & Tasks

The scenario and the corresponding privacy requirements in the experiment were derived from a real project in the context of the digitization of rural areas using the RE method described in [21]. In this method, workshops with users and experts of the problem domain are conducted with selected State of the Art RE methods in order to elicit relevant templates of privacy requirements. In the project, village citizens have access to digital services such as an online marketplace with local merchants, a delivery service where citizens deliver goods from local merchants to other citizens (called BestellBar) and a digital village bulletin board. The participants should imagine that they use these novel, digital services of this project and that this has potential privacy impact to them as personalized data is used in those services. The participants had the task to adjust the privacy requirements of these services to given privacy requirements. The requirements were not their own but specified by the authors of this paper. The presetting of the privacy requirements was necessary so that all participants could use the specification interfaces in a comparable way. This enabled us to compare the measured mistakes made by the participants.

The requirements were described as part of the six tasks. One task was, for instance; "When I place an order in the BestellBar app, I do not under any circumstances want to receive advertising from other providers that refers to the ordered product. They may not use my data." The requirements did not match one-to-one with the wording in the specification interfaces, because a one-to-one match would cause that the participants compare the buzzwords of the task and the interfaces but not the semantic content.

The scenario description and the tasks were provided on a digital handout, which the participants were advised to print out. The scenario description was supported by a short video that introduces the novel, digital services for citizens of a village. Four specification interfaces were created according to the selected specification paradigms presented in Section 2.2. We refer to these interfaces as the four specification paradigms in the following. The participants had to complete the same six tasks for each specification paradigm. The introduction material is presented in the supplementary experiment material [22].

All implementations of the specification paradigms in this experiment use the same templates, which is the outcome from the used RE method [21]. The paradigms template instantiation and wizard let the participant instantiate concrete privacy requirements from the templates. The paradigms default policies and security levels provide a limited list of already instantiated privacy requirements from the templates to choose from. In case of the paradigm security levels, the user can chose from three different sets of privacy requirements. All tasks in the experiment can be solved with all four specification paradigm implementations.

During the experiment design we had to decide whether we should provide a perfect match with the tasks for the paradigm security levels. This means that one of the security levels solves all tasks of the scenario. Such a perfect match is unlikely in real life. However, the lack of a perfect solution could confuse the participants in the experiment letting them abort. In addition, a massive influence on the experiment results (correctness and satisfaction) was expected. Thus, we decided to have a perfect match because we did not want to compromise the proper execution of the experiment.

4.3 Procedures and Instruments

Our experiment was created as a publically available online experiment. In order to avoid misuse, a participant could only start the experiment once with a unique eight digit participant id. It was possible to interrupt the experiment and continue with the participant id in the same place. However, it was not possible to repeat already executed steps. The experiment was provided in German and English.

Our experiment was structured as follows. First of all, the participants had to agree to an informed consent and confirm that they are at least 18 years old. Thereafter, the participants had to answer demographic questions about age, gender and educational level as well as their relationship to the authors' institutions and their research topics. The answers were used to determine whether the participants' characteristics and capabilities have an impact on the results of the experiment. Then, a self-assessment followed about one's own expertise and motivation in the areas of IT security and protection of one's own privacy as well as experience in dealing with digital services. Afterwards, the participants were asked to select the persona out of the five offered personas that they think fits best to them. All five personas of Dupree were described on the basis of nine to twelve original character traits [5] formulated in the ego-perspective. The order of the personas displayed was randomly determined. Thereupon, the scenario including the concrete tasks (privacy requirements) was explained by video and handout. Next, the participants were instructed that on the following pages they should set all the privacy requirements for each of the four different specification paradigms: default policies, security levels, template instantiation and wizard. The order in which the specification paradigms were presented to the participants was randomly determined to minimize learning effects. After each specification paradigm, the participants were asked whether they thought they did mistakes, how they liked the current type of specification in the current scenario and how they would like it transferred to real life. After completing the four specifications, the participants were asked to rank the four specification types according to their preference of using them in real life. Finally, participants should determine how well they can identify with the scenario and the chosen persona. Screenshots showing all steps of the experiment can be found in the supplementary experiment material [22].

4.4 Execution

We acquired the participants by means of a non-binding invitation by e-mail in the circle of friends and acquaintances of the authors as well as in the authors' institution. The participants were asked to forward the non-binding invitation to other persons. We sent each interested person a specific invitation email with a handout attached. The handout contained instructions for starting and conducting the experiment, the individual participant id and the scenario description. We sent 120 personal invitation emails and deleted them directly after sending in order ensure the anonymity of the participants. The online experiment was available for 14 days. Participants were informed about the approximate duration of the experiment of 30-40 minutes, but had no time limit for completion.

4.5 Data Analysis

All statistical analyses were conducted with SPSS 19 and Microsoft Excel. First of all, the plausibility of the self-selection of personas was checked by analyzing whether the self-reported security knowledge matches the persona classification by Dupree (see [5]). Moreover, we analyzed how well participants identified with the selected persona.

To answer RQ1, the number of mistakes was analyzed. The different paradigms required different numbers of decisions: One decision in security levels, six decisions in default policies, 18 decisions in template instantiation and 18 decisions in wizard. This means that the pure number of mistakes is not directly comparable, but the ratio of incorrect decisions had to be compared. To evaluate the differences between the paradigms, Wilcoxon signed rank test were used. We also performed a Kruskal-Wallis (suitable for small sample sizes) test ($\alpha = 0.05$) to investigate whether the persona has an influence on the objective correctness. The fundamentalist were excluded from analysis because of their small number.

To answer RQ2, we investigated whether there is an influence of the persona on the perceived correctness or not. The perceived correctness was measured by asking the participants after the use of each paradigm whether they think that they solved all tasks in the paradigm correctly (zero mistakes). A Fisher's exact test, which is a test for small sample sizes, was performed for the results of each paradigm.

5 RESULTS AND DISCUSSION

5.1 Participant Description

Out of 120 invitations sent, 61 persons finished the experiment with complete data sets. We did not find any indications that would have caused us to consider records as invalid. 43 percent of the participants are female. The participants' age ranges from 18 to 82 (M=40.54; SD=14.37). The majority of the participants (33 out of 61) hold a university degree as highest educational level, nine participants hold a doctoral degree, seven have an entrance qualification for higher education and eleven a secondary school leaving certificate as highest level of education. About half of the participants (54%) were related the authors' institution, 20 of them being scientific and eight non-scientific employees and five being students working with the authors' institution. 28 participants (46%) had no relation to the authors' institution. **Table 1** shows the distribution of the personas chosen by the participants. The largest group with 34 percent of the participants is the persona amateur. The fundamentalists make up the smallest group with five percent. The ratio of the other personas varies between 18 and 23 percent.

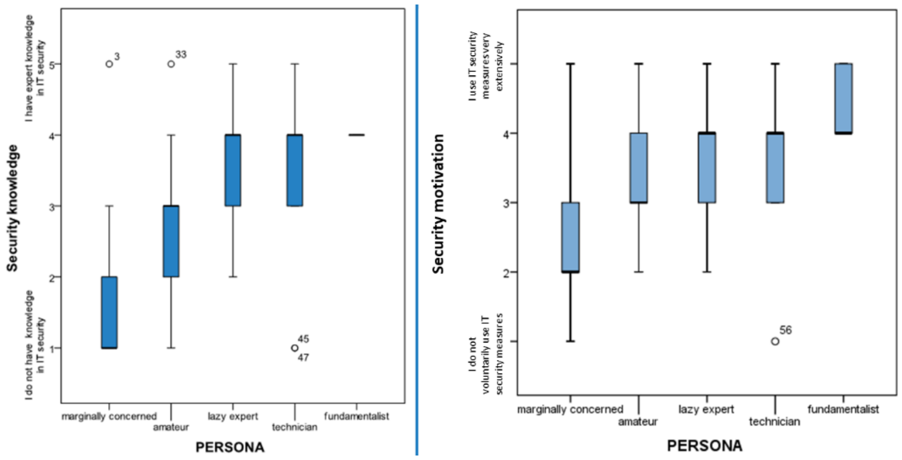

Fig. 2. Left: Knowledge to persona mapping | Right: Motivation to persona mapping

To verify the plausibility of the persona self-selection, we asked the participants to rate their IT security knowledge. The participants' security knowledge fits well to the chosen personas, except for the lazy experts (see **Fig. 2** left side). Based on Dupree's categorization (see **Fig. 1** right side), we expected the lazy experts to have higher self-estimated knowledge. The participants' security motivation fits to the model of Dupree as well (see **Fig. 2** right side). Moreover, we asked the participants, how well the chosen persona matches them on a scale from 1 (Not very well, but it matched best out of the five options) to 5 (I can identify myself very well with the persona). The participants responded on average with a score of 3.75. Not a single person reported the value 1.

Table 1. Chosen Personas

Persona	Number	Ratio	Persona	Number	Ratio
Marginally concerned	12	20%	Technician	14	23%
Amateur	21	34%	Fundamentalist	3	5%
Lazy expert	11	18 %	Total	61	

5.2 Experiment Results

The results regarding the objective correctness are presented first. Thereafter, the results of the perceived correctness in relation to the objective correctness are shown.

Table 2. Participants with 100% objective correctness and mistakes made by personas

Persona	Number of participants with all paradigms correct / n per persona	% of participants per persona	Mistakes made in relation to decisions				
				Default Policies	Security Levels	Template Instantiation	Wizard
			Degrees of Freedom	6	1	18	18
Marginally Concerned	1 / 12	8.33%	Average Mistakes	0.56	0.25	0.49	0.50
			Std. Deviation	0.36	0.45	0.29	0.29
Amateur	4 / 21	19.05%	Average Mistakes	0.12	0.05	0.12	0.12
			Std. Deviation	0.22	0.22	0.16	0.14
Lazy Expert	1 / 11	9.09%	Average Mistakes	0.15	0.00	0.16	0.21
			Std. Deviation	0.26	0.00	0.16	0.21
Technician	4 / 14	36.36%	Average Mistakes	0.17	0.00	0.15	0.11
			Std. Deviation	0.27	0.00	0.25	0.16
Fundamentalist	0 / 3	0%	Average Mistakes	0.00	0.00	0.06	0.13
			Std. Deviation	0.00	0.00	0.06	0.08
All participants	10 / 61	16.39%	Average Mistakes	0.22	0.07	0.20	0.21
			Std. Deviation	0.31	0.25	0.25	0.24

Objective Correctness. Different aspects were taken into account in the analysis of the objective correctness (see **Table 2**): First, we identified the number of the participants with perfect objective correctness. Secondly, the concrete number of mistakes in relation to the decisions per paradigm were analyzed. Fewest mistakes were made with security levels. Seven percent of participants chose the wrong security level. In the other three paradigms, which provided more decision options, about one in five decisions were taken incorrectly. Thus, for the whole population of the experiment there is no difference in objective correctness, except for a significant difference to the paradigm security levels (compared to default: $z= 3.83$, $p<0.01$, template: 4.22, $p<0.01$, wizard: 4.35, $p<0.01$). Only 10 out of 61 participants made no mistakes, thus they achieved 100 percent objective correctness in all paradigms.

The persona selection has a significant effect on the mistakes made in the paradigms default policies ($\chi^2 = 13.88$, $p < 0.01$), template instantiation ($\chi^2=14.10$, $p < 0.01$), and wizard ($\chi^2=17.04$, $p < 0.01$), and also on the security levels ($\chi^2=7.99$, $p <0.05$) but not that strong. The effect of the persona is likely given because of the significant difference of the marginally concerned to the other personas. For example, within the paradigm default policies, the amount of mistakes by the marginally concerned is significantly higher compared to the other personas (for each persona p<0.05). The effect sizes for all paradigms are strong (d<0.6; see details about the statistical results in the supplementary experiment material [22]).

Perceived Correctness in Relation to Objective Correctness. We asked the participants after each paradigm they used, whether they think that they solved all tasks correctly. The experiment results provide that the persona selection does not influence the perceived correctness in any paradigm (template: p=0.96; default: p=0.87; security level: p=0.85; wizard: p=0.62). This means that there is no difference in how optimistic or pessimistic the participants of the different personas are regarding these paradigms. In our experiment, we aimed at identifying which paradigm suits best for a correct self-estimation (perceived correctness) regarding the objective correctness. A self-estimation of a privacy requirements specification is rated as correct, if the participant did zero mistakes and was confident about the perfect solution or if the participant did at least one mistake and was confident that he did mistakes. Overall, 42 participants thought that they used all paradigms correctly, however, only eight of them made indeed no mistakes in all paradigms. Twelve persons reported mistakes in one paradigm and two persons even in all four paradigms. Thus, the perceived correctness is very high, regardless of the many mistakes that were made. Only four persons had a too pessimistic self-estimation. **Table 3** shows the correct estimations per paradigm for all participants and for each persona. Overall, the self-estimation was best with the security levels (78.7%) and worst with the wizard (29.5%). We found that more decisions during specification led to worse self-estimation.

Table 3. Accuracy of perceived correctness (Correct positive (P) and negative (N) estimations)

	Default Policies		Security Levels		Template Instantiation		Wizard	
	P/N	%	P/N	%	P/N	%	P/N	%
Marginally Concerned	2/1	25.0	8/1	75.0	1/2	25.0	1/0	8.3
Amateur	12/1	61.9	16/1	81.0	6/1	33.3	6/1	33.3
Lazy Expert	7/2	81.8	8/0	72.7	2/1	27.3	1/2	27.3
Technician	8/1	64.3	12/0	85.7	6/2	57.1	6/1	50.0
Fundamentalist	3/0	100	2/0	66.7	1/0	33.3	0/0	0.0
	32/5	60.7	46/2	78.7	16/6	36.1	14/4	29.5

Comparison of Results regarding Personas. The **marginally concerned** made using the security levels paradigm least mistakes and achieved best perceived correctness compared to other paradigms (Average Mistakes (AM): 25%, see **Table 2**; Correct Estimations (CE): 75%, see **Table 3**)). In all other paradigms, this group of people made more mistakes. The **amateurs** also achieved best results with the security levels (AM: 5%; CE: 81%). For the other paradigms, the AM values are equal at 12%. Regarding the perceived correctness, participants assessed themselves rather good with the default policies (CE 61.9%). Amateurs did rather few mistakes with the paradigms template instantiation and wizard, but the self-assessment is worse than with other paradigms. The **technician** achieved as all other personas better results in the paradigms security levels (AM: 0%; CE: 86%) and default policies (AM: 17%; CE: 64%). However, the technicians achieved best values regarding the perceived correctness and rather low rates of mistakes for the paradigms template instantiation (AM: 15%; CE: 57%) and wizard (AM: 11%; CE: 50%). The **lazy experts** are described by Dupree as people with a high level of knowledge and low motivation in terms of security and privacy (see **Fig. 1** right side). It is interesting to note that they performed worse than amateurs and technicians in many direct value comparisons. The values for the default policies (AM: 15%; CE: 81.8%) and security levels (AM: 0%; CE: 72.7%) are best. Since only three participants have chosen the persona **fundamentalist**, no conclusions can be made about this persona. Still, the results reflect the persona scheme of Dupree [5].

5.3 Threats to Validity

We did not control the participants during or after the experiment, which is a threat to internal validity. We cannot exclude the possibility that the participants talked about the experiment with other participants before their participation, nor that the participants could not find the necessary information or concentration to solve the tasks adequately. Distraction might increase the number of mistakes. However, we adequately instructed participants with a text handout, a scenario video and instructions in various steps in the experiment as we would have done in a controlled setting. We did not find any hint for an inadequate introduction (e.g., in the feedback at the end of the experiment). Thus, we assess this threat as low. A participant who could not identify with the provided privacy requirements well, maybe had lower motivation to take effort in correctly using the paradigms in the experiment. This may negatively affect the objective correctness and is a threat to internal validity.

The experiment tried to represent the use of privacy requirements in real life. In reality, participants would have their own individual requirements. However, we had to preset the privacy requirements in order to measure the correctness as the discrepancy between the participants' results and the sample solution. Thus, we cannot be sure whether the same correctness values would be achieved in the real world with own privacy requirements. This poses a threat to external validity. The paradigm security levels in combination with the given tasks does most likely not reflect the reality since the preset tasks matched perfectly to one of the security levels. This is rarely the case in real life and therefore limits the external validity to some extent. However, we decided to propose a perfect solution, as the lack of the perfect match may have influenced the measured correctness and irritated the participants, which would have been a threat to internal validity. Furthermore, the experiment was conducted in a scenario that represents a single use case for privacy requirements (mono-operation bias). Further experiments that confirm our results in different scenarios would increase the generalization of the results and therefore the external validity. The number of participants per persona is quite small, especially the number of fundamentalists (three persons). In addition, a large

number of participants are academics. This does not reflect the overall population. Those aspects are threats to external validity.
The selection of the specification paradigms is based on our observations of the paradigms most commonly used in practice. We cannot rule out the possibility that there are other paradigms that could lead to better results in a comparable experiment. This implies a threat to conclusion validity with respect to our recommendations of best suitable specification paradigms. For the specification of privacy requirements the participants use concrete ‘tools’, which are implementations of the specification paradigms. This mixes findings on specification paradigms and corresponding tools. To minimize this threat to conclusion validity, usability experts supported us to make the ’tools’ as unobtrusive as possible. We discuss the generalizability of the experiment results in the following section.

5.4 Discussion

We wanted to investigate the relation between the selected persona and specification paradigm used in relation to objective correctness (RQ1) and self-estimation regarding perceived correctness (RQ2) with our research questions.

With respect to RQ1, we identified that all personas did least mistakes with the specification paradigm security levels. The number of mistakes related to decisions differs only marginally between the other paradigms. However, the persona marginally concerned differs significantly from the others with respect to objective correctness as they did more mistakes. The cumulated mistakes are higher than expected by the authors. This raises the question about the difficulty of the tasks to be solved. It was possible to solve all tasks without mistakes, because 10 out of 61 participants achieved the perfect objective correctness (zero mistakes in total). No one explained that he did not understand the tasks or the scenario in free text comments at the end of the experiment.

Regarding RQ2, we found that the perceived correctness is related to the number of decisions of a paradigm. More freedom led to worse perceived correctness in our experiment. However, there is no significant difference in how personas perform regarding perceived correctness in these paradigms. We did not expect that only few participants (8 out of 61) estimated perceived correctness rightly. Most of the others overestimated themselves and only four underestimated their correctness. Overestimation could in practice frustrate a user of privacy settings, as the system is not acting as expected. This could reduce trust in the privacy settings interface and its providing company. The participants underestimating their achieved correctness might appreciate the correct specification and the effect by the system, but they also might be frustrated because they have the feeling of not having control over the system.

Our experiment relies on the personas developed by Dupree [5]. We decided to go for these personas since they were developed based on empirical data. The personas mainly differ regarding motivation and security knowledge but also include more valuable information (e.g. valuing convenience more than security). Moreover, they contain concrete security behaviors such as use of strong passwords. We assume that such concrete information ease the self-classification compared to a scale with short statements, which are prone to a subjective interpretation (i.e. expert knowledge might be interpreted differently). Our two questions in the experiment about security knowledge and motivation had the purpose to control whether the persona selection is reasonable. However, we do not consider these to questions as sufficient to replace the personas. In practice, it would be preferable to have a small selection questionnaire for the user to persona mapping. However, to the best of our knowledge, that does not exist.

In the study by Dupree [5], the number of fundamentalists was the smallest by far, such as in our experiment. More fundamentalists are needed to draw conclusions about an appropriate specification paradigm. The other personas were represented by 11, 12, 14, and 21 participants, respectively. The numbers seem small as well but were enough to properly apply statistical analyses with the chosen tests. Nevertheless, the experiment need to be repeated with more participants to improve the generalizability of the results.

Many participants are academics or related to an academic work environment (69% academics, 54% employees of the authors' institution, 93% german-speaking participants). Obviously, the group of participants does not reflect the overall population (e.g., 15% academics in Germany). We cannot rule out that this had an influence on the results and a negative impact their generalizability. It seems unlikely to us that the level of education has a direct impact, but indirect effects seem reasonable. The level of education is related to certain jobs and interests and by this to knowledge about IT- security. More precise questions have to be asked in future to properly investigate the relation of education to correctness. Questions could be 'is your job related to IT-security or privacy?' and 'do you spend time in your spare time to learn about privacy?'

6 CONCLUSION AND FUTURE WORK

In this article, we have shown that appropriate specification interfaces can be assigned to users to promote the correct specification of privacy requirements and to give users confidence that they have made the right decisions. To this end, we have categorized the common types of specification interfaces used in practice as specification paradigms and have them used by different user types (personas) according to predefined tasks within a scenario. Through the results, we can recommend specification paradigm assignments to personas to achieve the highest possible objective and perceived correctness. In summary, we can clearly recommend the security levels for all personas. In addition, amateurs, lazy experts and technicians performed well with default policies. In case of necessity for fine-grained specifications, template instantiation and wizard can be effective enough for technicians. Due to the small number of fundamentalists, we cannot give recommendations for this persona.

The main focus of the overall experiment is to identify potential for increasing effectiveness, efficiency and satisfaction of privacy policy specification interfaces for users. This paper shows that effectiveness can be increased for personas by the selection of the right specification paradigm. We show in [20] that the specification paradigm also influences efficiency and satisfaction. In our results, effectiveness and efficiency of specification paradigms are aligned, satisfaction behaves contrary. People do not like "security levels" but perform efficiently and effectively with this paradigm. Vice versa, people like the paradigms “wizard” and “template instantiation”, but are more ineffective and inefficient with them. This poses a dilemma for the provider that needs to select the appropriate specification paradigm for the privacy specification interfaces of the own software product. High effectiveness and efficiency may be desired by users, however the low satisfaction with the paradigm may hinder users to specify privacy requirements at all. Contrary, a satisfying tool that leads to incorrect privacy settings may limit the trust in the provider. Besides that also other obligations might be fulfilled, such as legal requirements or the necessity of the provider to collect data due to the business model of the software product. Thus, with current results we cannot give generic recommendations for the specification paradigms selection. Providers must carefully balance pros and cons before selecting a paradigm based on the personas which best reflect the users.

To confirm our results, we need to perform non-exact replications of our experiment including a larger sample of participants from all user types and additional scenarios. We need to find out whether optimizations in the implementations of the paradigms can positively influence the objective and perceived correctness. Therefore, we also need to explore the use of additional paradigms and discuss the current look and feel as well as the interaction process of the used paradigms.

Acknowledgements

The research presented in this paper is supported by the German Ministry of Education and Research projects “Nationales Referenzprojekt für IT-Sicherheit in der Industrie 4.0 (IUNO)” (grant number 16KIS0328) and “Transparente und selbstbestimmte Ausgestaltung der Datennutzung im Unternehmen (TrUSD)” (grant number 16KIS0898). The sole responsibility for the content of this paper lies with the authors.

References

1. European Commission: Special Eurobarometer 431 - Data Protection. http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_431_en.pdf (2015)
2. Symantec: State of Privacy Report 2015. https://www.symantec.com/content/en/us/about/presskits/b-state-of-privacy-report-2015.pdf (2015)
3. Rudolph, M., Feth, D., Polst, S.: Why Users Ignore Privacy Policies. A Survey and Intention Model for Explaining User Privacy Behavior. In: 19th International Conference on Human-Computer Interaction (HCII), Las Vegas, USA (2018)
4. Kumaraguru, P., Cranor, L.: Privacy indexes: a survey of Westin's studies. http://repository.cmu.edu/isr/856 (2005)
5. Dupree, J.L., Devries, R., Berry, D.M., Lank, E.: Privacy personas: clustering users via attitudes and behaviors toward security practices. Conference on Human Factors in Computing Systems, (2016)
6. Digman, J.M.: Personality Structure: Emergence of the Five-Factor Model. Annual Review of Psychology (1990)
7. Keirsey, D.: Please understand me 2. Prometheus Nemesis Book Company (1998)

8. Myers, I.B., McCaulley, M.H., Most, R.: Manual: A guide to the development and use of the Myers-Briggs Type Indicator, vol. 1985. Consulting Psychologists Press Palo Alto, CA (1985)
9. Urban, J.M., Hoofnagle, C.J.: The Privacy Pragmatic as Privacy Vulnerable. In: Workshop on Privacy Personas and Segmentation. SOUPS, Menlo Park, CA, July 9-11 (2014)
10. Smith, H.J., Milberg, S.J., Burke, S.J.: Information privacy: measuring individuals' concerns about organizational practices. MIS Quarterly, 167–196 (1996)
11. Morton, A., Sasse, M.A.: Desperately seeking assurances: Segmenting users by their information-seeking preferences. In: Privacy, Security and Trust (PST), 2014 Twelfth Annual International Conference on, pp. 102–111 (2014)
12. Malhotra, N.K., Kim, S.S., Agarwal, J.: Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. Information systems research 15(4) (2004)
13. Information Systems Security Research Group, University of Kent: Permis. http://sec.cs.kent.ac.uk/permis/
14. Johnson, M., Karat, J., Karat, C.-M., Grueneberg, K.: Usable Policy Template Authoring for Iterative Policy Refinement. In: IEEE International Symposium on Policies for Distributed Systems and Networks, POLICY, Fairfax, Virginia, USA (2010)
15. Uszok, A., Bradshaw, J., Jeffers, R., Suri, N., Hayes, P., Breedy, M., Bunch, L., Johnson, M., Kulkarni, S., Lott, J.: KAoS policy and domain services: Toward a description-logic approach to policy representation, deconfliction, and enforcement. In: IEEE 4th International Workshop on Policies for Distributed Systems and Networks, POLICY. (2003)
16. Liu, Y., Gummadi, K.P., Krishnamurthy, B., Mislove, A.: Analyzing Facebook privacy settings: User expectations vs. reality. ACM conference on Internet measurement, (2011)
17. Fang, L., LeFevre, K.: Privacy Wizards for Social Networking Sites. In: Proceedings of the 19th International Conference on World Wide Web. ACM, New York, NY, USA (2010)
18. Cranor, L.F.: P3P: making privacy policies more useful. IEEE Security Privacy (2003)
19. Cranor, L., Langheinrich, M., Marchiori, M.: A P3P Preference Exchange Language 1.0 (APPEL1.0). https://www.w3.org/TR/P3P-preferences/ (2002)
20. Rudolph, M., Polst, S.: Satisfying and Efficient Privacy Settings. Mensch und Computer (2018)
21. Rudolph, M., Feth, D., Doerr, J., Spilker, J.: Requirements Elicitation and Derivation of Security Policy Templates—An Industrial Case Study. In: 24th International Requirements Engineering Conference (RE), Beijing, China, pp. 283–292 (2016)
22. Supplementary Experiment Material including extended Figures for this Paper: http://s.fhg.de/yU6