

Are Users More Willing to Use Formally Verified Password Managers?

Carolina Carreira[1,2(✉)], João F. Ferreira[3], Alexandra Mendes[4], and Nicolas Christin[1]

[1] Carnegie Mellon University, Pittsburgh, USA
[2] INESC-ID and IST, University of Lisbon, Lisbon, Portugal
carolina.carreira@tecnico.ulisboa.pt
[3] INESC-ID and Faculty of Engineering, University of Porto, Porto, Portugal
joao@joaoff.com
[4] INESC TEC, Faculty of Engineering, University of Porto, Porto, Portugal
alexandra@archimendes.com

Abstract. Formal verification has recently been increasingly used to prove the correctness and security of many applications. It is attractive because it can prove the absence of errors with the same certainty as mathematicians proving theorems. However, while most security experts recognize the value of formal verification, the views of non-technical users on this topic are unknown. We designed and implemented two experiments to address this issue to understand how formal verification impacts users. Our approach started with a formative study involving 15 participants, followed by the main quantitative study with 200 individuals. We focus on the application domain of Password Managers (PMs) since it has been documented that the lack of trust in PMs might lead to lower adoption. Moreover, recent efforts have focused on formally verifying (parts of) PMs. We conclude that formal verification is seen as desirable by users and identify three actionable recommendations to improve formal verification communication efforts.

Keywords: Password Managers · Usable Security · Formal Methods · Formal Verification · User Study · HCI · User Trust · Security Adoption · Verified Software · Software Assurance · Human Factors in Security · Cryptography Usability · Secure Software Adoption

1 Introduction

Formal verification has developed substantially in recent years and has been applied to many domains. Recent work in this area includes the formal verification of security properties in trusted execution environments [19], critical parts of the Linux kernel [28], the open-source TLS implementation used in numerous Amazon services [7], and even compilers [22,24]. Formal verification works by rigorously proving the correctness of code against a formal specification and thus

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
D. Bianculli and E. Gómez-Martínez (Eds.): SEFM 2025, LNCS 16192, pp. 185–202, 2026.
https://doi.org/10.1007/978-3-032-10444-1_12

ensuring that software components behave as intended, even under adversarial conditions. An extensive survey conducted in 2020 [13] revealed that experts in formal methods widely believe these approaches offer enhanced code quality, strengthened cybersecurity, and more manageable maintenance efforts. Although security experts generally recognize the value of formal verification, the views (or even awareness) of end-users are unknown [3,5].

Previous research has explored user expectations and perceptions of security in areas such as differential privacy [8] and cryptocurrency systems [27]. Similarly, prior work on formal methods has also successfully evidenced trends, limitations, and future paths for formal methods [13]. However, to the best of our knowledge, **the impact of formal verification on end-user willingness to use formally verified products is still unexplored.** Understanding users' views on formal verification can help foster the adoption of formally verified code (with all the advantages verified code can provide) and aid developers in communicating with end-users and industry partners.

Our work addresses this gap in the literature by studying formal verification's impact on users of a concrete application domain: Password Managers (PMs). PMs are applications that help users generate and manage their passwords. Most PMs offer password security features like password generation, multi-factor authentication, and secure storage with a primary password.

We chose this domain because text passwords are one of the most used security mechanisms, and PMs are an essential security tool to manage them. However, previous work suggests that users have trust issues with PMs [18,33] and are reluctant to use them due to a lack of understanding of their security properties [29]. In parallel, recent work on formal verification has been applied to PMs [14]. We designed and deployed two studies to understand users' views on formal verification. The first is a *formative interview study*. The *main study* is a more extensive quantitative study that builds upon the formative and aims to explore some of its findings in more detail.

Contributions. Overall, our main contributions are:

Addressing a critical gap in the literature: Our research provides an empirical evaluation of the influence of formal verification on end-users within a specific application domain—PMs. This work makes a contribution by addressing a gap in the formal methods literature—the study of end-users of formally verified products.

Impact of formal verification on user adoption: Our findings suggest that formal verification has a positive effect on users' willingness to adopt PMs. In both our formative user study and main study, participants showed an increased willingness to use PMs with formally verified components and identified formal verification as a desirable feature.

Identification of features for formal verification: We systematically identify a hierarchy of PM features that users consider most critical for formal verification. We identify several critical features that are particularly relevant for participants. We argue that developers and practitioners should focus their

formal verification efforts on features that users deem most critical for their trust and usage intentions, such as vault security and password generation.

Recommendations to improve formal verification communication efforts: Finally, we contribute three actionable recommendations for industry practitioners. We argue that practitioners can increase the appeal and effectiveness of verified security software by: i) addressing user-identified priorities, ii) enhancing transparency around the function and benefits of formal verification, and iii) engaging in broader educational initiatives to educate users about formal verification.

2 Scope and Research Questions

Our study aims to test users' perceptions of formal verification. To this end, we choose the domain of PMs and address the following Research Questions:

RQ1. How does formal verification impact users' willingness to use PMs?
RQ2. What features would users like to see formally verified in a PM?
RQ3. Do users value the guarantees formal verification can provide in PMs?

Summary of Methods. Our overarching goal is to understand how formal verification in PMs influences the decisions and trust of non-expert users.

The **formative study**, characterized by its qualitative and exploratory nature, focuses solely on RQ1. It aims to capture users' initial reactions and understandings of formal verification in PMs through in-depth interviews and interactions with a PM prototype. This prototype, an extended version of Bitwarden [2] incorporates a formal verification icon and explanations to educate users about formal verification's role in enhancing PM security. It gathers insights on general themes around PMs and formal verification, setting the groundwork for the subsequent larger-scale survey. The emphasis here is on identifying and exploring users' perceptions, which directly contributes to addressing RQ1 by revealing how users view and understand formal verification in the context of PMs.

The **main study** builds on the findings from the formative study, expanding the scope to address all RQs mentioned above. With 200 participants, this larger-scale survey aims to quantitatively assess the broader implications of formal verification on users' willingness to use PMs. This study aims to validate the themes identified in the formative study at a larger scale and further explore how formal verification impacts users' choices and trust in PMs. We use a combination of statistical analysis and scenario-based questions to better understand the role of formal verification across different aspects of PM usage and answer our RQs.

Crucially, we do not wish to "explain" formal verification to our participants; we merely want to communicate its consequences and impact on the PM. An overview of the methodology adopted can be seen in Fig. 1.

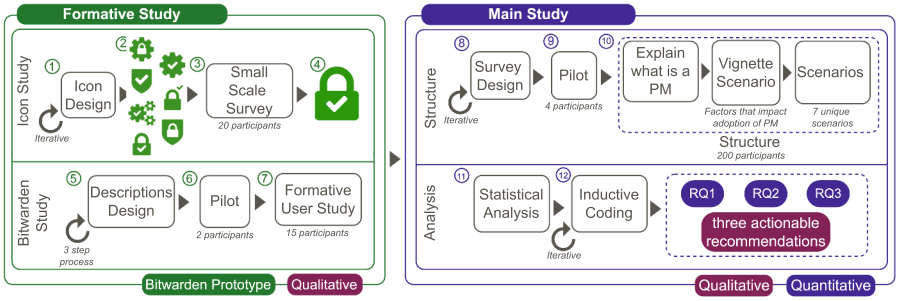


Fig. 1. Overview of the Methodology.

Ethical Considerations. In both studies, we did not collect any personal data. All participants were over 18 years old and were shown a consent form, which they had to accept before starting the survey. In the first author's institution (the lead institution for this project), studies with these properties do not need to be submitted to the Ethics Committee (as confirmed by the Chair of the Ethics Committee). The local Ethics Committee recommends reviewing the information and checklists shown in the European Commission's guide on ethical self-assessment.[1] All studies described in this paper followed these guidelines.

3 Background and Motivation

Formal verification. In general, it is hard to build secure computer-based systems. Formal methods offer the promise of software that does not have exploitable bugs and has been used in a wide range of domains to prove the correctness and security of applications [7,12,14,19,28]. The idea of formal verification is mathematically modeling a system and then using formal methods to prove that the model satisfies specific properties or specifications. For example, Chudnov et al. [7] describe the development and operation of a continuously checked proof ensuring key properties of the TLS implementation used by many AWS services. Nelson et al. [28] describe their experience applying formal methods to a critical component in the Linux kernel, the just-in-time compilers. Their results show that building a verified component within a sizeable unverified system is possible with careful specification and proof strategy design. Another line of work investigates the application of formal methods to the specification and evaluation of password composition policies. Johnson et al. [20] introduce Skeptic, a toolchain that applies formal verification and power-law modeling to automatically select and justify password composition policies. Formally verified compilers are also available. An example is the CompCert [25] compiler, which compiles code from a large subset of the C programming language to PowerPC

[1] European Commission's guide on ethical self-assessment can be accessed here https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/how-to-complete-your-ethics-self-assessment_en.pdf.

assembly code and guarantees that safety properties proved on the source code hold for the executable compiled code as well [25].

PMs. Passwords have become increasingly complex as most websites adopt strict password security policies. Previous work suggests users struggle with generating unique passwords [17], and password reuse is a problem—Pearman et al. [30] show that 40% of users reuse 81–90% of their passwords. The basic features that all PMs provide to users are: (a) A secure encrypted vault to save passwords; (b) A password generator to generate unpredictable passwords that comply with complex password security policies. Moreover, most modern PMs also provide other features, including: (c) "Autofill" to automatically fill passwords in a website's login screen; (d) Cloud synchronization to keep passwords safe and synchronized across devices. Despite being recommended by experts [10], PM usage is not widespread. Users complain about not knowing that PMs exist and not understanding how they work [6,18,29,33]. Users also do not trust PMs [29]. As an intermediary tool, PMs must be reliable, consistent, and predictable. Previous work [6,29] recommends education as a tool for enhancing adoption and trust in PMs.

Usable security. An effective security mechanism is one that is used correctly. Usable security can be traced back to Saltzer and Schroeder's 1974 paper [36] that introduces the term "psychological acceptability" for access-control systems.

In the last decade, there has been a lot of work on usable security across many domains [4,18,27,29]. Prior work explored various dimensions of security warnings, from SSL warnings in web browsers to end-to-end encryption guarantees. A notable contribution is the series of studies by Felt et al. [11] examining the impact of SSL warnings on user behavior within Google Chrome. Their work focused on experimenting with SSL warnings and assessing user adherence and comprehension. Similarly, Akhawe and Felt's [1] study provided evidence on the effectiveness of browser security warnings, revealing insights into how users perceive and react to these alerts in real-world settings. Other relevant, usable security work includes previous efforts to communicate about security using labels similar to "nutrition labels [21]." In this realm, previous research designed privacy labels for several domains, such as app stores [26] and IoT devices [9]. Related work has also evaluated the effectiveness of commonly deployed password strength meters and found significant inconsistencies between meter classifications and actual password resistance to cracking: many passwords labeled as "strong" were, in fact, easily guessed, and some "weak" passwords resisted cracking [31].

Limitations of Existing Approaches. A 2020 survey [13] of formal methods experts studied their views of formal methods. However, to the best of our knowledge, no previous work has focused on (non-expert) users of formally verified PMs, nor, more generally, on the impact that formal verification may have on end users. Previous work on formal verification of PMs is scarce, but efforts have been made to formally verify the random password generation algorithm

of a PM [14] or look at usability [3]. However, these studies concentrated on the technical aspects of formal verification without examining its usability implications or had a smaller sample size.

4 Formative Study

Our goal with this formative study was to understand users' overall views on formal verification. To achieve this, we implemented two significant extensions to a PM: a formal verification icon and an explanation.

4.1 Method

We created a proof-of-concept prototype by extending an established PM's—Bitwarden's browser extension—interface and testing it for user acceptance.

Formal verification icon. We designed an icon to communicate formal verification (④ in Fig. 1), and placed it on all features that could be formally verified: (a) Password vault; (b) Primary password input box; (c) Password generator; (d) Clipboard, by its settings. We derived the color green used in the UI from Bitwarden's blue (`#175DDC`) and grey (`#7C7C7C`), ending up with the shade (`#0BDB0B`). The design process followed three phases: (a) Initial brainstorming with security researchers that resulted in 15 icon designs, later narrowed to 7 (② in Fig. 1). (b) Contextual tests of the icons within the PMs interface. (c) A survey with 20 users with two preference tests [15] and ratings in the icon's intended context.

Descriptions of formal verification. To explain the role of formal verification in the PM, we began by identifying all verifiable features. We developed the explanations in three phases: (1) We began with one-on-one discussions with formal verification researchers, gathering explanations, and removing jargon (⑤ in Fig. 1). (2) We then met with the team, gathered feedback, and applied it. (3) Finally, after another round of feedback, we reached a consensus and finalized the explanations. We compared the base PM (i.e., without any extensions or formal verification) and the extended PM (i.e., with interface updates).

Structure. The user tests were divided into the following parts. First, we provided users with a brief introduction to the study with a "*Pre-Task Questionnaire*" about their experience with PMs and demographics. We asked participants to perform everyday tasks in a PM (e.g., save a password). And finally, we asked them to fill out the "*Final Questionnaire*". Each questionnaire was used as a base for semi-structured interviews. We collected data through questionnaires and observation.

Recruitment and Participant demographics. Our sample comprised 15 participants, 10 for the extended interface and 5 for standard interface. Participants were recruited through the authors' network and received no compensation. Most of the participants (60%) had higher education. Of the 15 participants, only 2 had a technical background related to IT. The most frequent age group was 25–34 (40%). Overall, 60% of users were younger than 34. Gender in our sample was evenly divided. Each interview took around 50 min, and 90% of participants were unfamiliar with formal verification. We performed two pilot interviews to refine the protocol and interview script (⑥ in Fig. 1).

4.2 Results

To understand if participants knew the icon's meaning (see ④ in Fig. 1), we asked them to explain it. Of the 10 participants, 5 mentioned formal verification, and the other five mentioned concepts related to the security of the PM (e.g., "*the icon means that the passwords were safe*" (P1)).

Some users correctly identified the formally verified features. Another aspect we explore is whether participants understand which features are formally verified. Specifically, we found that: (a) 6 out of 10 identified the generator and the storage as formally verified; (b) 3 out of 10 stated that the whole PM was formally verified; when asked why one user stated "*I saw the icon in several places*" (P2); (c) and, one user could not explain, stating they did not know.

Users may be more willing to use a formally verified PM. When comparing users' responses regarding trust in a formally verified PM *vs* one that is not formally verified, users stated trusting more the formally verified PM.

Limitations and Motivation for Second User Study. A key limitation is the small sample size, which may not reflect a diverse population. Furthermore, our explanation of formal verification may have led participants to confuse security and verification. To avoid that issue in our main study, we completely refrain from explaining formal verification; instead, we communicate about the guarantees it provides. Another limitation is using a specific PM, which might threaten the study's generality. So, we decided not to use any specific PM interface in our main study. After gathering the main insights from this explorative study, we designed and deployed a large-scale quantitative study.

5 Main Study

To address all the mentioned in Sect. 2, we designed and conducted a 200-participant online survey to understand users' perceptions of formal verification (⑧ in Fig. 1).

Table 1. Percentage of participants that *agreed* or *strongly agreed* that the factor would impact their willingness to use a PM.

Results	Factors
69.0%	being inexpensive
73.5%	having support materials (e.g., tutorials)
76.5%	being certified by Password Manager Security Group
78.0%	being free
81.5%	being made by a trustworthy and familiar company
86.5%	being mathematically correct
87.0%	being easy to use for first-time/beginner users

5.1 Structure

First Section. One of our goals was to test whether formal verification in a PM affects users' willingness to use it. First, we ensured participants knew what a PM was (prior work suggests users may not be familiar with PMs [29,37]). To do this, we show participants a vignette scenario where they read a news story about what a PM is. After viewing this scenario, we show participants a survey question addressing RQ1. We ask "*Imagine that you were thinking about using a PM. What would make you more willing to use the PM?*" We then present participants with different factors to which they respond using a 5-point Likert scale [23]. Among these factors, we include formal verification. The reason we added other factors in addition to formal verification was to prevent participants from realizing that the focus of the study was formal verification and prevent common biases where the users are inclined to agree with the researcher. We also wanted to understand if users' preference for formally verified PMs differed from other factors (see Table 1). One of the factors used was a certification by the "*Password Manager Security Group*". This factor did not represent an actual entity. It was added to understand if participants were inclined to agree that any factor was impactful, without understanding, or because it sounded good.

Finally, we omit the term "*formal verification*" by using analogies when possible. The specific analogy we chose is related to something most participants had to deal with in their education—mathematics. Instead of saying that the PM is formally verified, we state that it is "*mathematically correct, that is, its features are as trustworthy as a mathematical proof.*" We do this to follow best practices [34] and exclude jargon from our survey, as most users are unfamiliar with formal verification. Media outlets have previously used mathematical analogies to explain formal verification to broader audiences (e.g., BBC [35] and Quanta Magazine [16]).

Second Section. After asking participants if they found that formal verification (among other factors), we presented participants with their answers and asked, respectively, "*In the previous question, you stated you felt more/less willing to*

use a PM that is mathematically correct (...). Please state your reasons." We hoped to understand why participants valued (or not) formal verification in a PM and what they associated it with.

Third Section. To answer RQ2 and RQ3, we gather common PM features (e.g., Password Generator and Clipboard clearing). For each of these, we present scenarios representing the impact formal verification can have on each feature. For example, for the Password Generator, the policy compliance scenario is: "*Imagine that you are creating a new account on a website (e.g., Twitter, Facebook). To increase security, you ask the PM to generate a password with seven characters and at least two numbers. However, the password generated does not include any numbers.*". After each scenario, and using a 5-point Likert scale, we ask users if that scenario would make them stop using a PM. A feature may have multiple scenarios if formal verification can impact it differently. An example is, again, the password generator, where formal verification may help with policy compliance but also with the guarantee that the generator is truly unpredictable (i.e., all passwords have the same probability). We followed design guidelines [23] to iteratively develop seven unique scenarios

5.2 Recruitment

We recruited participants through Prolific.[2] We recruited 200 participants who were paid 2.65 GBP per submission (with one submission per user and an average time between 8 and 9 min).

Participant demographics. About half of our sample identified as male and half as female. Our sample skewed younger than the general population, as most participants were less than 34 years (87%), and the average age was 26. Most participants knew what a PM was before the study (77.5%), but a slight majority of them stated that they did not use one currently (52%). The rest of the users were divided between those who said they were currently using a PM (40%), those who had used one in the past but not anymore (3.5%), and those who did not know (3.5%). The three PMs that participants mentioned the most were Google Chrome PM (41%), Bitwarden (15%), and Apple Keychain (14%).

Nonetheless, due to a lack of understanding about what a PM is, some participants answered that they had never used a PM when they used a built-in PM (e.g., when they received prompts to save a password in a browser). We saw that, of the 52% of participants who had never used a PM, 67% of them saved passwords when prompted (e.g., in built-in PMs). **These results seem to imply that users use PMs without realizing they are using them,** maybe due to a lack of technical knowledge or low computer literacy. In total, **77% of participants used PMs to store and manage passwords.**

[2] Prolific is a crowd-sourcing platform that enables large-scale user studies by connecting research and users https://prolific.com.

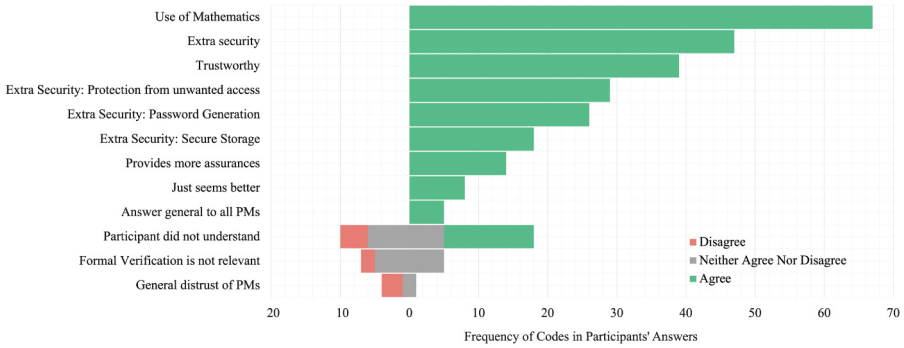


Fig. 2. Frequency of codes representing the reasons participants gave for "agreeing" or "disagreeing" that they felt willing to use a mathematically correct PM.

5.3 Analysis

To understand if there was a significant difference between different survey questions, we used Friedman's test and Wilcoxon Signed-Rank Test with Bonferroni continuity correction [23]. We coded the open answers from the survey using an inductive coding strategy (⑫ in Fig. 1). Two coders created the codebook iteratively using inductive and hierarchical coding [23]. We coded 200 answers, achieving a Cohen's Kappa of 0.81, indicating very high inter-rater reliability [23]. All answers were coded individually, and regular meetings happened to discuss ambiguities and updates to the codebook. The final codebook consisted of 13 codes and each participant's answer could have multiple codes. We used a hierarchical coding scheme for some of the codes, for example, the code "*Extra Security*" had as sub-codes: "*Extra security: password generation*", for when participants mentioned that the PM could create strong passwords and "*Extra security: secure storage*", for when the participants stated that their passwords were safer inside the PMs' storage. We also included more general codes for when participants were not specific in their answers or showed a lack of understanding about what was being asked ("*Answer general to all PMs and not specific to formally verified PMs*" and "*Just seems better*").

5.4 Results

In this section, we address each RQ and its respective study insights.

Formal Verification and Willingness to Use a PM (RQ1). In the survey's first question, we asked participants whether formal verification impacted their willingness to use a PM. To reduce biases, we compared seven positive factors (see Table 1). The factors are all things that could be perceived as desirable to participants, so it is no surprise that most participants "*agreed*" or "*strongly agreed*" would be more likely to use a PM with any of those factors. What we

want to understand is if their opinion is the same across all factors or if there are some factors that users value more. A Friedman test showed that there was a statistically significant difference in participants' answers between the different PM's features with a $\chi^2(2) = 51.42$ and a $p < 0.05$. However, the following pairwise comparison of users' answers using the Wilcoxon rank sum test did not find a significant difference between most factors ($p > 0.05$).

Comparing all the factors, we found that more than 86% of participants considered the use of formal verification (as described in Sect. 5.1) and being easy to use as important (they "*agreed*" or "*strongly agreed*" that these factors would make them more willing to use a PM). The factors that users considered less critical (while still being essential for most) were being inexpensive (i.e., it costs money even if it is not expensive) and having support features (e.g., tutorials and help pages) with 31% and 27%, respectively, not thinking they would be more willing to use a PM with them. This information can be found in Table 1. Our results suggest that users are more willing to use a formally verified PM than a non-verified one. With this information in mind, we now aim to understand why participants answered our survey the way they did.

Understanding Users' Reasons. We separated answers into three categories: participants who valued formal verification, had no opinion, and found it undesirable (see Fig. 2). Most participants who valued formal verification in their PMs cited the **use of mathematics** as a reason (67 participants (35%)). Participants seem to associate mathematics with a "*logical*" and precise behavior. One even stated that they preferred to use a formally verified PM "*because mathematics doesn't lie*" (P89). Another stated, "*I like to use such products or services that have scientific proof of their effectiveness*" (P147).

"Security" codes were also frequent (see Sect. 5.3). Some participants mentioned they believe formal verification added security (e.g., "*I think it would be more secure.*" (P91)), and others gave more detailed explanations (e.g., "*The main reason to start using a PM is (...) the passwords generated are the strongest (...)*" (P5)). Over 50% of the participants mentioned security as a reason why they valued formally verified PMs, with some mentioning more than one specific subcode of security in their answers (e.g., one participant stated that they preferred a formally verified PM because "*... it will be more difficult to get hacked since it will create more secure passwords*" (P162)).

Overall, participants mostly agreed that formal verification was something to be desired in a PM. The most frequent reasons for this were related to security. On the other hand, participants who did not find formal verification relevant had a general distrust of PMs or found that formal verification was not essential for them. For example, one of the participants directly mentioned, *"I just don't believe in that kind of app"* (P22). Lack of understanding was also a frequent code present in the answers of participants who had a neutral response to this survey question (see the bars in grey in Fig. 2). Lack of understanding of technical concepts has been the cause of users' trust issues with other security software in the past [32]. Nonetheless, most participants who

Table 2. P-values of pairwise comparisons of users' answers in the scenarios question using the Wilcoxon rank sum test and Bonferroni continuity correction.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7
Scenario 2	2e-16	-	-	-	-	-	-
Scenario 3	0.00013	2e-16	-	-	-	-	-
Scenario 4	0.47137	2e-16	1.2e-05	-	-	-	-
Scenario 5	0.00011	3.3e-10	3.2e-13	0.00220	-	-	-
Scenario 6	3.2e-05	1.3e-10	1.6e-14	0.00097	0.89316	-	-
Scenario 7	6.4e-10	2.8e-06	2e-16	2.2e-07	0.03915	0.03915	-
Scenario 8	1.1e-14	0.00138	2e-16	1.8e-11	0.00024	0.00019	0.08562

agreed that formal verification was important to them stated that security had something to do with it.

RQ1. How does formal verification impact users' willingness to use PMs?

- Users seem to be more willing to use a formally verified PM than a non-formally verified PM;
- Formal verification seems to have positively affected some users' trust;
- Users primarily value formal verification for its reliance on mathematics and the additional security it provides.

Formally Verified Features (RQ2). After learning that formal verification seems to impact users' interest in using PMs, we are interested in learning their priorities, thus addressing RQ2. We realize that users may not have the expertise to judge what is essential and what should be verified in a PM. However, our goal is to understand what features, if verified, would be more impactful for users. We want to understand how to meet users' wants and maximize the impact of formal verification. A Friedman's test showed that there was a statistically significant difference in participants' answers between the different scenarios with a $\chi^2(2) = 331.08$ and a $p < 0.05$. The results of a pairwise comparison of users' answers using the Wilcoxon rank sum test and Bonferroni continuity correction can be seen in Table 2. Our results suggest that most of the differences between scenarios are statistically relevant ($p < 0.05$).

Overall, participants seem to consider some features more critical than others. Our results suggest that the secure vault and its encryption are the most important—96% of participants stated they *agreed* or *strongly agreed* that they would stop using a PM if the respective scenario (S3) occurred. The most important features to verify seem to be the secure vault (S3), password generator (S1), login in the PM (i.e., primary password security in S4), clipboard (S6), synchronization across devices (S7 and S8) (see Table 3).

Table 3. Percentage of participants that *agreed* or *strongly agreed* the scenario would make them stop using a PM. S# corresponds to Scenario #

Results	Scenarios
54.50%	S2 Policy Compliance
70.50%	S8 Ransomware/ Deleting your vault
76.00%	S7 Synchronization
79.00%	S5 Autofill
83.00%	S6 Clipboard Clearing
86.50%	S4 Primary Password Exposure
92.50%	S1 Unpredictability
96.00%	S3 Vault Exposure

RQ2. What features would users like to see formally verified in a PM?

- The password vault's security seems to be the feature that should be prioritized in a future implementation of a formally verified PM;
- Surprisingly, verifying that the PM's passwords will not be lost does not seem to be important for users, as scenarios where there was no third-party access to the password were not as impactful;
- Ideally, all the features mentioned in this section should be formally verified, as over 50% of participants found their scenarios impactful.

Formal Verification Guarantees (RQ3). The scenario's question also provides insights into whether participants value the guarantees of formal verification in PMs. As stated before, most scenarios were impactful for users, so in general, our results suggest that participants value the guarantees that formal verification provides. We can divide our scenarios into the following categories:

- Scenarios where a third-party learns all the passwords: "Vault Exposure" (S3) and "Primary Password Exposure" (S4);
- Scenarios where a third party, over time, can learn manypasswords: "Unpredictability" (S1), "Autofill" (S5), and "Clipboard Clearing" (S6);
- Scenarios where the user loses access to all the passwords stored in the PM's vault: "Synchronization" (S7) and "Deleting your vault" (S8);
- A scenario where the generator does not behave as intended and the passwords are not compliant with the specification: "Policy Compliance" (S2).

Two of the three most impactful scenarios allowed a third party to learn all user passwords (S3, S4). In these scenarios, 86.5% and 92.5% of participants stated they *"agreed"* or *"strongly agreed"* that they would stop using a PM if it

happened. These scenarios involved the lack of formal verification, which would otherwise prevent password exposure to third parties.

However, the second, fourth, and fifth most impactful scenarios (S1, S5, and S6) described a situation where a third party, over time, could learn several users' passwords. In these, over 79% of participants *"agreed"* or *"strongly agreed"* that they would stop using a PM if they happened. The other type of scenario, where the user loses access to all the passwords they have stored in the PM's vault (S7 and S8), was not as impactful for participants as the ones mentioned before. Finally, the last type of scenario described a situation where the PM did not work as intended (S2), but its malfunction did not lead to a loss of passwords. It just inconvenienced users. This was the scenario that fewer participants found impactful. Interestingly, the least important feature in our hierarchy, the unpredictability of the password generator (S1), is a feature that previous work has made efforts to formally verify [14].

RQ3: Do users value the guarantees that formal verification can provide in PMs?

- Our results suggest that users value the guarantees that formal verification can provide in PMs;
- Our data also seems to indicate that some guarantees are more important than others. For example, guaranteeing that the passwords are not leaked to third parties seems more relevant for participants than guaranteeing that they are synced in the cloud.

6 Discussion

In this section, discuss the main insights and suggest recommendations.

User Awareness and Perception. An insight we got was that there is an overall low level of awareness among users about what formal verification is. Despite this, users react positively to formal verification in PMs. There appears to be a gap between the technical understanding of formal verification and the perceived value it adds to PMs. Some of our participants liked formal verification due to an inherent trust in the mathematical and technical rigor it implies, suggesting that even a superficial understanding or awareness of formal verification may influence user preferences.

User features prioritization. A significant insight from this study is that users perceive some PM features as more important. The security of the password vault and the reliability of password generation were highlighted as particularly critical features for users. This insight underscores the need for developers to prioritize formal verification efforts on features that most directly impact user trust and perceived security. Understanding user priorities can guide the allocation of development resources towards the aspects of PMs that users value more and enhance the effectiveness and attractiveness of formally verified PMs.

6.1 Recommendations

One of our goals was to identify how to increase the impact of formal verification in software. To achieve this, and based on the results of our user studies, we present several recommendations for developers of formally verified products.

1. Increase user familiarity with formal verification. Our results suggest that users do not know what formal verification is. With this in mind, developers of formally verified software should try to convey to users the role of formal verification in their software. We suggest that these efforts should be focused on the specific consequences of formal verification in software instead of theoretical concepts. Avoiding technical language may facilitate the application of best practices such as using metaphors and avoiding jargon [34]. Increasing this transparency may enable users to understand the advantages of formally verified software and, thus, increase its adoption. If formally verified software is valued by users, it may also increase its investment value for businesses (for example, for PMs such as Bitwarden).

2. Don't overstate/mislead users when explaining formally verified software. As mentioned before, some users associated formal verification with security (see Sect. 4.2). While this can sometimes be true, security is not necessarily related to formal verification. As such, we suggest that future efforts to communicate about formal verification should take care to prevent overstating the impact it has. To build trust, communication efforts should carefully distinguish the benefits of verification from broader security properties.

3. Make an effort to understand the users' priorities in formally verified software. Even if you do not intend to make formal verification a selling point of the product, we suggest that it is important to understand what are the users' priorities. Understanding users' wants can provide relevant insight into what can make a product more desirable. Moreover, since formally verifying software is time-consuming and expensive, understanding users' priorities may help focus verification efforts on the features that are more impactful.

6.2 Limitations

User studies such as these may suffer from bias. Bias can arise from the questions or even the questionnaires. To mitigate our limitations, we performed cognitive interviews[3] to validate the study until no more errors or typos were detected (we did two cognitive interviews for each study). We also followed best practices by offering "don't know (DK)" or "prefer not to answer" responses [34]. Moreover, we asked users to explain their understanding of the topic to mitigate the Dunning-Kruger Effect (e.g., when they stated that they understood what formal verification is, we asked them to explain what it is) and included attention check questions. For example, we include one of these questions among the factors that

[3] *Cognitive interviews* involve asking respondents to think aloud as they complete a survey and asking them questions about each survey item [34].

may impact users' willingness to use a PM (see Sect. 5.1) and another among the scenarios in the third section of the survey (see Sect. 5.1). We also removed as much jargon as possible. Jargon includes terms like "memory", "encrypted" and "formal verification" itself [34] and tried to mitigate the Hawthorne effect. To prevent bias, we randomized the order in which the factors that may impact users' willingness to use a PM are shown (see Sect. 5.1), and the order of the scenarios in the third survey section (see Sect. 5.1).

6.3 Future Work

Exploring formal verification's impact on users' willingness to use PMs opens up several avenues for further research.

Exploration in Other Domains. Our study within the context of PMs suggests a positive user reception towards formal verification. Extending this research to other domains where formal verification is applied could offer an insightful research opportunity. For example, investigating user perceptions in the context of autonomous vehicles, medical devices, and blockchain technologies—all of which rely on the integrity and security assurances that formal verification provides—could reveal domain-specific user attitudes and expectations. These studies could help understand whether the positive inclination towards formal verification observed in PM users is universal or if adjustments in communication and implementation strategies are needed based on the application domain.

Development of New Ways to Communicate about Formal Verification. One significant finding from our study is the general lack of user awareness about formal verification. This gap presents a research opportunity to develop and test new communication methods about formal verification for the public. Inspired by initiatives like the IoT security labels [9], a "Formal Verification Label" could be designed to provide at-a-glance information about the verification status of a product. Such a label could include simplified symbols or ratings that indicate the extent and areas of formal verification applied, making it easier for users to understand the claims of formally verified software products.

7 Conclusion

Formal software verification can be expensive and time-intensive, but our work suggests that it may positively impact users. In our study, we propose several directions for future research while providing concrete insights into which features should be prioritized by practitioners, such as enhancing the security of the password vault and ensuring the reliability of password generation. Moreover, our work has shed light on a previously mostly unexplored area of research, combining formal verification with usable security. We argue that many paths for future work could be explored: virtually any domain where formal methods could be applied could benefit from studying user perception on the subject.

Acknowledgments. This work was financed by National Funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) under grant PRT/BD/153739/2021 and project UIDB/50021/2020 (DOI: 10.54499/UIDB/50021/2020). Work done in the context of the PassCert project, a CMU Portugal Exploratory Project funded by FCT with reference CMU/TIC/0006/2019.

References

1. Akhawe, D., Felt, A.P.: Alice in warningland: a large-scale field study of browser security warning effectiveness. In: 22nd USENIX Security, pp. 257–272 (2013)
2. Bitwarden: Bitwarden (2023). https://bitwarden.com Accessed 10 Jun 2025
3. Carreira, C.: Studying users' willingness to use a formally verified password manager. In: International Conference on Integrated Formal Methods, pp. 343–346. Springer (2022)
4. Carreira, C., Ferreira, J.F., Mendes, A.: Towards improving the usability of password managers. INFORUM (2021)
5. Carreira, C., Ferreira, J.F., Mendes, A., Christin, N.: Exploring usable security to improve the impact of formal verification: a research agenda. First Workshop on Applicable Formal Methods (co-located with Formal Methods 2021) (2021)
6. Chiasson, S., van Oorschot, P.C., Biddle, R.: A usability study and critique of two password managers. In: USENIX Security Symposium (2006)
7. Chudnov, A., et al.: Continuous formal verification of Amazon S2N. In: International Conference on Computer Aided Verification, pp. 430–446. Springer (2018)
8. Cummings, R., Kaptchuk, G., Redmiles, E.M.: I need a better description: an investigation into user expectations for differential privacy. In: ACM SIGSAC (2021)
9. Emami-Naeini, P., Agarwal, Y., Cranor, L.F., Hibshi, H.: Ask the experts: what should be on an IoT privacy and security label? In: SP. IEEE (2020)
10. ENISA: European union agency for cibersecurity authentication methods (2025). https://www.enisa.europa.eu/topics/cyber-hygiene Accessed 10 Jun 2025
11. Felt, A.P., et al.: Improving SSL warnings: comprehension and adherence. In: CHI (2015)
12. Ferreira, J.F., Johnson, S., Mendes, A., Brooke, P.: Certified password quality: a case study using coq and Linux pluggable authentication modules. In: 13th International Conference on Integrated Formal Methods (2017)
13. Garavel, H., Beek, M.H.t., Pol, J.v.d.: The 2020 expert survey on formal methods. In: Formal Methods for Industrial Critical Systems: 25th International Conference, pp. 3–69. Springer (2020)
14. Grilo, M., Campos, J., Ferreira, J.F., Mendes, A., Almeida, J.B.: Verified password generation from password composition policies. In: 17th International Conference on Integrated Formal Methods (2022)
15. Harley, A.: Usability testing of icons (2016). www.nngroup.com/articles/icon-testing/ Accessed 10 Jun 2025
16. Hartnett, K.: Hacker-proof code (2020) . https://www.quantamagazine.org/formal-verification-creates-hacker-proof-code-20160920/ Accessed 10 Jun 2025
17. Inglesant, P.G., Sasse, M.A.: The true cost of unusable password policies: password use in the wild. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp. 383–392 (2010)

18. Ion, I., Reeder, R., Consolvo, S.: ...No one can hack my mind: comparing expert and non-expert security practices. In: Eleventh Symposium On Usable Privacy and Security (SOUPS 2015), pp. 327–346 (2015)
19. Jangid, M.K., Chen, G., Zhang, Y., Lin, Z.: Towards formal verification of state continuity for enclave programs. In: USENIX Security Symposium (2021)
20. Johnson, S., Ferreira, J.F., Mendes, A., Cordry, J.: Skeptic: automatic, justified and privacy-preserving password composition policy selection. In: 15th ACM Asia Conference on Computer and Communications Security (2020)
21. Kelley, P.G., Bresee, J., Cranor, L.F., Reeder, R.W.: A "nutrition label" for privacy. In: SOUPS, pp. 1–12 (2009)
22. Kumar, R., Myreen, M.O., Norrish, M., Owens, S.: CakeML: a verified implementation of ML. ACM SIGPLAN Notices **49**(1), 179–191 (2014)
23. Lazar, J., Feng, J.H., Hochheiser, H.: Research methods in human-computer interaction. Morgan Kaufmann (2017)
24. Leroy, X.: Formal certification of a compiler back-end or: programming a compiler with a proof assistant. In: ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, pp. 42–54 (2006)
25. Leroy, X.: Formal verification of a realistic compiler. Commun. ACM **52**(7), 107–115 (2009)
26. Li, Y., Chen, D., Li, T., Agarwal, Y., Cranor, L.F., Hong, J.I.: Understanding IOS privacy nutrition labels: an exploratory large-scale analysis of app store data. In: CHI (2022)
27. Mai, A., Pfeffer, K., Gusenbauer, M., Weippl, E., Krombholz, K.: User mental models of cryptocurrency systems-a grounded theory approach. In: Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020), pp. 341–358 (2020)
28. Nelson, L., Van Geffen, J., Torlak, E., Wang, X.: Specification and verification in the field: Applying formal methods to BPF just-in-time compilers in the Linux kernel. In: OSDI '20, pp. 41–61 (2020)
29. Pearman, S., Zhang, S.A., Bauer, L., Christin, N., Cranor, L.F.: Why people (don't) use password managers effectively. In: SOUPS (2019)
30. Pearman, S., Thomas, J., et al.: Let's go in for a closer look: observing passwords in their natural habitat. In: ACM SIGSAC (2017)
31. Pereira, D., Ferreira, J.F., Mendes, A.: Evaluating the accuracy of password strength meters using off-the-shelf guessing attacks. In: ISSREW. IEEE (2020)
32. Presthus, W., O'Malley, N.O.: Motivations and barriers for end-user adoption of bitcoin as digital currency. Procedia Comput. Sci. **121**, 89–97 (2017). https://doi.org/10.1016/j.procs.2017.11.013
33. Ray, H., Wolf, F., Kuber, R., Aviv, A.J.: Why older adults (don't) use password managers. In: USENIX Security Symposium (2021)
34. Redmiles, E.M., Acar, Y., Fahl, S., Mazurek, M.L.: A summary of survey methodology best practices for security and privacy researchers. Technical report (2017)
35. Rubens, P.: How playing computer games can make the world safer (2015). https://www.bbc.com/news/business-33519194 Accessed 10 Jun 2025
36. Saltzer, J.H., Schroeder, M.D.: The protection of information in computer systems. Proc. IEEE **63**(9), 1278–1308 (1975)
37. Stobert, E., Biddle, R.: The password life cycle: user behaviour in managing passwords. In: SOUPS 2014 (2014)