
Proceedings Paper:
Petrie, Helen orcid.org/0000-0002-0100-9846, Sreekumar, Gayathri and Shahandashti, Siamak F. orcid.org/0000-0002-5284-6847 (Accepted: 2024) Understanding users' mental models of Federated Identity Management (FIM): use of a new tangible elicitation method. In: IFIP International Symposium on Human Aspects of Information Security & Assurance (HAISA 2024). International Symposium on Human Aspects of Information Security & Assurance, 09-11 Jul 2024 , SWE (In Press)

Understanding users' mental models of Federated Identity Management (FIM): use of a new tangible elicitation method

Helen Petrie [1[0000-0002-0100-9846]] Gayathri Sreekumar [1[0009-0002-9888-722X]] and Siamak Shahandashti [1[0000-0002-5284-6847]]

[1] University of York, York YO10 5GH, United Kingdom
helen.petrie@york.ac.uk, gayathrisreekumar.mec@gmail.com, siamak.shahandashti@yor.ac.uk

Abstract. The number of passwords users require to interact with online accounts continues to grow, as the services they interact with online become more and more common. Federated Identity Management (FIM) offer an easy option for users to authenticate themselves to many accounts using just one password from an Identity Provider such as Facebook or Google. Previous research has shown that users are reluctant to use such systems and have inaccurate mental models of how they work, but much of the research is now over a decade old. An initial exploratory study with 12 users asked them to create a mental model of a particular concrete FIM scenario, using a new tangible elicitation method involving felt icons and a flocked board, based on the Fuzzy-Felt toy for young children. It was found that almost all participants had inaccurate mental models of FIM which may lead to hesitancy to use such systems: they believe much more information is passed to the website they wish to login to and they mis-understand the route taken by the information that is passed between their browser, the Identity Provider and the target website. The implications of these results and the new method of eliciting mental models are discussed.

Keywords: Federated Identity Management (FIM), Mental Models, Fuzzy-Felt Method.

1 Introduction

The number of passwords users require to interact with online accounts continues to grow, as the services they interact with online become more common. Often the information stored in these accounts is personal and valuable, be it financial information or personal photos, so ensuring the information is secure is important. Yet research has repeatedly shown that users tend to create weak passwords or re-use the same or similar passwords on different accounts. Analysis by NordPass [11] found that the most commonly used password in 2022 was "password", used by nearly 5 million users, with "123456" in second place with over 1.5 million users.

Federated Identity Management (FIM)[1] is an attractive alternative to remembering many different passwords and many websites now offer this process, often using services provided by Google or Facebook. However, research has suggested that users are sometimes hesitant about using FIM or that they use it but have concerns about the security of the process and this may be because they have inaccurate mental models of how it works and how secure it is. One might legitimately ask, why does it matter whether users have accurate mental models of the processes involved in FIM? It is not necessary to understand the inner workings of complex systems to be able to use them effectively. However, if inaccurate mental models are either leading users to take unnecessary risks online or conversely to not use convenient and usable systems or not use them appropriately, then that is affecting the adoption and safe use of such systems. Users' inaccurate mental models of FIM may be an interesting case in point: if they do not realise how secure the process is, they may not use it and resort to less safe methods of authentication.

Our previous research using an online survey [13] found that while the majority of British participants reported using FIM at least some of the time but both users and non-users appeared to have inaccurate mental models of the security of the process. Therefore, in this study we used a different method to assess participants' mental models of FIM, by asking them to create a tangible diagram of a specific FIM scenario. This will allow us to gather more data about users' understanding of FIM, but also compare the assessment of their understanding using different methods.

The rest of this paper sets out an overview of previous research of users' attitudes and understanding of FIM, then presents the methodology for our initial exploratory study and in particular the method for creating the tangible diagrams of what happens in a FIM process, followed by the results and conclusions of the work.

2 Background

Federated Identity Management (FIM) processes allow users to authenticate themselves to a range of systems through an Identity Provider (IdP), rather than having to authentication separately for each system. FIM grew out of the Lightweight Directory Access Protocol (LDAP) developed in the early 1993 [26], but only became widely used with the growth of many online services in the late 2000s. For example, Facebook launched its first FIM system in 2008 [1]. Consequently, there was an initial burst of studies of the acceptability and usability of the process for users in the early 2010s, and a steady if small number of studies since then. Early unpublished usability research was conducted by Yahoo! and Google on their implementations of FIM processes and found usability problems with each of them (Freeman, 2008, Sachs, 2008, reports no longer accessible, see [25]). The first published research was by Sun et al. [23, 24] who conducted a small laboratory study with nine participants. They found that participants'

[1] A more restricted form of FIM is Single Sign-On (SSO) when used within same entity or domain (e.g. within Google). SSO as a term which is more familiar to the general public and is often used instead of FIM, so we referred to the process as SSO with our participants. Social login is another term used in the literature, but it is not widely used by the public in the UK. In this paper for simplicity we will use FIM to refer to both FIM and SSO.

existing password management strategies reduced the perceived usefulness of FIM, that some participants expressed concerns about "single point of failure" risks, most held incorrect beliefs that their credentials were being passed to the target website, and many were reluctant to use FIM for services that contained valuable personal information. In follow-up research, Sun et al. [25] designed a prototype FIM user interface to address some of these issues. However, an evaluation with 35 participants found that one third of participants would still opt to use separate passwords for each system and a further third would decide depending on the value of the information being held and the trustworthiness of the target website.

A number of other laboratory studies have been conducted [1, 4, 6, 17]. Engelman [6] presented three variations of Facebook Connect (the Facebook FIM in 2013), the original presentation and two variations with the same information but different layouts. 87 participants used one of these to login to three different websites. Over 75% of participants proceeded to log in with Facebook Connect, with the layout of information and website not having a significant effect. However, nearly a third of participants had an inaccurate mental model of what information was transferred to the target website. It may be that the high rate of using Facebook Connect in this study was due to the participants trust in the researchers and the "demand characteristics" of the laboratory situation [12], problems which have been noted in several studies of usable security [19 - 21].

Arianezhad et al. [1] were particularly interested in differences how users' security expertise affects their attention to security indicators in FIM and their willingness to use it. They measured eye-tracking behaviour as well as interviewing participants. Their small study (9 security expert participants, 9 novices) found that the experts spent longer studying the security indicators than the novices, but neither experts nor novices had a good understanding of FIM. Brostoff et al. [4] investigated the acceptance of FIM for online government services (then under consideration) in the UK. In a first study in which prototype FIM information complied with UK government guidance, 50% of participants said they would not login with the FIM and 34% of participants felt threatened rather than reassured by the privacy statement. With a redesigned interface, only 20% said they would not login. Ruoti et al [17] compared seven different available web authentication processes. They found that transparency of information increased usability, but also led to confusion and a decrease in trust. Most interestingly, they found that participants preferred FIM to other authentication processes, but also wanted to augment it with site specific passwords.

Saint-Louis and McEwen [18] studied participants' mental models of FIM by providing them with magnetized buttons for logos of a range of entities and asking them to connect them with lines drawn on a magnetized white board. They presented a display of 30 buttons in their paper, but participants were asked to undertake 15 FIM related tasks and create representations of their mental model of each, so presumably all the buttons were required at some point. This certainly increased the complexity of the task for participants. These researchers do not report any results on the accuracy of participants' mental models, as they are more interested in their method of eliciting the mental model.

A number of other researchers have conducted online studies about use and acceptance of FIM. Bauer et al. [3] investigated Google, Facebook and Google+ FIMs and found that participants' understanding of the information shared was based largely on their preconceptions and not the actual information provided in the FIM dialogs (this study was conducted with MTurk participants and raised doubts about how much attention they were paying, compared to what they would do in their real use of FIM processes). Participants also expressed a strong desire to be informed about what information is being shared. Gafni et al. [7] conducted an online survey of users and non-users of FIM processes in Israel. They found that while familiarity with and convenience of the process predicted FIM use, ease of use had no impact, and not surprisingly privacy and security concerns predicted non-use. Jiang [8] also conducted an online survey in China investigating individual characteristics which predict FIM use. She found that men were more likely to use FIM than women, older people and those with more privacy knowledge and self-protection skills are less likely to use them than others. Most recently, Cho et al. [5] conducted an online study which asked participants to login to four mock-ups of apps which had differing levels of valuable information using either an FIM (Facebook, Google or Twitter), with their email, or by setting up a new account. The study produced rather surprising results: for the low value app (a class reunion), participants were equally split between FIM and their email account, although these were more popular than setting up a new account; but for the high value apps (a serious matchmaking app and a "hook up" app for casual liaisons), again Facebook or email were preferred over a new account, but for the "hook up" app participants preferred Facebook to their email, whereas for the matchmaking app they preferred their email. It might have been expected that for the "hook up", users would want the highest level of security, but it may be that in this instance immediate privacy (not revealing one's email address) was more important than overall security. The authors conclude that users have different layers of sensitivity, so decisions about using a particular authentication mechanism are quite nuanced.

Stobert and Biddle [22] conducted interview studies with both novice and experts in computer security, showing them a number of screenshots of login screens. One of these screenshots offered a choice of logging in with Facebook or with username and password. Of the 27 novices interviewed, only two mentioned the Facebook option, one participant said she would use it, because she had difficulty remembering even her reused passwords; the other participant said she would not use it because she did not want any extra information cluttering her Facebook page (which shows an interesting mis-understanding of how FIM works). None of the other participants commented on the Facebook option and they were not prompted to do so, as the purpose of the interviews was to investigate people's use of passwords rather than FIM. In the 15 interviews with experts, FIM did not appear to get mentioned at all.

Balash et al [2] conducted two large online surveys and found that more than half the respondents used FIM but many expressed concerns about the process giving access to personal information such as email addresses. Pratama et al. [14] also used a large online survey to investigate individual differences in attitudes to FIM, finding that there was an age difference, with older participants less aware of the security of FIM than younger participants, but that there was no difference between men and women.

Morkonda et al. [10] in another large online survey found that 55% of participants would opt to use FIM for authentication, with those who would not choose FIM most frequently citing privacy concerns. Interestingly after being provided with more information about what information would be shared by the process, 28% of participants said they would change their option, but 11.5% said they would change to a different FIM option (e.g. Google to Apple), 9.0% said they would change from a non-FIM option to an FIM one and 7.5% opted out of the FIM options. Thus, providing more information did not uniformly make participants report they would be more likely to use FIM. In our online survey [13] with 98 participants in the UK, 75% use FIM but both users and non-users rate it as moderately high risk and have numerous security concerns about it.

The availability of FIM has grown considerably in the past few years, so we would expect users to be much more familiar with the concept and perhaps have developed a better understanding of the process and more confidence in it. Yet research continues to show that only 50% to 75% of users are taking up the option [4, 6, 10, 13], even though it is generally more secure than having individual passwords for different accounts (if these are not strong). Therefore, it is important to understand why a substantial percentage of potential users decide not to use it. Although asking people in surveys provides some insight into their mental models of the processes involved, we investigated whether a more accurate understanding could be developed by asking them to create a tangible representation of the process. In addition, this method of creating a tangible representation of users' mental models might be useful in investigating other issues in cybersecurity research.

3 Method

3.1 Design

Studying people's mental models is not easy, as this information may not be stored verbally but in some more abstract, relational form. Thus, although researchers often use questionnaires and interviews to elicit mental models, drawing techniques are often used as well. However, our experience is that asking research participants to draw also has a number of problems. Participants may be unconfident about their drawing skills and embarrassed to draw for a stranger (i.e. the researcher), and trying to make a visual representation of unfamiliar concepts such as servers, the cloud and websites may make them even more unconfident and stressed. Both these factors may mean participants put considerable cognitive effort into the process of drawing rather than thinking about their mental model.

Therefore, we propose a simpler and hopefully more enjoyable method for eliciting participants' mental models, which involves the use of small felt icons and a flocked board (see Figs. 1, 2) and combines the use of these materials with a concurrent think aloud protocol [16], a very well-established method in human-computer interaction and other fields to elicit information from participants about what they are doing and thinking. These materials are based on a young children's toy known as Fuzzy-Felt[2]. Thus, one contribution of this research is both results from a small study about current users' mental models of FIM, but also an initial trial of the Fuzzy-Felt method. We believe this method will be useful for researchers in studying users' mental models of a range of topics in usable security and beyond, as well as FIM. We discovered (after we had conducted this study) that another group of researchers [15] had also used Fuzzy-Felt icons and a flocked board. They were interested in where people located domestic appliances in their home and whether they programmed them and how often, so a very different, but interesting purpose, to that of the current study. They were not interested in people's mental models, but the Fuzz-Felt was a useful way of grounding the discussion of different appliances in the home.

Thus in this initial exploratory study participants were asked to create a visual representation of their mental model of what they thought happened during a particular single sign-on (FIM) event, described to them in a non-technical scenario (see section 2.3, below), To create the representation they were given a set of tangible materials, a Fuzzy-Felt board and a range of felt icons representing all the entities that might be involved in FIM, including some distractor entities (i.e. digital entities which are not involved in FIM, but which might plausibly be involved), and some blank felt icons and a felt-tip pen with which to create further icons if they wished. Participants were also provided with WikiSticks[3] to represent connections in their representation (see Figs. 1, 2). While participants created their representation, they were asked to "think aloud" what they were doing and what their mental model was, to ensure that the researcher understood what they were creating. After they had created the representation, participants were asked how confident they were in their understanding of the FIM process and what information they thought FIM systems stored. As part of the debriefing, they were shown what actually happens in a FIM event such as the one in the scenario.

3.2 Participants

An opportunistic sample was used, with the only selection criteria being that participants should be 18 years or older and have seen or used an FIM login before. 12 participants took part in the study, 6 women and 6 men, with a mean age of 26.7 years (range: 22 to 46 years). All participants were educated to master's degree or above. Four participants were professionals working at the University of York, the other eight were students. The students were studying a wide range of subjects including English,

[2] Fuzzy Felt was developed in the UK as an educational toy for young children, it was first sold in 1950. The toy consists of a flocked board onto which a number of felt shapes are placed to create different pictures (https://en.wikipedia.org/wiki/Fuzzy-Felt).

[3] WikkiStix (wikki-stix.co.uk) are sticky waxed sticks 12 cm long which adhere to the flocked board.

psychology and computer science, but none were specializing in computer or online security. Three participants (one of the professionals and two of the students) had some previous experience working in positions related to online security. The student participants were offered an Amazon gift voucher worth GBP10 (approximately USD12.50) for their time. The working professionals were offered coffee and biscuits.

Participants were asked to rate their expertise with computers and the internet on a rating scale from "not at all expert" (scored as 1) to "very expert" (scored as 7). The median rating was 4.5 (range: 1 to 7), which was not significantly different from the midpoint of the rating scale (one sample Wilcoxon ranked signs test T = 0.63, n.s.), so the participants thought they were "averagely" expert in this area. Participants were also asked to rate their expertise with online security issues on the same scale. The median rating was 4.0 (range: 1 - 5), again not significantly different from "averagely" expert (T = -1.19, n.s.). Participants were asked about what education they had about online security issues: nine were considered "self-taught", mentioning online reading and research and word of mouth; of the other three, one had taken courses at university and the other two had learnt in working in areas related to security.

3.3 Materials and Equipment

A standard Fuzzy-Felt flocked board was used to represent cyberspace (see Figs. 1, 2). 14 different felt icons (see Table 1) were created to represent entities of cyberspace: different coloured felt was cut into 1-inch squares and labelled with the initials of each entity (see Fig. 1). The range of icons created was based on our previous survey which asked respondents an open-ended question about what happens in FIM [13]. 12 of the icons would involved in an FIM process, but 2 were "distractor" icons, entities which are not involved in FIM processes, but which people often think are involved. A display of all the entities was created with the description and a reference icon (which were left on the display even if the participant used all three available icons; if they were all used, the researcher would create some more on the fly, but this never happened). This meant participants always had access to the meaning of the icons while creating their representation. Participants were given a pile of extra blank felt pieces and a felt tip pen in case they wanted to add an entity which was not present in the display. WikkiStiks were provided to represent connections between entities in the representation.

A mobile phone mounted on a tripod was used to record participants' creation of their representation, along with their think aloud protocol and any prompts from the researcher.

Fig. 1. Set up at the start of the study. From the left, WikkiStixs and extra unlabelled felt icons; the 14 icons of entities which could be used in making a representation, with short labels on the icon and longer explanations next to them; the Fuzzy-Felt flocked board with two items, one of a sender of a message and one of a receiver (only one was needed for the main sessions, this arrangement was for the demonstration by the researcher, see section 2.3); smartphone on tripod set up to record the session.

A recruitment questionnaire collected demographic information from participants and asked them to rate themselves on their expertise with computers and the internet, and on their expertise about online security issues (both on scales from "not at all expert" scored as 1 to "very expert" scored as 7). Participants were asked where they learnt about online security issues.

The researcher used an example scenario to illustrate how to create a representation using the materials and provide a "think aloud" while doing so. This scenario was not related to FIM, but involved sending an email from her Gmail account to her friend's Apple account. This was to ensure participants understood how to use the materials and in particular how to provide the "think aloud" while doing so.

The main scenario was read to the participants and placed on the table so they could refer to it at any time while making their representation:

> Imagine you are interested in travelling to Edinburgh to attend the Edinburgh Fringe Festival. When you try to book your train tickets in trainline it requires you to create an account for buying the tickets. Since you have little time, you decide to go ahead and sign in with your Google account (note that you are not already signed into your Google account). Imagining the Fuzzy-Felt board as the cyberspace, can you use the different felt pieces provided to map out what you think happens when you sign in with Google to the Trainline website?

After creating the visual representation, a short post-study questionnaire asked participants if they thought FIM systems stored information about them and if so, what information was stored and where.

The third author created an accurate visual representation (using the Fuzzy-Felt materials) and think aloud protocol of what actually happens in the FIM scenario, which was used to explain the process to participants at the end of their session and to provide

a reference representation and information against which to code their representations and think aloud protocols for accuracy.

2.3 Procedure

Ethical approval was given by the University of York Physical Sciences Ethics Committee. The researcher (either the first or second author) started by giving an overview of the study, allowed the participant to ask any questions and asking them to sign the informed consent form. Then the researcher explained the materials to be used and demonstrated an example of using the materials to create a visual representation and provided a think aloud protocol while doing so. The participant was again asked whether they had any questions.

The participant then read the scenario and a printed copy was available throughout the study for them to refer to if they wished. The researcher started the representation for the participant by placing computer and browser icons next to the person icon (male and female icons were used as appropriate, see Fig. 2) and prompted them to start creating their representation and thinking aloud. The researcher also prompted the participant as needed if they were not thinking aloud or if the researcher thought something was not sufficiently clear.

When the participant was satisfied with their representation, they completed post-study the questionnaire. Then the researcher explained the actual FIM process involved in the scenario, answered any questions, thanked the participant for their time and for the students, arranged for the gift voucher. The study took about 30 to 45 minutes for each participant.

2.4 Data Analysis

Participants' think aloud protocols and any interventions by the researcher were transcribed for analysis in conjunction with the Fuzzy-Felt visual representations.

In order to understand the accuracy of participants' representations and think aloud protocols, the first author used a 7-point Likert item to code their accuracy in comparison with the reference representation and information (see Table 1). The second author independently scored all the representations and think aloud protocols and any differences between scores were discussed and resolved. The representations and think aloud protocols were also analysed to identify which misconceptions participants had.

Table 1. Coding of level of accuracy for the representation and think aloud protocol information.

Accuracy level	Explanation
1	No understanding at all, got it completely wrong
2	Not a very clear understanding, 6 or more misconceptions
3	A slight understanding, 5 misconceptions
4	Some understanding, 4 misconceptions
5	Good understanding, 3 misconceptions
6	Fairly good understanding, two misconceptions
7	Very good understanding, a single misconception allowed

3. Results

Figure 2 shows the mental models created by two participants, neither accurate. In terms of the accuracy, only one participant produced a completely accurate representation (with a score of 7); 4 of the 12 participants (25%) were able to create a representation that was close to the actual working of the FIM (scored 5 or 6); and 5 participants (41.7%) had only a slight idea of how FIM authentication works or less (scored 1 to 3).

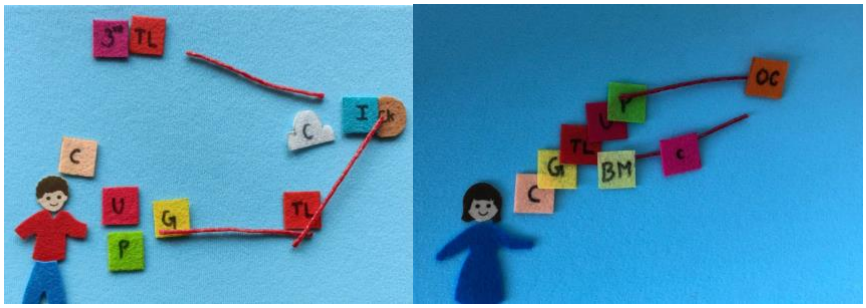

Fig. 2. Fuzzy Felt representations of FIM by two participants.

Table 2 summarizes the misconceptions made by participants in creating their representations and the frequency of each misconception, which reflect inaccurate mental models. Participants often made more than one misconception, so the numbers add up to more than 12 (the number of participants).

The most common misconception (made by 10, 83.3% of participants) was not realizing the full data flow path aFIMciated with FIM login. Participants had a simpler mental model in which the data flow in FIM login is one way (see representation on the left in Fig. 2). However, the actual FIM process is more complex, the user inputs their credentials (username and password) to the Google server and the server sends a code and the user's username only to the user's browser. The browser then sends a unique code and the username to the Trainline website. This may well be a reason for hesitancy in using FIM. However, if users realized that Google sends a code back to the browser

and then the browser sends the code and username to the Trainline, they might have more confidence in the process.

Table 2. Misconceptions made by participants and their frequency.

Misconception	**N % of participants reporting**
Did not realise data travelled back to the user's browser	10 (83.3%)
Thought Google shared user's credentials (username and password) with the Trainline website directly	7 (58.3)
Thought credentials were directly sent to the Trainline website rather via a Google server and the user's browser	5 (41.6)
Unsure of the role of Google in FIM login	6 (50.0)
Thought all information associated with Google account is stored in user's browser memory and cookies	3 (25.0)
Thought Trainline has access to all the information in the user's Google account	3 (25.0)
Thought all information associated with user's Google account is stored in the cookies of the Trainline website	2 (16.6)
Thought there was a connection established between the Trainline website and user's browser/computer memory	2 (16.6)
Thought there was a connection established between the user's Google username and password and the Trainline website	2 (16.6)
Considered two factor authentication as part of SSO login process	1 (8.3)

A number of the other misconceptions involved providing the user's password and other personal information to the target website. This accounted for over half of the misconceptions reported. Overall, only one participant had a clear understanding of what information was passed to the target website and one participant had some understanding, but contracted themselves during their think aloud protocol. This is undoubtedly a major reason in hesitancy in using FIM processes.

Participants were also confused about the role of cookies in FIM. This may be a case of a "little knowledge is a bad thing". Cookie banners try to provide information, whereas typically FIM pages do not provide any information about what information will be passed between the IdP and the target website. Participants may be trying tothe information they have learnt about cookies to FIM. One participant also became confused between two factor authentication and FIM, and described a novel system combining the two, which involved the user receiving a token on a different device and entering it into the target website.

Although the sample size was small, we did investigate whether there was a relationship between participants' self-rated level of expertise with computers and the internet or with online security and the accuracy of their mental model. It should be noted that these two self-report measures did correlate significantly with each other (Spearman rank correlation $r_s = 0.66$, $p = 0.019$), so participants who said they were more expert with computers and the internet were also more likely to say they were expert about online security. However, there was no correlation between the participants' self-ratings of online security expertise and the researchers' ratings of the accuracy of their representations of FIM ($r_s = 0.37$, $p = 0.24$). There is a suggestion of

a correlation between the participants' self-ratings of their computer and internet expertise and the researchers' ratings of the accuracy of their representations of FIM (r_s = 0.49, p = 0.09). This relationship is plotted in Fig. 3, note that there are only 10 points, as 2 participants gave scores of 3 for their self-rating and received 3 for their representation, and two gave 5 for their self-rating and received 6 for their representation. So, for 10 out of the 12 participants there is a very good positive relationship between self-rating of computer/internet expertise and rating of their representation. But two participants give themselves high self-ratings of their computer/internet expertise, but only received ratings of 2 (not a very clear understanding, 6 or more misconceptions) for their representation. Examining the data for these two participants, one cannot pick out a particular reason why they should be so different from the rest of the sample. Of course, it is completely reasonable for someone to have very good computer/internet expertise and not understand how FIM works, but it is intriguing that for over 80% of the sample, there is a very good relationship. This warrants further investigation, but clearly needs a larger sample of participants, and more nuanced measures of their computer/internet and online security expertise. These do exist, but we did not want to extend the length of the study to include them.

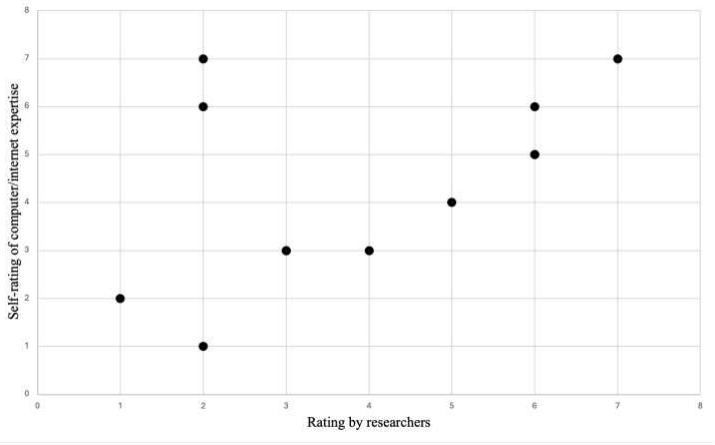

Fig. 3. Relationship between participants' self-rating of computer/internet expertise and researchers' ratings of the accuracy of their representation of FIM.

4. Discussion and Conclusions

The aim of this study was to further understand people's mental models of FIM processes, in terms of how accurate they are and whether they would lead people to be overly cautious in terms of their decisions of whether to use FIM. It was also an initial exploration of the use of the Fuzzy-Felt method for eliciting people's mental models of FIM processes, which might also be useful for studying people's mental models of other

aspects of online security, and indeed their mental models of other aspects of the online world.

Although the sample size was very small, the results are clear that almost all participants had a poor mental model of FIM, with many misconceptions. However, in examining participants' mental models we revealed two reasons why users' may be hesitant to use FIM, one of which has been widely reported in the research literature, but the other has not. We would argue that this is because we asked participants to create a visual representation of their mental model, which only two previous studies on use and understanding of FIM have done [18, 25]. The first reason is that almost all participants did not understand how little information is shared with the target website, many thinking that their password is shared with the IdP, or indeed all their personal information. This confirms previous studies, particularly those by Sun et al. [23, 24] and Brostoff et al. [4] which suggest that a decade later, users still do not have a basic understanding of FIM, which may lead to hesitation to use it.

But in addition, we found that users did not understand that the data flow in FIM and believed that data flowed directly from the IdP to the target website, without the user's browser involved. There is a hint of this misconception in the research by Sun et al [25] (the figure with a participant's incorrect mental model drawing shows exactly this misconception), but it is not discussed in the paper, which only considers whether users' mental models were correct or not. Clearly, the link between this misconception and hesitancy about FIM needs further investigation. However, on the basis of both these findings, we suggest that IdPs should provide more clear information to users about what information is passed to the target website and what precautions are taken. Interestingly, this advice is at odds with the findings of Ruoti et al [17], who found that greater transparency of information led to confusion and a decrease in trust. On the other hand, Karegar et al. [9] created a tutorial and a user interface to inform users about the privacy issues of FIM. They found that both approaches improved participants' knowledge about FIMs. Whether this would lead to greater use in real life in the future remains an open question.

The results on the relationship between computer and internet expertise and accuracy of representations and hence mental models, an interesting possibility, but requires further research with a larger sample and better measures.

On the use of the Fuzzy-Felt method to elicit participants' mental models of FIM, we did not conduct a formal evaluation of the method, as this was a first exploratory study. However, many of the participants commented that it was fun and easy to use the materials. As researchers, although it involved more time and effort to set up than simply asking participants to create a drawing of their mental model, we found it a satisfying way to work with participants. Combined with the use of the participants' think aloud protocol, we believe it is an effective way to elicit mental models. We are already planning a further study, of a different aspect of online security, in which we will compare simply drawing a mental model with creating it use Fuzzy-Felt (both with think aloud), to investigate whether the Fuzzy-Felt method is more effective, more efficient (for participants) and more enjoyable than drawing.

The study has a number of limitations, which should be noted. As already mentioned, the sample size was very small, but in addition it was very heterogeneous,

with some students and some staff. We had originally aimed for a sample of all students of a limited age range, but participant recruitment was difficult, so a wider net was cast. In some sense it yielded a more robust sample, but a more homogenous sample might have yielded clearer findings particularly on the relationship between self-rated computer/internet and security expertise and accuracy of mental model. In addition, participants only created their mental model of one concrete FIM scenario. There may have been something about this particular scenario which influenced the results. Using a range of scenarios would have been better, but we did not want to extend the length of the study for our participants.

In conclusion, this study has shown that users still have inaccurate mental models of what happens in FIM, in two particular ways, what information is passed to the target website and how it is passed. These may both contribute to a reluctance to use FIM processes. In addition, we found the Fuzzy-Felt method to elicit the participants' mental models fun and effective, if not as efficient for researchers as drawing and plan to provide more evidence of this in the future.

Acknowledgments. We would like to thank all the participants who gave their time to help with this research.

Disclosure of Interests. The authors declare that they have no competing interests.

References

1. Arianezhad, M., Camp, L. J., Kelley, T., Stebila, D. Comparative eye tracking of experts and novices in web single sign-on. Proceedings of the third ACM conference on Data and application security and privacy, San Antonio, Texas, USA. (2013).
2. Balash, D.G., Wu, X., Grant, M., Reyes, I., & Aviv, A.J. Security and privacy perceptions of third-party application access for Google accounts. Proceedings of the 31st USENIX Security Symposium (2022).
3. Bauer, L., Bravo-Lillo, C., Fragkaki, E., Melicher, W. A comparison of users' perceptions of and willingness to use Google, Facebook, and Google+ single-sign-on functionality Proceedings of the 2013 ACM workshop on Digital identity management. (2013).
4. Brostoff, S., Jennett, C., Malheiros, M., & Sasse, M. A. Federated identity to access e-government services: are citizens ready for this? Proceedings of the 2013 ACM workshop on Digital identity management. (2013).
5. Cho, E., Kim, J., Sundar, S. S. Will You Log into Tinder using your Facebook Account? Adoption of Single Sign-On for Privacy-Sensitive Apps Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (2020).
6. Egelman, S. My profile is my password, verify me! the privacy/convenience tradeoff of Facebook connect SIGCHI Conference on Human Factors in Computing Systems.. (2013).
7. Gafni, R., Nissim, D. To social login or not login? Exploring factors affecting the decision. Informing Science and Information Technology, 11, 57 – 72 (2014).
8. Jiang, J. Social login acceptance: a DIF study of differential factors. 22nd Pacific Asia Conference on Information Systems. Association for Information Systems (AIS). (2018). https://aisel.aisnet.org/pacis2018/20 [accessed 6 April 2024]

9. Karegar, F., Gerber, N., Volkamer, M., Fischer-Hübner, S. Helping john to make informed decisions on using social login. Proceedings of the 33rd Annual ACM Symposium on Applied Computing, Pau, France. (2018).
10. Morkonda, S.G., Chiasson, S., & van Oorschot, P.C. Influences of displaying permission-related information on web single sign-on login decisions. Computers & Security, 139, 103666 (2024)
11. NordPass Security. Top 200 most common passwords. https://nordpass.com/most-common-passwords-list/ (2024). [accessed 6 April 2024]
12. Orne, M.T. On the social psychology of the psychological experiment: with particular reference to the demand characteristics and their implications. American Psychologist, 17, 776-783, (1962).
13. Petrie, H., Sreekumar, G. Passwords and single sign-on: use, security, and understanding for online accounts. Proceedings of 37th International British HCI Conference (in press).
14. Pratama, A.R., Firmansyah, F.M., Rahma, F. Security awareness of single sign-on account in the academic community: the roles of demographics, privacy concerns, and Big-Five personality. PeerJ Computer Science, 8:e918 http://doi.org/10.7717/peerj-cs.918 (2022).
15. Rode, J.A., Toye, E.F., Blackwell, A.F. The fuzzy felt ethnography—understanding the programming patterns of domestic appliances. *Personal and Ubiquitous Computing,* 8, 161–176 (2004).
16. Rogers, Y., Sharp, H., Preece, J. Interaction design: Beyond human-computer interaction (6th edition). Wiley. (2023).
17. Ruoti, S., Roberts, B., Seamons, K. Authentication Melee: A Usability Analysis of Seven Web Authentication Systems Proceedings of the 24th International Conference on World Wide Web, Florence, Italy. (2015).
18. Saint-Louis, H., McEwen, R. Diagrammatic mental representation: a methodological bridge. Visual Studies, 37:5, 664-680, (2022).
19. Schechter, S. E., Dhamija, R., Ozment, A., Fischer, I. The Emperor's New Security Indicators. Proceedings of the 2007 IEEE Symposium on Security and Privacy. (2007).
20. Sotirakopoulos, A., Hawkey, K., Beznosov, K. "I did it because I trusted you": Challenges with the study environment biasing participant behaviors. Paper presented at SOUPS Usable Security Experiment Reports (USER) Workshop. (2010). http://lersse-dl.ece.ubc.ca/record/238/files/238.pdf [accessed 6 April 2024]
21. Sotirakopoulos, A., Hawkey, K., Beznosov, K. On the challenges in usable security lab studies: lessons learned from replicating a study on SSL warnings. Proceedings of the Seventh Symposium on Usable Privacy and Security, Pittsburgh, Pennsylvania. (2011).
22. Stobert, E., Biddle, R. The Password Life Cycle. ACM Transactions on Privacy and Security, 21(3), Article 13 (2018).
23. Sun, S.-T., Pospisil, E., Muslukhov, I., Dindar, N., Hawkey, K., Beznosov, K. OpenID-enabled browser: towards usable and secure web single sign-on. CHI '11 Extended Abstracts on Human Factors in Computing Systems, Vancouver, BC, Canada. (2011).
24. Sun, S.-T., Pospisil, E., Muslukhov, I., Dindar, N., Hawkey, K., Beznosov, K. What makes users refuse web single sign-on? an empirical investigation of OpenID. Proceedings of the Seventh Symposium on Usable Privacy and Security, Pittsburgh, Pennsylvania. (2011).
25. Sun, S.-T., Pospisil, E., Muslukhov, I., Dindar, N., Hawkey, K., & Beznosov, K. Investigating Users' Perspectives of Web Single Sign-On: Conceptual Gaps and Acceptance Model. ACM Transactions on Internet Technologies, 13(1), Article 2. (2013).
26. Yeong, W., Howes, T., Kille, S. X.500 Lightweight Directory Access Protocol. July 1993. https://www.rfc-editor.org/rfc/rfc1487