The U in Crypto Stands for Usable: An Empirical Study of User Experience with Mobile Cryptocurrency Wallets

Artemij Voskobojnikov
voskart@ece.ubc.ca
University of British Columbia

Oliver Wiese
oliver.wiese@fu-berlin.de
Freie Universität Berlin

Masoud Mehrabi Koushki
mehrabi@ece.ubc.ca
University of British Columbia

Volker Roth
volker.roth@fu-berlin.de
Freie Universität Berlin

Konstantin Beznosov
beznosov@ece.ubc.ca
University of British Columbia

ABSTRACT

In a corpus of 45,821 app reviews of the top five mobile cryptocurrency wallets, we identified and qualitatively analyzed 6,859 reviews pertaining to the user experience (UX) with those wallets. Our analysis suggests that both new and experienced users struggle with general and domain-specific UX issues that, aside from frustration and disengagement, might lead to dangerous errors and irreversible monetary losses. We reveal shortcomings of current wallet UX as well as users' misconceptions, some of which can be traced back to a reliance on their understanding of conventional payment systems. For example, some users believed that transactions were free, reversible, and could be canceled anytime, which is not the case in reality. Correspondingly, these beliefs often resulted in unmet expectations. Based on our findings, we provide recommendations on how to design cryptocurrency wallets that both alleviate the identified issues and counteract some of the misconceptions in order to better support newcomers.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; • **Security and privacy** → **Usability in security and privacy**.

KEYWORDS

Cryptocurrency, natural language processing, thematic analysis, review analysis

ACM Reference Format:
Artemij Voskobojnikov, Oliver Wiese, Masoud Mehrabi Koushki, Volker Roth, and Konstantin Beznosov. 2021. The U in Crypto Stands for Usable: An Empirical Study of User Experience with Mobile Cryptocurrency Wallets. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3411764.3445407

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445407

1 INTRODUCTION

Over the past years, the cryptocurrency domain has grown substantially. As of 2020, there exist over 6,000 different cryptocurrencies with a combined worth of over $350 billion USD [11]. The number of users, while difficult to measure, is estimated to be 60 million worldwide [42] and is growing daily. This user base has also changed significantly and is no longer only made up of cypherpunks and computer experts [28, 53] as was the case in the early days of Bitcoin. One of the main reasons for this phenomenon is widely accessible software tools that lowered the entry barriers and made it possible to purchase and own cryptocurrencies with a single click.

Managing cryptocurrencies using such tools, however, has been found to be challenging for the users. Eskandari et al. [16] conducted cognitive walkthroughs with bitcoin wallets and identified shortcomings, such as misunderstood metaphors and abstractions, that could lead to user errors. Voskobojnikov et al. [61] interviewed cryptocurrency users and found that the majority of them found current software tools difficult and unintuitive to use. Despite these studies and the evident prevalence of user experience (UX) shortcomings, it remains unknown what features of current wallets contribute to the poor UX, why, and to what extent.

To address this knowledge gap, we analyzed reviews of mobile cryptocurrency wallets (referred to as mobile wallets from now on). We first built a crawler and collected 45,821 reviews of the five most popular mobile (both Android and iOS) wallets, which, only on Android, have over nine million installations. Then, by using supervised machine learning (ML) techniques and natural language processing (NLP), we identified 6,859 reviews that included information relevant to UX. We then analyzed these relevant reviews using Thematic Analysis [21].

We uncovered a wide range of issues that appear to negatively affect the UX. Both *general* and *domain-specific* UX issues were found to have an effect, with some leading to dangerous errors and irreversible monetary losses. General issues, such as app crashes and freezes, while known to be prevalent in mobile apps in general [22, 23, 34], had a more severe effect on mobile wallets, and often resulted in users losing access to their assets.

Other issues, however, were unique to cryptocurrencies and their tools. For example, some reviewers had conventional payment systems in mind when using the apps and believed that, similar to online banking, transactions had fixed fees, were reversible, and could be canceled, none of which are the case in reality. Further, reviewers also had limited understanding of the underlying technology and were confused about the fundamental building blocks,

including public key cryptography, recovery mechanisms, transactions, and even the role of a wallet. Often, these misconceptions resulted in reviewers blaming the wallet providers even when the latter were not at fault. Based on our results, we recommend that future wallets mimic the functionality of traditional payment systems, create distinct profiles for advanced users and novices, and provide more guidance during setup and use.

Our work makes the following contributions. We investigate UX issues of mobile wallets, their prevalence, and their effect on users. We identify users' misconceptions about the building blocks of cryptocurrencies that lead to dangerous errors and monetary losses. We provide design recommendations for addressing the identified UX issues and potentially enhancing UX in the domain.

2 BACKGROUND AND RELATED WORK

2.1 Cryptocurrencies and Wallets

Cryptocurrencies are digital currencies that are cryptographically secured and have no centralized governing party. The most prominent example is arguably bitcoin [46], a cryptocurrency that was released in 2009. As of January 2021, it is worth over $700 billion USD [11], making it the most valuable cryptocurrency in terms of market capitalization. In the following years, many other cryptocurrencies were created, some of which allow people to create their own tokens on top of the respective network. The Ethereum blockchain [63] is the most popular platform that allows the creation of tokens and contributed significantly to the now over 6,000 different cryptocurrencies [11]. These tokens are bound to that specific network, and whenever tokens are sent from one address to another, the transaction fees are paid in the cryptocurrency whose network the tokens exist on. These fees are paid to the miners, i.e., to the entities that process transactions and include them in the respective blockchain [3]. Here, transactions with higher fees are prioritized and are included more quickly than ones with lower fees.

Despite the large number of different cryptocurrencies, they all make use of public key cryptography. Addresses consist of cryptographic hashes of the public key, and private keys are needed to be able to transfer funds, as they are used to sign a transaction [3]. These cryptographic key pairs are often stored in so-called wallets, which allow the users to manage their keys. These wallets are the primary interface for cryptocurrency users and can be grouped into *custodial* and *non-custodial* wallets [18]. Custodial wallets take care of the key management, whereas non-custodial wallets allow the end users to manage the keys themselves. In this paper, we focus on non-custodial mobile wallets (both iOS and Android) and report on the UX challenges users face when using them.

2.2 UX and Cryptocurrencies

Cryptocurrency users are known to struggle with the management of cryptographic keys. Studies have shown that users have inadequate mental models of the underlying cryptography [19, 43] and often employ poor security practices leading to dangerous errors and, in the worst case, monetary losses [18, 35, 61].

Dangerous errors, however, can also occur as a result of poor wallet UX. Eskandari et al. [16] conducted the first and only usability study of cryptocurrency wallets. Through a series of cognitive walkthroughs of six bitcoin wallets, the authors identified issues that could potentially affect users, including, but not limited to, complex metaphors, highly technical terminology, and a general lack of guidance. Due to the nature of the study, however, it remains unclear what effects such issues have on the UX.

Our contribution is twofold. Firstly, in a large body of app reviews of mobile wallets, we identify and categorize UX issues, and assess their severity and prevalence. Secondly, guided by the call for blockchain design patterns in the work of Elsden et al. [14], we propose design recommendations that address the identified issues. These recommendations can inform future design of wallets and can potentially contribute to an enhanced UX in the domain.

2.3 Review Analysis

App reviews provide rich, contextual information about an app as perceived by its users. Such information can include feature requests, bug reports, and functional error reports [23, 34], which, when analyzed, can help improve the app. Due to the sheer number of reviews, however, this analysis can become very tedious.

To reduce this burden, automated classification approaches have been proposed by scholars. Here, ML and NLP techniques were successfully applied to retrieve informative reviews related to general app maintenance [10], user feedback [20, 22], and UX [4, 27, 40, 45, 50]. For example, Hedegaard and Simonsen [27] proposed an approach to automatically classify reviews based on the dimensions of usability and user experience (UUX) found in literature. Results suggest that close to half of the reviews included information related to UUX, some of which could be used to improve the apps. Inspired by the previous work, we employ a combination of ML and NLP techniques to identify reviews that contain information about the UX in the context of cryptocurrencies. In our case, these techniques are used to narrow down the review corpus to a humanly manageable size, which can then be analyzed qualitatively.

3 METHODOLOGY

We used a three-stage approach (see Figure 1) to investigate the shortcomings of cryptocurrency wallets as perceived by the users. For data collection, we built a crawler to generate a list of reviews (and their metadata) from both the Apple App Store and Google Play Store. As review data is noisy [45] and only a subset of reviews is relevant to UX [27, 49], we decided to automate this selection process by using a combination of heuristics and an ML classifier. Lastly, we performed qualitative analysis of the subset of the reviews selected during the second stage. This third stage allowed us to provide a rich, contextual, in-depth analysis of the data. In the following sections, we describe each stage in detail.

3.1 Data Collection

3.1.1 App Selection. There exist two options when it comes to managing cryptocurrencies: custodial and non-custodial wallets. The latter allow its users to manage the cryptographic keys themselves, which, as prior work has found [18, 55, 61], users often struggle with. To investigate the underlying reasons for these struggles, we focused on non-custodial wallets and excluded custodial options,

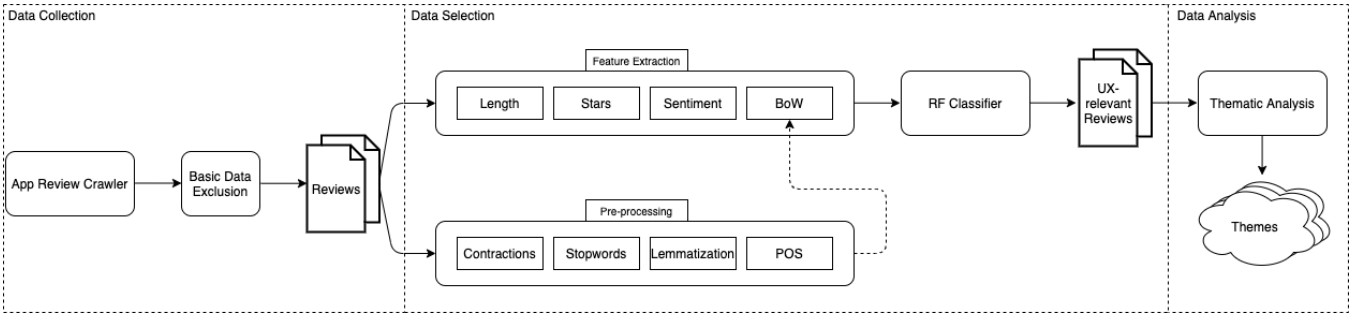


Figure 1: Overview of the data collection, selection, and analysis approach (POS = part of speech, BoW = bag of words, RF = random forests)

i.e., apps that only serve as an interface to a cryptocurrency exchange, such as Coinbase[1] or Binance.[2] In particular, we decided to investigate mobile wallets, as we found them to be the second-most-popular wallet type, after desktop wallets [2], where reviews are scarce or non-existent.

The number of available mobile wallets, however, is overwhelming. In the Google Play Store alone, there are over 250 different apps with tens of thousands of reviews, which make it infeasible to analyze them manually, even after selecting only UX-relevant reviews. Therefore, we selected the top five mobile wallets in the Google Play Store, with the most reviews as of April 2020: Blockchain Wallet,[3] Trust Crypto Wallet,[4] Electroneum,[5] Coinomi Wallet,[6] and BRD Bitcoin Wallet.[7] We then selected the same applications and their corresponding reviews in the Apple App Store. Four of the top five wallets on Android were also among the top five wallets on iOS. The selected mobile wallets support a variety of different cryptocurrencies, such as bitcoin, Ether, and various tokens, and we believe that insights gained from analyzing these reviews will be transferable to other types of software wallets.

3.1.2 Review Crawler. Similar to prior work [15, 60], we implemented a custom crawler that collected review data for a given app from both the Apple App and Google Play stores. The software adhered to the "ethical crawling" framework [57] by only downloading data from websites that permitted crawling according to the site *robots.txt* file, and by reducing the number of requests to minimize website traffic and prevent exorbitant operator costs. Besides reviews for any given application, the crawler also retrieved metadata associated with every collected review. This metadata includes the rating of the review, the date it was posted, and store-specific information including the number of times a review was rated as helpful in the Google Play Store and the review title in the Apple App Store. Overall, we collected 45,821 reviews for the five chosen mobile wallets.

3.1.3 Basic Data Exclusion. As with app reviews in general, our dataset had noisy data that reduced the overall quality [9, 22]. Guzman and Maalej [22] suggest that short reviews are often non-informative and might only include praise or dispraise. To improve the overall quality of our data set (similarly to McIlroy et al. [45]), we excluded those reviews that had only three or fewer words.

While we only crawled reviews from the English versions of the Apple and Google Stores, some reviews turned out to be not written in English. To further reduce possible noise, we used the Google Cloud Translate API[8] to identify non-English reviews. We also excluded these from the final corpus. This reduced the review corpus further, to 26,109.

3.2 Data Selection

Prior work has shown that only a subset of all reviews is relevant to UX. The number of relevant reviews, however, can range and depends on the platform and application area. For example, Hedegaard and Simonsen [27] analyzed reviews on a dedicated review site epinions.com and found that up to 49% of all reviews contain information relevant to UX. A study of reviews in the App Store [49], on the other hand, only classified one-third as related to UX.

While these numbers differ significantly, they both suggest that only some reviews seem to be relevant to UX, whereas others are not. Therefore, our goal was to identify such relevant reviews for manual analysis from the selected corpus of 26,109 reviews. To assess how many reviews are relevant in our corpus, we manually labeled a random sample of 1,000 and found that only 37.5% of reviews were relevant to UX. Because of the high rate of irrelevant reviews, we sought ways to focus our efforts on analyzing relevant samples. To this end, we employed a combination of NLP and ML techniques and used the 1,000 reviews as a training set (Section 3.2.3). The following subsections explain our automated classification approach for the identification of UX-relevant reviews.

3.2.1 Pre-processing. Prior to building our machine learning model, we pre-processed the review corpus to further reduce noise and lower the number of superficial features. We first *expanded contractions* in each review and removed emoticons, numbers, and other symbols. Next, we removed stop words, i.e., common words that

[1] Coinbase: www.coinbase.com
[2] Binance: www.binance.com
[3] Blockchain Wallet: https://www.blockchain.com/wallet
[4] Trust Crypto Wallet: https://trustwallet.com/
[5] Electroneum: https://electroneum.com/
[6] Coinomi Wallet: https://www.coinomi.com/en/
[7] BRD Bitcoin Wallet: https://brd.com/

[8] Google Cloud Translate: https://cloud.google.com/translate

Table 1: Features of a review

Raw Review	Pre-processed Review	Character Count	Rating	Sentiments
Horrible app! I haven't received my bitcoin money after a month of purchasing it. I emailed Blockchain and no one has gotten back to me yet. $200 down the drain! Thanks for nothing!	['horrible', 'receive', 'money', 'month', 'purchase', 'email', 'drain', 'thanks']	49	1	$neg: 0.18$ $neu: 0.74$ $pos: 0.08$ $compound: -0.5684$

Table 2: Examples of tagged reviews

Class	Review Text	Explanation
Relevant to Cryptocurrency UX	Takes at least 3 hours for the bitcoins to arrive in my bread wallet	User mentions dissatisfaction with bitcoin transactions being slow.
Relevant to General UX	The application does not open after upgrade! iPad Air 2 iOS 8.4	This review is not specific to cryptocurrency wallets, but to apps in general.
Irrelevant to UX	The future of crypto instant payment!	No relevance to the app or UX.

rarely add any meaningful information for the purpose of automated text classification. Such words include articles, pronouns, prepositions, and in our case also the names of the apps, words such as "exchange," "blockchain," and names and abbreviations of cryptocurrencies. Lastly, we tokenized the filtered reviews and tagged every word according to its *part of speech*. For both steps, we used the Natural Language Toolkit (NLTK)[9] and only included nouns, verbs, adjectives, and adverbs. Based on these tags, we then *lemmatized* each word to reduce different forms of the same word to a common basic lemma.

3.2.2 Feature Selection. Similar to past studies on review classification [22, 27, 40, 47], we used the *bag-of-words* (BoW) model to represent the reviews. In BoW, all unique words in the text corpus are added to a vocabulary, and each document (review in our case) is represented as a vector. This vector contains the word occurrences for the respective document, and the length of this vector is equal to the vocabulary size [44]. We further employed the "term frequency – inverse document frequency" as the weighting scheme, which decreases the weight of words that occur frequently in many documents, and increases the weight of words that occur frequently in a single document. We excluded words that appeared in more than 10% of the documents, to eliminate very common words. This threshold produced the best results during our experiments.

Occurrences of word sequences, instead of single words, can also be used as features. We have experimented with both character and word n-grams and determined that word bi- and tri-grams ($n = 2, 3$) yield the best results for our corpus and task at hand. For example, some occurring n-grams in our corpus are *"zero balance" ($N = 607$)*, *"waste time wallet" ($N = 589$)*, and *"wallet scam" ($N = 569$)*.

We also treated metadata as features in our classification. Inspired by the findings of Pagano and Maalej [49], we have included both the rating (1 to 5) as well as the length of each pre-processed review in characters.

[9]Natural Language Toolkit: https://www.nltk.org/

Lastly, we also considered the associated sentiment for each review as a feature. Prior work has shown that sentiment scores can be effectively used to classify reviews containing UX information [22, 40]. We employed VADER (Valence Aware Dictionary for Sentiment Reasoning) [31], a tool for sentiment analysis of social media texts. Texts on social media are often short and sparse, which also holds true in the case of app reviews.

VADER is a rule-based sentiment lexicon. Texts are assigned *positive*, *neutral*, and *negative* scores that add up to 1. These values can also be represented as the *compound* score, which is computed by summing the scores for each word and then normalizing the sum to be in the interval from -1 (most negative) to +1 (most positive). A score between -1 and -0.05 indicates a negative sentiment, a score higher than -0.05 and lower than 0.05 a neutral sentiment, and a score between 0.05 and 1 indicates a positive sentiment [31]. Table 1 shows a review with all metadata. We have experimented with both one (compound) and two sentiment scores (positive and negative), and the classification results with different sets of features can be found in the supplemental material (Table B.1).

3.2.3 Training Set. After defining the feature set, we proceeded with training the classifier. We randomly selected 1,000 reviews (500 for Android and 500 for iOS) from the set of 26,109 and manually classified them based on their relevance to UX. To ensure consistency, two researchers independently classified the 1,000 reviews and followed a coding guide (Section A in the supplemental material) throughout the process. UX relevance was assessed based on definitions used in prior work [1, 25, 29, 32, 51] and a review was considered relevant only if it contained information about perceptions and experiences unique to wallets (and its features) or characteristics of cryptocurrencies. During the manual classification process, we classified the reviews into three groups: *relevant to cryptocurrency UX*, *relevant to general UX*, and *irrelevant to UX*.

We followed an iterative approach when manually classifying the 1,000 reviews. We split the reviews into three sets ($n_1 = 300, n_2 = 300, n_3 = 400$) and classified each set independently. After each set, Cohen's kappa was calculated to assess the inter-rater reliability

Table 3: Number of classified/analyzed reviews per wallet and platform. Numbers from the training set are in parentheses.

	Automatically Classified Reviews			Analyzed Reviews		
	Play Store	App Store	Wallet Total	Play Store	App Store	Wallet Total
Blockchain Wallet	2,309	257	2,566	353 *(56)*	344 *(87)*	697 *(143)*
Electroneum	1,659	26	1,685	329 *(32)*	35 *(9)*	364 *(41)*
Coinomi Wallet	1,001	23	1,024	327 *(30)*	31 *(8)*	358 *(38)*
Trust Crypto Wallet	468	59	527	310 *(13)*	79 *(20)*	389 *(33)*
BRD Bitcoin Wallet	385	297	682	314 *(17)*	400 *(103)*	714 *(120)*
Store Total	5,822	662	6,484	1,633 *(148)*	889 *(227)*	2,522 *(375)*

(IRR) and disagreements, which were caused by the different levels of familiarity of the coders with cryptocurrencies and crypto wallets, were discussed and resolved. For n_1, IRR was $\kappa = 0.61$, for n_2, we achieved $\kappa = 0.62$, and for n_3, a value of $\kappa = 0.65$. All of these values indicate a substantial agreement among coders [39]. Table 2 shows examples of the three kinds of reviews that we encountered during the process.

The manual classification occurred on a review-based level, meaning that if a review had parts that were both relevant and irrelevant, it was classified as relevant. This classification is coarse; however, it was acceptable in our case, as it is only an intermediate step in our approach. This coarse classification might also be the reason for the high percentage of reviews classified as relevant to UX. Out of the 1,000 reviews, we classified 375 as *relevant to cryptocurrency UX*, 557 as *relevant to general UX*, and 68 as *irrelevant to UX*. For our study, we were only interested in the UX issues particular to mobile wallets, and we therefore considered reviews containing only general UX comments as irrelevant. Overall, we manually classified 375 reviews as relevant to mobile wallet UX and the rest (625) as irrelevant. Cohen's kappa for these two classes of reviews was $\kappa = 0.69$.

3.2.4 Machine Learning Model. After finalizing the training set, we selected random forests (RF) as our machine learning model. We have compared several models that are known to perform well for short text [52] and, for our data, RF performed the best (the comparison between models can be found in Section B in the supplemental material). The combination of sentiment scores, BoW model, and metadata (length + rating) had an f1 score of 0.78 for the positive class (relevant UX reviews) and is comparable to prior work on binary classification of reviews relevant to UX [22, 27, 40].

3.2.5 Validation. We have also evaluated the performance of our binary classifier by measuring the area under the curve (AUC) for the receiver operating characteristic (ROC) [24]. The ROC curve is not sensitive to imbalanced data [26] and is therefore appropriate for evaluating our classification approach. A perfect classifier would have a corresponding AUC-ROC value of 1, whereas a binary classifier that operates by chance would have a value of 0.5. When we performed a 10-fold cross validation, our classifier achieved a mean AUC value of 0.90, which, according to Hosmer et al. [30], is indicative of an outstanding discrimination. The ROC curves can be found in Figure B.1 in the supplemental material.

3.3 Data Analysis

The final step in our approach was the qualitative analysis. From the 25,109 reviews, our classifier identified 6,484 (25.8%) as relevant to mobile wallet UX (see Table 3). Combined with the 375 relevant reviews from the training set, the overall number was 6,859. Similar to the Thematic Analysis approach described by Guest et al. [21], the lead analyst created an initial codebook. In our case, this resulted in 121 codes from the 375 relevant reviews of the training set.

We followed an iterative coding approach with two researchers. We selected the wallet with the lowest number of relevant reviews and randomly sampled the same number from each other wallet. As such, we selected 230 reviews (10 ∗ 23 (# reviews for Coinomi iOS) = 230), and split them randomly between the two analysts. These reviews were split into equal-sized batches and were coded sequentially, i.e., one analyst waited until the other finished coding and then used the updated codebook to code their batch. We have not calculated the IRR, as it is suggested that in Thematic Analysis such measures do not indicate objective coding, but merely the fact that coders are able to code in the same subjective manner [59]. Instead, after every coding round, three researchers together discussed the codes, created code groups, and identified themes. Overall, after 17 review batches, 325 codes, 26 code groups, and five themes were identified. Out of the 2,522 classified reviews that we have coded, only 64 (~ 2.5%) were coded as false positives, which again is an indicator of the high accuracy of our classifier.

We stopped the analysis once it became clear that we had reached thematic saturation [21]. This occurred after 1,285 reviews. To ensure that this was truly the case, we coded another 1,237 reviews, and while there were new codes, no new themes were identified. The saturation graph can be found in Figure C.1 in the supplemental material.

4 RESULTS

In the following sections, we first briefly describe the metadata of the reviews that were classified as relevant by our approach and then present each identified theme. We include illustrative reviews where appropriate.

4.1 Review Corpus

The reviews that were classified as relevant to UX are longer and more detailed than the irrelevant reviews. They have more than twice as many words, with a standard deviation more than twice as large. The shortest relevant review is four words long and the longest 527 words. In Table 4, the relevant reviews are compared

Table 4: Metadata for different types of reviews

		Word Count		Character Count		Rating		Sentiment		
Review Type		M	SD	M	SD	M	SD	Neg	Neu	Pos
Relevant ($N = 6,859$)		33.12	28.24	178.13	153.3	2.36	1.52	17.5%	38.0%	44.5%
	Android ($N = 5,970$)	31.05	24.82	166.78	134.25	2.34	1.50	18.3%	38.5%	43.2%
	iOS ($N = 889$)	46.98	42.36	254.34	231.65	2.49	1.66	12.3%	34.5%	53.2%
Irrelevant ($N = 19,250$)		12.95	13.45	69.39	73.46	4.23	1.37	10.1%	14.3%	75.6%
All ($N = 26,109$)		18.25	20.53	97.96	111.54	3.74	1.63	15.2%	17.4%	67.4%

against the irrelevant ones. Other metadata, such as review age, can be found in Section E in the supplemental material.

We performed an independent samples t-test on the lengths and ratings of the reviews. The results show that both the difference in length ($t(7992.5) = 56.79, p < 0.001$) and rating ($t(11020) = -89.47, p < 0.001$) between relevant and irrelevant reviews are statistically significant. The worse rating of relevant reviews is also reflected in the associated sentiments. The difference in negative sentiments between the relevant (M = 0.09, SD = 0.01) and irrelevant reviews (M = 0.04, SD = 0.01) was also statistically significant ($t(12162) = 35.17, p < 0.001$). Less than half (44.5%) of the relevant reviews have a positive sentiment, compared to the 75.6% of the irrelevant ones. The analysis was conducted using VADER [31], and discrete ratings were based on the compound score.

4.2 General UX Issues

Some reviews complained about shortcomings that are not unique to mobile wallets. Reviewers reported performance issues, including freezes and crashes that led to the inability to access funds in the worst case, e.g., *"[redacted wallet] is crashing every time I try to access it. Sometimes I can enter a number or two for my PIN before it crashes, but otherwise I cannot access my wallet at all [...]" (R6758).*

App updates also often resulted in negative effects for reviewers. Instances where an update led to a loss of functionality or access were reported for all wallets. In all such cases, reviewers appeared to be emotionally distressed, as illustrated in the following review: *"My app worked perfectly fine till you 'updated it' to 'fix crashes' now i can't even get the money out of my wallet. FIX THE APP ITS BEEN A WEEK ALREADY" (R522).* Some of these updates also occurred automatically and surprised some reviewers: *"[...] updating an app for security is not the same as completely changing what people get used to out of nowhere. being blindsided is not good [...]" (R1359).*

Common interface problems were also found in the reviews. Some issues were related to color schemes in the wallet interface, typos, or an inconsistent use of icons and were arguably merely an inconvenience, whereas others resulted in a direct functionality loss. Some examples for the latter were overlays that prevented reviewers from creating wallets or making transactions: *"[...] the app would not position the screen (lay out) properly...preventing me from touching the (APPROVE) button. Preventing the successful completion of the transaction. THIS APP IS FREEKING USELESS!!!!!!" (R440).* In this case, the reviewer did not lose cryptocurrency; however, there were also cases where poor interface design led to monetary losses. One example for this was double-spends, where reviewers sent a transaction multiple times by mistake. Because all transactions are irreversible by design, it resulted in a direct loss: *"This app is the worst. It just made me double send all my [cryptocurrency] by not reacting to the Send button properly. Goodbye [redacted wallet], I trust you no more [...]" (R1086).*

Therefore, it appears that while the abovementioned issues are not unique to mobile wallets, they become more severe whenever money is at stake.

4.3 Domain-Specific UX Issues

We identified a variety of shortcomings unique to mobile wallets and present them in three groups: *before*, *during*, and *after* use.

4.3.1 Before Use. The wallet initialization process was found tedious by some reviewers. Some of the wallets that we investigated supported the option to purchase cryptocurrencies and therefore required reviewers to go through an identity verification procedure (also called know your customer) to prevent money laundering and fraud. This process, however, was reported to be cumbersome and sometimes involved downloading another application: *"Now I have download another app that is called Yoti. I have to put in MORE information about myself. I have to always take a selfie for any and every transaction. I don't want to have to use another app just to use another app [...]" (R1455).* Even when downloading another app was not necessary and the verification was done in the respective wallet, it could often take up to multiple weeks to get verified, which resulted in reviewers getting frustrated: *"Since 4 months...Still showing verification under process...Your information is being reviewed...Am I going to be verified?" (R4105).* This is in line with the findings of Voskobojnikov et al. [61], who identified the verification process as one of the entry barriers.

Those reviewers who were able to get through the verification process also faced challenges. Reviews mentioned a lack of guidance during the setup that made it challenging to create a wallet in the first place. This was mostly reported by reviewers who considered themselves to be novices: *"I've been trying to learn about buying crypto, and even as an IT professional I've been a bit intimidated and confused. It took me several days of trying figure out how exactly to get money into the wallet [...]" (R382).* This lack of guidance was also prevalent for specific functionalities of the wallet, such as the private key import/export: *"[...] Tutorials on how to import export would be nice and I'm a little nervous to try it out [...]" (R567).*

4.3.2 During Use. Reviewers were unaware of which cryptocurrencies were supported by the wallet they were using. In the majority of cases when reviewers had installed a wallet without realizing that it did not support their particular cryptocurrency, reviewers

requested the support in a future version. However, there were also instances where reviewers sent an unsupported cryptocurrency to their wallet, resulting in it getting stuck: *"I sent ICN tokens (300 of them) into my iconomi [...] wallet, and now [...] I can't send them out as it tells me it's an unambiguous address [...] but i need erc20*[10] *functionality not just storage" (R6334).*

Irrespective of the cryptocurrency, reviewers also reported incorrect balances being displayed in their wallets. Some reviewers verified the transaction status with the help of blockchain explorers, such as Etherscan[11] or blockchain.com,[12] and found a mismatch between the actual status on the blockchain and the status in the wallet: *"We received a Bitcoin Cash transaction into this wallet 2 days ago, and the official blockchain shows that this transaction has received 76 confirmations. However, this app refuses to show the transaction as confirmed. It shows the transaction as pending with zero confirmations [...]" (R732).* Pending transactions were also often misunderstood, as illustrated by the following review: *"But it is still pending showing no confirmation and I am not able to transfer my ether to anybody these people have totally hijacked my wallet and I am on their mercy now" (R5542).* In this case, the user believed that by not being able to send their funds, a malicious party must have gained access to their wallet. In reality, however, pending transactions refer to transactions that have not been processed by the network yet.

One prominent reason for a transaction to stay in the pending state is the transaction fee (set by the user who initiates the transaction) being too small. For cryptocurrency transactions, transaction fees can be set by the user and have a direct influence on how quickly a transaction is processed by the network – the higher the transaction fee, the more quickly a transaction is included in a block. Reviewers reported limited possibilities when it came to setting these fees in their wallets: *"Recommended fees are crazy one is too high the other is so low it would never process [...] I should have more than two fee options and should be allowed to set my own. Your [economy] fee would never get picked up by miners and your standard fee is crazy [...]" (R408).* The main problem with inadequate fees is that even if a transaction fails, the fees are not refunded and this might result in a monetary loss: *"I've lost 40$ to unconfirmed transactions, my friends have lost 50$,70$, and 25$ [...]" (R873).*

High transaction fees were also caused by the apps' failure to support upgrades in the blockchain protocols. In the case of Bitcoin, one feature could have lowered the fees as explained in the following review: *"The fee for a bitcoin transaction on this app are ridiculous high, they have yet to add segwit*[13] *which will lower the fees and allow faster transaction times. If they would incorporate [segwit] these problems would not happen. I am very annoyed [...]" (R702).*

Moreover, the lack of transparency with regard to blockchain upgrades also resulted in functionality loss, with some reviewers not being able to send their transactions because of incompatible addresses: *"do you guys support the new bech32*[14] *format to send, keep getting incorrect bitcoin address error [...]" (R889).*

Inadequate transaction fees, however, were not the only problem, as some reviewers did not even know how to make transactions in the first place. Similar to the setup phase, some reviews mentioned an overall lack of guidance: *"I'm new at this but how do you add money to your wallet" (R608).* This lack of guidance was not only prevalent for transactions, but also for the recovery process. To recover a wallet, users need to input a 12-word seed phrase that was generated during the setup stage, and without such a phrase, a recovery is impossible. Some reviewers explained they had lost their phrases, whereas others did not understand how to use them: *"How can I recover my blockchain app [in case] I mistakenly delete it?" (R4127).* Reviewers also mentioned having difficulties with recovering access to their wallets in the case of new phones, or whenever they lost access otherwise: *"I recently uploaded some funds and now I can't even access them because [it keeps] crashing I don't want to delete and reinstall because I've had this wallet for over a year and I don't [have] the [seed phrase] anymore [...]" (R405).*

4.3.3 After Use. Some of the wallets required creating an account with the wallet provider, which some reviewers wanted to delete. In all such cases, however, they were not able to do so: *"I'm giving the 1 star would give it 0 if I could, I don't see any way to delete my account I don't use this anymore and trying to delete it so if anybody knows how to let me [know]" (R190).*

4.4 Misconceptions of Users

Some shortcomings can be attributed to reviewers' misconceptions of blockchain characteristics and cryptocurrency building blocks, such as transactions. Reviewers attributed high fees, which were explained in Section 4.3.2, to wallet providers: *"I had around $30 worth of Bitcoin in my wallet. As I went to transfer the coin, I was charged a $10 transaction fee. That's 1/3 of my balance. Not only that, but I was charged a fee when I had my Bitcoin sent to Bread. I don't want to call them criminals or crooks because that's extreme. I just want to let everyone know how much in fees you will be charged if you use this service. Never again!" (R402).* Contrary to the reviewer's belief that the wallet provider is setting the transaction fees, the minimum required fees are determined by the network congestion and transaction size, and are received by the miners [3].

The fee structure was also not clear to some reviewers. To transfer tokens that are based on Ethereum, users have to pay the transaction fee in its currency, referred as Ether (often abbreviated as "ETH"), which was unclear to some: *"tries charging you ETH to make non-ETH trades and on top of that won't tell you how much ETH is even required. They are straight up extorting money from users! Avoid at all costs until they fix this and give us our money back [...]" (R3051).* Even after the developer responded to the review and explained the fee structure, the reviewer was still convinced they were correct: *"No exchange charges you coins that are not part of the 2 currencies being exchanged. Period. Fix your broken platform that locks people out of their $$."* This hints at the reviewer's belief that cryptocurrency exchanges operate in the same way as mobile wallets, which is not the case. Trades on a cryptocurrency exchange are happening

[10] ERC-20 is the standard for tokens on Ethereum: https://eips.ethereum.org/EIPS/eip-20

[11] Ethereum blockchain explorer: https://etherscan.io/

[12] Bitcoin blockchain explorer: https://www.blockchain.com/explorer

[13] Segregated Witness (SegWit) was a protocol proposal improvement that, among other benefits, resulted in lower transaction fees: https://github.com/bitcoin/bips/blob/master/bip-0141.mediawiki

[14] Bech32 is a standard for wallet addresses that support SegWit: https://github.com/bitcoin/bips/blob/master/bip-0173.mediawiki

off-chain, i.e., are not incorporated into the blockchain, whereas transactions that are initiated from a mobile wallet are.

Reviewers were also confused about addresses. Some believed that their funds were tied to the mobile app: *"This app has deleted all my funds and account. DO NOT DOWNLOAD!!!" (R884).* Such reviews hint at a misconception, as wallet addresses are immutable by design, with wallets being merely an interface to manage keys. Others believed that addresses were static and would not change, which is not the case for hierarchically deterministic wallets.[15] These wallets generate new addresses after every incoming payment, and in some instances, this led reviewers to believe that monetary losses have occurred: *"My Blockchain address was changed. I work for a company so they usually pay me in bitcoin every Friday. I usually receive the current weeks payment the next week. But due to the fact that my address was changed I couldn't see my payment on my wallet. I was shocked when I went to check and find out that my wallet address has been changed and my money for the week was lost" (R713).*

Some reviewers also wanted to cancel transactions, which the design of the underlying blockchain does not allow. For example, this was reported in the context of transactions that were sent to a wrong address: *"[...] if i unfortunately sent any token on wrong [address] there [is] no cancel option! Add cancel option" (R928).* A cancellation option was often requested, with some reviews mentioning conventional banking as an example: *"I never had an issue with Blockchain unlike Breadwallet but I [just] wish it was a way to stop payment [...] like you do a stop payment on a checking account" (R853).* Recovery alternatives that were known from other payment systems were also requested: *"No way to restore your data and money once you lose your [seed] phrase. Please enter some other way of recovery like email backup" (R6271).* Such recovery methods do not exist for non-custodial wallets, and using conventional systems as a reference point when dealing with cryptocurrencies could lead to errors and losses.

The comparison to conventional systems was also found in other contexts. Some reviewers were surprised by the transaction fees: *"Not user friendly. Always wanting to get some sort of fee. I can't transfer funds without a cost. Just want to cash out and delete" (R3106).* Others were taken by surprise when experienced fluctuations of the stored value, when expressed in fiat currency: *"I'm still new to the world of crypto currency, I thought this App is for storing value. But it looks my balance fluctuates with the Bitcoin price. I really need an urgent help before I lose all of my money" (R3096).*

A prevalent belief was also that the wallet providers have an influence on the transactions. Pending transactions were one example where reviewers asked for help from the developers: *"[...] Been 5 DAYS NOT ONE CONFIRMATION. I DONT CARE IF ITS THE MINERS THAT CONFIRM THE TRANSACTION THIS IS THEIR APP AND SHOULD HAVE SOME TYPE OF PROTOCOL FOR THIS [...]" (R90).* In case of failed transactions and lost fees, reviewers also expected help from the support team: *"Tried to exchange currency once it failed and I lost all of the [associated] fees, no help from support as to why it failed, I'll be using a different wallet from now on" (R5654).*

Support was also wanted in case of incorrectly displayed balances in the wallet. Some of these mismatching balances, however, were caused by syncing problems with the blockchain. Mobile wallets, and cryptocurrency wallets in general, need to sync with the actual blockchain to display the newest transactions. Some reviewers did not know the purpose of the sync, whereas others were impatient and blamed the developer for coins that were not displayed: *"I do not recommend this app, they will steal your bitcoins and say that your wallet is out of sync, when you try to sync it, it just says the same thing. WALLET is a complete rip-off !!!!" (R3271).*

The majority of issues were believed to be the developer's fault, whereas in reality, the wallet had no influence whatsoever, as illustrated in the following review: *"Uninformed people are writing one star reviews because of high transaction fees and delays, when in reality [the wallet] has nothing to do with any of these issues" (R443).*

4.5 Security and Privacy

Issues related to security and privacy were brought up in app reviews. Mobile wallets allow anyone with the possession of the phone to transfer funds, and reviewers expressed their concerns when it came to access to the app: *"This app is working very well except it opens straight into your send/receive function. A PIN would be nice to ensure only I can open the app" (R854).* PIN codes, however, were viewed by some to be inferior to other authentication methods, such as biometrics: *"I was hoping they would finally include fingerprint authentication, but they still use pin and have a more dysfunctional app [...]" (R844).* What's clear from our data is that reviewers were uncomfortable about their mobile wallets being accessible to anyone who opens them.

Measures to ensure that only the actual owners were able to send funds were also requested. For example, reviewers called for secondary protective measures, such as two-factor authentication: *"[...] security could be better (I'd like option to turn on 2FA when transacting on mobile app)" (R4497).* Additional layers of access control, however, were not always welcomed, as illustrated by the following review: *"Forcing a screen lock = not cool. I keep my phone with me at all times, just like my physical wallet, and I don't want to turn on an annoying screen lock just for this app [...] Look, I get it, most people need to have their hand held through the process of securing their stuff, and adding an extra layer of security is covering your own behinds. [...] Just please don't force overbearing (and redundant) security features on users who don't want or need them" (R3333).*

Reviewers were also worried about the security of their private keys. Wallets that allowed access to private keys were perceived to be more secure: *"This is only iOS wallet that puts you in control of your private keys (that I know of). Which makes it (potentially) one of the most secure mobile wallets" (R383).* At the same time this potential single point of failure was also found worrisome, with some reviewers losing funds because of a stolen phone or a lost seed phrase: *"Some has stolen my phone and my blockchain wallet is inside the phone, I have so much money inside the bitcoin wallet but I've forgotten my password" (R720).*

Physical safety was also considered to be a risk by some reviewers. This again relates to the fact that there is no centralized entity or safety net and in the case of losses, the funds are gone for good. Having a balance displayed in the wallet, for example, was found risky by some: *"[the wallet] should not display the accounts Total at*

[15] Deterministic wallets: https://en.bitcoin.it/wiki/Deterministic_wallet

first screen. Make you vulnerable to a $5 wrench[16] *attack" (R6742).* Similarly, having to enter the PIN to accept transactions was seen as a physical risk as well: *"I have a background in law enforcement [...] I've investigated a quite a few robberies during my career and know how the future of robberies will happen [...] If I meet someone from craigslist to sell something [...] I have to enter my PIN to share my public key. So, let's say I'm meeting a stranger and now, I enter my pin and hold it up for them to scan... And WHAM! They snatch the phone out of my hand, run off with my phone and can now steal my bits cause I entered my pin for the bad guy. We don't need a PIN number to protect us from receiving money [...]" (R552).*

Lastly, privacy concerns were also found in the reviews. As previously mentioned, reviewers had to undergo a know your customer verification process to be able to purchase cryptocurrency and found the sharing of personal data alarming: *"Need to disclose too much personal information [...] It just doesnt make any sense [...]" (R1837).* Deleting such personal information was found impossible, which resulted in distress for some: *"I am 17 years old and I'm trying to delete this app after an elderly man tried to trick me into opening an account. I'm scared because my Id information is on this app and I cannot delete it...can anyone help me?" (R770).* Questionable app permissions, such as the access to the contact list, microphone, or camera access also raised privacy concerns: *"Why do you need access to my microphone?? Can you [please] cite the international regulation you mentioned? What are 'localisation improvements?' Thanks. WHAT ARE LOCALISATION IMPROVEMENTS?? CITE THE INTERNATIONAL REGULATIONS???" (R99).* This is in line with prior work that identified the prevalence of overprivileged permissions in Android apps [17].

4.6 Trust

Reviews also included comments on factors that led reviewers to trust or distrust the wallets and their developers. Some reviewers explained they were using different types of wallets depending on the amount: *"If you're like me, you keep most of your BTC offline, or near offline, and transfer a little to [the Blockchain wallet] for some spending money [...]" (R836).* Others explicitly mentioned only trusting mobile wallets with small amounts, as illustrated in the following review: *"Just like any other wallet on a phone don't put all your funds here. It is just for small spending and transfers" (R6745).*

Some wallets were open source, which appeared to have an influence on trust. Reviewers valued the ability for the public to validate the source code: *"[...] the code is available to audit. It super important to feel safe using the app and this is the only one I trust" (R684).* However, whenever a wallet was closed source, such as the Coinomi wallet, some reviewers appeared to be hesitant: *"Coinomi's website lies by saying 'source-available.' They went close-source approx one year ago. This is not trustless. Who knows what they are now doing with your private keys" (R6319).* Bad publicity also appeared to influence the trust negatively: *"[...] the main reasons I don't trust this wallet is the defensive nature [that] the [developers] take towards negative comments, the fact that they leaked seeds [and] tried to threaten to cover it up [...]" (R5780).* This incident is well documented and was reported by multiple blockchain news outlets [38, 56].

[16] A hypothetical scenario where an attacker can physically force someone to give up information instead of breaking the underlying cryptography (https://xkcd.com/538/)

Unsatisfying UX also led reviewers to question the motives of the developers and their apps. For example, missing transactions made reviewers believe that the wallet was a scam: *"[redacted wallet name] is a scam. The ceo needs to be arrested. I sent [cryptocurrency] to Coinbase over a month ago and they have not arrive yet. This is one thing the president needs to look into, is not letting blockchain have any dealings in the USA" (R6660).* Various other reasons, such as the lack of a quick response from the customer service, high transaction fees, or even the volatility also led to some reviewers concluding that the app must be fraudulent: *"Beware of this scam application, it [is] deducting my money, from $155 to $89. What is deducting money from the account without any transaction. This app is totally scam scam scam scam" (R5569).*

Other prevalent issues that were presented in the previous sections also contributed to reviewers distrusting the apps. The inability to access the wallet or pending transactions, of which some might be attributed to unsynced wallets, were believed to be indicators of scams: *"Transferred a small amount from coinbase to this 12 hours ago, still nothing showing in the wallet. Has been confirmed sent but nothing on this end yet. [...] Think this is a scam of some kind" (R6071).* Even the most trivial things, such as accepting the terms and conditions, made some reviewers question the legitimacy of the wallet: *"Would like allow me to accept terms and conditions. Wondering if it is a scam" (R6796).*

4.7 Theme Prevalence

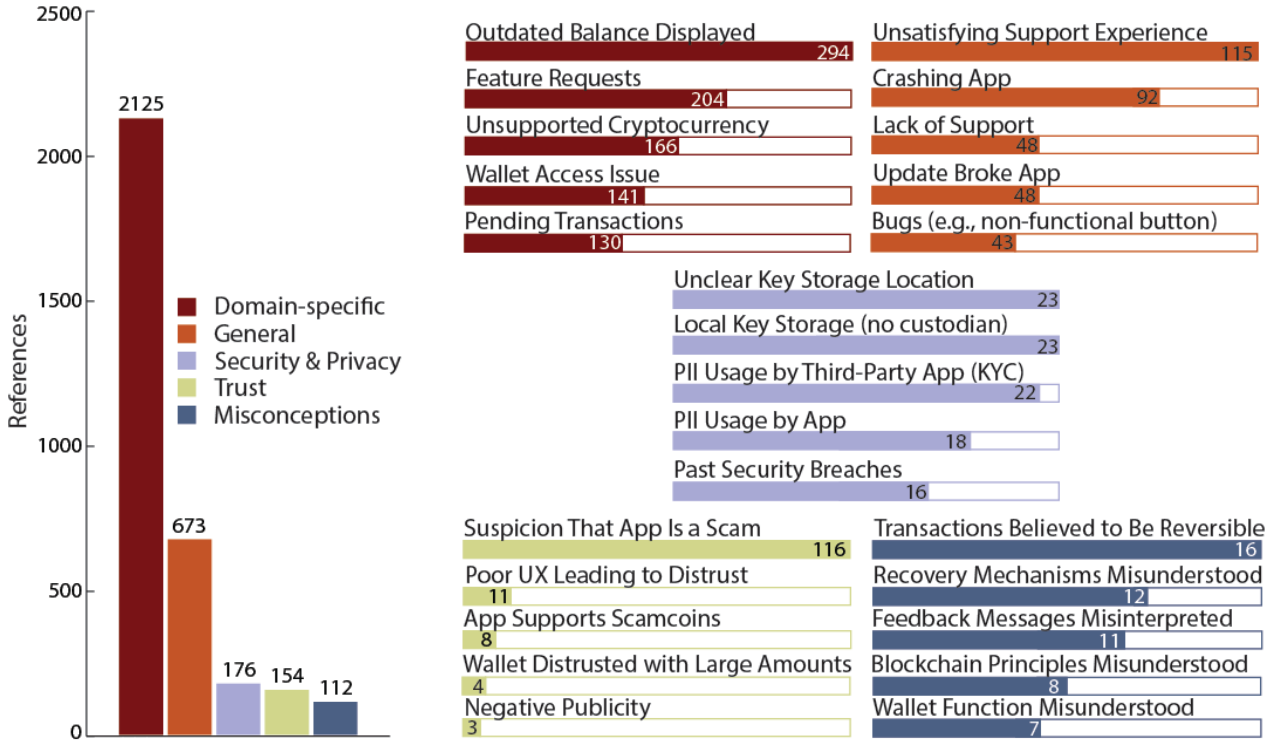
Domain-specific issues were the most prevalent in our analysis. With 2,125 referenced review segments, these issues were more than three times as common as the next theme, i.e., general issues (see Figure 2(a)). Among domain-specific issues, outdated balances were the most prevalent, which, together with pending transactions, were found in over 400 reviews. For general issues, reviewers reported a poor (or nonexistent) support experience and further mentioned crashes and freezes that made it impossible for them to use the application in the first place. The other three themes were encountered less often (Figure 2(b)) but also appeared to influence the overall UX as outlined in Sections 4.4, 4.5, and 4.6.

4.8 Praised Features of Non-custodial Wallets

While the overwhelming majority of review segments were negative ($N = 3,104, \sim 96\%$), we also identified aspects about the wallets that were praised by the reviewers ($N = 136, \sim 4\%$). Pending transactions and outdated balances were among the most prevalent issues in the review corpus (see Figure 2(b)), which, as mentioned in Section 4.3, were sometimes caused by fixed transaction fees. Therefore, it is not surprising that some reviewers found customizable fees beneficial: *"I get to adjust the fees however I please, which has saved me a TON OF TIME AND MONEY when it comes to Price & Trx Spikes so i avoided being stuck on the blockchain [...]" (R6668).*

Security features were also welcomed by reviewers. Some found it more secure that the app would lock itself after a time period, whereas others praised additional authentication measures, such as biometrics or PINs: *"Simple and easy way to send and receive bitcoin. Supports Touch ID and has a pin. I really love this app" (R857).*

Reviewers also valued the non-custodial property of mobile wallets, as illustrated in the following review: *"[...] most other wallets*



(a) Number of referenced review segments per theme

(b) Five most frequent codes per theme (PII: personally identifiable information)

Figure 2: Theme statistics

are not wallets at all, they are just online banking apps for BTC. Not this! This is an actual wallet where you control the private keys. There is no central server, so you are responsible for your own security (this is a good thing) [...]" (R142). Similarly, the ability to import private keys into a wallet allowed users to seamlessly access their funds, leading to a better UX: *"Thanks so much! Had a few coins stuck in a wallet and instead of waiting for the wallet to sync I just imported my private key into [redacted wallet name]" (R6445).* Such features could therefore have the potential to prevent access issues and stuck funds, which were both mentioned in hundreds of reviews.

5 DISCUSSION

We have successfully applied a novel combination of automated review classification and manual, qualitative analysis to identify UX issues of mobile wallets. This hybrid methodology saved a considerable amount of time, as most reviews that were included in the qualitative analysis (> 97.5%) were indeed relevant, compared to only 37.5% that were found relevant in a random sample. We identified various issues that, while wide-ranging, can be assigned to one of two groups: *general* or *domain-specific*. All investigated wallets had these shortcomings, some of which resulted in dangerous errors or monetary losses for the reviewers. Our main contribution is the identification of previously unreported UX issues, both on the interface and conceptual levels, and their effects. In this section, we categorize these UX issues, discuss their implications, and outline possible design improvements that could alleviate them.

5.1 General UX Issues

Unsurprisingly, the investigated wallets had issues that are commonly found in mobile apps in general, both on Android and iOS. For example, reliability issues, such as crashes or freezes, are widely reported in the literature [22, 23, 34], yet their prevalence and impact in mobile wallets is unknown. For all mobile wallets that we investigated, the private keys are stored locally on the mobile device and are managed with the help of the app. For all apps, however, reviewers reported to have lost access to their funds because of apps crashing, slowing down, or not opening after new updates. Non-custodial wallets therefore present a unique case where user interface (UI) issues that would be harmless in many apps can have a disastrous impact on the UX and can lead to monetary losses.

To address this, wallet developers can clearly specify how to recover private keys in the case of non-functional applications. Mobile wallets, and non-custodial wallets in general, are interfaces that facilitate key management and are interchangeable as long as they support a particular cryptocurrency. Often, reviewers did not know how to access or recover their assets without the mobile app. Providing more guidance and possibly technology support, e.g., in the form of key import/export, could address this.

We have also encountered reviews mentioning common UI issues that led to a loss of functionality. For example, reviewers reported missing buttons that prevented them from sending transactions or instances where two transactions were sent because of an unresponsive interface. Similar to performance issues, these interface shortcomings can also have grave consequences. Following common usability guidelines or heuristics, such as the ones proposed by Nielsen and Molich [48], can help in reducing or even eliminating general issues, and can contribute to a better UX in the cryptocurrency domain. Here, *error prevention (Heuristic 5)* is particularly important, because of the transaction irreversibility and the lack of a safety net for recovering from user-induced errors.

5.2 Domain-Specific UX Issues

The second group of identified issues is unique to the cryptocurrency domain and needs special attention. Here, we differentiate between issues directly related to the UI and issues that are the result of users' misconceptions regarding cryptocurrencies. Two prominent examples of the former were incorrect wallet balances and unclear transactions statuses, some of which were caused by unsynced wallets. As a result, reviewers believed that losses had occurred, when in reality, the blockchain status displayed by the wallet was simply outdated. We recommend that wallets clearly display the current status of the blockchain and further state that syncing issues might be a reason for "incorrect" balances or "missing" transactions.

Transaction fees were another major problem encountered in reviews. The fees were perceived as either too high or too low, with reviewers rarely finding the fees appropriate. Transactions with low fees might not be processed by the miners at all, whereas users could overpay on transactions with too high fees. Mai et al. [43] further found that the implications of varying fees might be misunderstood by the users and suggest that wallets should provide preset fee options, as it is currently done in the Blockchain wallet and Coinomi. However, as our findings suggest, many reviewers found these options restricting and, at times, inadequate, especially when they could not be changed. We therefore recommend that wallets always allow users to customize their transaction fees based on a recommendation that averages the fees of the recently processed transactions that were included in the blockchain. Two of the five investigated wallets, BRD Bitcoin wallet and Trust Crypto wallet, did not (and still do not, as of December 2020) have a custom fee option, which was poorly received by their users.

Solving these domain-specific issues is not a trivial task, and one option could be the development of usability heuristics specific to the cryptocurrency domain. Prior work has shown that domain-specific heuristics can help in identifying more severe usability issues unique to the domain (e.g., [5, 6, 13, 33, 36, 41]), and we believe that tailored heuristics can do the same for cryptocurrency wallets. These heuristics could not only be used for discount usability testing, but also for informing the design of wallets.

Reviewers also had their fair share of misconceptions about the building blocks of cryptocurrencies that went beyond the interface. The most prevalent misconception of reviewers was that wallet developers are the ones who set and receive the transaction fees. This belief was particularly prevalent in those cases in which reviewers perceived the fees as very high, leading them to assume that the app was fraudulent. Others believed that transactions could also be canceled or reversed, which is also not the case in reality. Prior work has found that users do not understand the implications of varying fees on the processing speed [43], and our findings further suggest that the nature of transactions and fees in general might also be unclear for some. We do not report misconceptions regarding the underlying cryptography, as these are well documented [19, 43, 55, 61] and were confirmed in our study.

It further appears that traditional systems, such as online banking, serve as a reference point for some cryptocurrency users who had these misconceptions. This can be problematic because of the vastly different natures of these two payment systems. Contrary to traditional currencies and systems, which allow the reversal and cancellation of transactions in certain cases [8], it is impossible to do the same with cryptocurrencies. This lack of a safety net makes user errors, such as incorrectly addressed transactions, more costly. Cryptocurrencies also differ in other regards, such as inconsistent fees or dynamic addresses, and therefore clash with conventional payment systems [7].

Our results suggest that some users have these traditional systems in mind when using cryptocurrencies and are surprised when key characteristics or functionalities, such as the cancellation of payments, are not implemented. When using wallets, users also expected some type of recovery mechanisms that would allow them to regain access to their funds in case of lost seed phrases. Cryptocurrency wallets should therefore try to mimic existing payment systems and features that users are already familiar with. Designing such systems is technically feasible, and we give detailed recommendations in Section 5.4 that could help in meeting some of the users' expectations.

Poor UX also led reviewers to question the intentions of the developers. Even the smallest errors or bugs, as well as the lack of customer support, made reviewers believe that the respective application was a scam. This is unsurprising considering the number of fraudulent cryptocurrency exchanges and startups operating in the space [12]. Even in app reviews, we have identified several instances where reviewers tried to scam others, as illustrated in the following example: *"[...] please send some btc to my address I need to help some leprosy patients in my locality in Nigeria. All items and photos will be posted on all trust social network as a way trust is giving back to the society. [redacted bitcoin address]. Thank you in anticipation" (R1242).* Fraudulent activities in the domain are prevalent [58] and make it particularly difficult for users to build trust. Losing trust, on the other hand, is easy and can be caused by common UX issues, such as a crashing app or a prompt asking the user to accept a set of terms and conditions. Therefore, it appears that trust in the cryptocurrency domain not only includes social actors directly participating in transactions as suggested in prior work [54], but also the developers of tools that help users in managing their cryptocurrency. A detailed investigation of factors influencing the choice of tools can be a subject for future studies.

Overall, it appears that the development of mobile wallets is a very difficult and unforgiving task. The developers have to consider not only common UX issues, but also domain-specific ones, both of which can lead to irreversible damages for the respective users. Misconceptions on the users' side make it particularly difficult to

design a wallet with a satisfying UX, as developers are blamed for every little mistake or bug, even though in many cases they can hardly be held responsible. Combined with the already distrustful user base due to the overwhelming number of scams and frauds in the domain, developers are fighting an uphill battle, as they have to address issues that are related to both UI and misunderstandings of technological intricacies of cryptocurrencies. We believe that providing functionality that is known from conventional systems and otherwise clearly stating the differences, such as transactions, addresses, and fees, can prevent future monetary losses and might further contribute to an improved overall UX.

5.3 Limitations

In this study, we have only investigated mobile wallets, and our findings are therefore not necessarily applicable to other wallet types. Similarly, the selected wallets do not support all cryptocurrencies in existence, and it is possible that some UX issues were therefore missed. However, we believe that while features might differ across cryptocurrencies and wallets, the key functionalities stay the same, and those include the management of cryptographic keys. A comparison of UX for different wallet types and cryptocurrencies could be an interesting avenue for future research.

Further, we have used an automated approach to filter non-relevant reviews. While our classifier has high accuracy, it is still possible that valuable reviews highlighting unreported UX issues were not included for manual analysis.

Some of the identified issues might have been version specific and may have been fixed in forthcoming updates. Due to the limited metadata in the app stores, however, we were not able to retrieve the version in which reviewers encountered these issues. Despite this limitation, we believe that the insights from this study provide a sufficiently accurate overview of the UX issues that are encountered in mobile wallets these days, especially since both general and domain-specific issues were found for all wallets and platforms.

5.4 Design Recommendations

5.4.1 Mimic Existing Payment Systems. Some users were surprised when features known from other payment systems were not implemented in mobile wallets. Certain features, such as reversible transactions, cannot be implemented as they clash with the fundamental principles of cryptocurrencies; others, however, can. A prevalent problem that was reported across all wallets in hundreds of reviews was pending transactions. Both Ethereum and Bitcoin allow users to "overwrite" an existing unconfirmed transaction that has not yet been processed by the network. For Bitcoin, this mechanism is called replace-by-fee[17] and allows users to send a modified transaction with a higher fee than the initial, possibly stuck, transaction. The miners would then process the newer transaction more quickly to maximize profits. Sending such transactions is relatively complex, and wallet developers could automate this process. Users could then replace "stuck" transactions with one click and would not have to wait days or even weeks just for a transaction to fail. One caveat of this approach, however, is that the user must pay the fees for both transactions. The wallet UI should therefore clearly communicate this prior to sending a replacement transaction.

[17]Replace-by-fee in Bitcoin: https://en.bitcoin.it/wiki/Replace_by_fee

Users also struggled with the recovery mechanism. Some had lost their seed phrase, whereas others had not saved it in the first place. Our work extends prior findings on self-errors of cryptocurrency users [35, 55, 61] and shows the prevalence thereof during the wallet recovery phase. To alleviate this problem, encrypted cloud backups might be a viable option. During the setup, users could be given the option to encrypt their seed phrase with a password and store it in a cloud storage of their choice. In the case of lost seed phrases, users could then simply import their encrypted file and decrypt it in the wallet. Similarly, password managers could also be used to store seed phrases to ensure that funds could be recovered in case of non-functional apps or forgotten passwords. Both solutions, however, pose a potential new attack vector, as users would have to rely on a third party. To guarantee that such wallets remain non-custodial, the seed phrases would have to get encrypted on the user's device so that they were only stored by the service provider in an encrypted form. Reliance on a third party in such "hybrid" approaches might be acceptable, particularly for newcomers with small amounts of cryptocurrency. Alternatively, as a way to avoid storing seed phrases online at all, the wallets could provide backup reminders and ask users to enter their seed phrases periodically to ensure that they still have access.

5.4.2 Allow Wallet Personalization. Wallet personalization could also alleviate some of the identified issues. Users reported fixed transactions fees that led to pending transactions or lost funds. Customizable transaction fees, possibly with recommended values provided by the wallet, could have prevented some of these issues.

Distinguishing between advanced and new users could also be accomplished on the interface level. Personalized interfaces are known to enhance the overall UX [37] and we believe that different profiles for advanced versus new users can have the same effect for cryptocurrency wallets. Newcomers could be shown only the default values, whereas experienced and expert users could have advanced options, such as custom transaction data, fees, and key import/export. Particularly the latter features should be communicated clearly, as our analysis has shown that some reviewers were confused or unsure of where the keys are located (see Figure 2(b)).

The security settings implemented by cryptocurrency wallets were also found to be problematic. Shoulder surfing was perceived as a risk by users, with some being afraid of disclosing their wallet balance and others having the same concerns about their pin. A simple toggle option could be used to hide the wallet balance, whereas biometrics, e.g., a fingerprint, could be used to authenticate the wallet owner. Such features have to be implemented with caution, as some users complained about additional authentication measures being overbearing.

5.4.3 Improve Users' Understanding of Cryptocurrencies. New users appeared to be overwhelmed when using cryptocurrency wallets (Section 4.3.1). Blockchain characteristics were often unclear, particularly for users who used conventional payment systems as a reference point. Perhaps tutorials and sandboxes could help to familiarize newcomers with cryptocurrencies and could potentially prevent costly user errors, which are known to be prevalent in the domain [35, 54, 61].

Users also often expected help from the developers whenever something went wrong. Often, however, the raised issues were

the result of misconceptions on the users' end. Addressing these misconceptions is complicated, and there is no silver bullet. Yet guiding the users through their first transactions could improve their understanding. We believe that guiding users through a complex process safely, also referred to as *safe staging* [62], could be applied successfully to cryptocurrency wallets. Embedded videos or animations could further be used to explain the most important elements and characteristics, such as private keys, addresses, and irreversibility of transactions. These guides could also explain the differences when compared to conventional systems and could not only help the users but also save time for the developers and their customer support teams.

6 CONCLUSION

We collected 45,821 app reviews of mobile cryptocurrency wallets and employed ML and NLP techniques to select reviews relevant to UX. We then identified themes illustrating common UX issues as experienced by the users. These can be grouped into common and domain-specific UX issues, with both types leading to functional and monetary losses. Further, users appeared to rely on their understanding of traditional payment systems, such as online banking, when using cryptocurrencies and faced challenges when doing so. Security, privacy, and trust challenges were also uncovered that were caused by the UX shortcomings.

Our findings provide an overview of the challenges that cryptocurrency users are facing and the underlying misconceptions prevalent among users. We suggest that to make the domain more usable, future wallets need to mimic conventional payment systems and users need to be more supported before and during use.

ACKNOWLEDGMENTS

This research has been supported in part by a gift from Scotiabank to UBC. We would like to thank our anonymous reviewers for all the feedback and suggestions they provided to improve the paper.

REFERENCES

[1] ISO 9241-11. 2018. *Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts.* Standard. International Organization for Standardization.

[2] Svetlana Abramova, Artemij Voskobojnikov, Konstantin Beznosov, and Rainer Böhme. 2021. Bits Under the Mattress: Understanding Different Risk Perceptions and Security Behaviors of Crypto-Asset Users. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.*

[3] Andreas M Antonopoulos. 2014. *Mastering Bitcoin: unlocking digital cryptocurrencies.* O'Reilly Media, Inc.

[4] Elsa Bakiu and Emitza Guzman. 2017. Which feature is unusable? Detecting usability and user experience issues from user reviews. In *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW).* IEEE, 182–187.

[5] Lisa Benson, Dean Elliott, Michael Grant, Doug Holschuh, Beaumie Kim, Hyeonjin Kim, Erick Lauber, Sebastian Loh, and Thomas C Reeves. 2002. *Usability and Instructional Design Heuristics for E-Learning Evaluation.* Association for the Advancement of Computing in Education (AACE).

[6] Enrico Bertini, Silvia Gabrielli, Stephen Kimani, Tiziana Catarci, and Giuseppe Santucci. 2006. Appropriating and assessing heuristics for mobile computing. In *Proceedings of the working conference on Advanced visual interfaces.* 119–126.

[7] Vavrinec Cermak. 2017. Can Bitcoin Become a Viable Alternative to Fiat Currencies? An Empirical Analysis of Bitcoin's Volatility Based on a GARCH Model. *An Empirical Analysis of Bitcoin's Volatility Based on a GARCH Model (May 2, 2017)* (2017).

[8] Paola Ceruleo. 2014. Bitcoin: a rival to fiat money or a speculative financial asset? *Master's Thesis* (2014).

[9] Rishi Chandy and Haijie Gu. 2012. Identifying spam in the iOS app store. In *Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality.* 56–59.

[10] Ning Chen, Jialiu Lin, Steven CH Hoi, Xiaokui Xiao, and Boshen Zhang. 2014. AR-miner: mining informative reviews for developers from mobile app marketplace. In *Proceedings of the 36th international conference on software engineering.* 767–778.

[11] CoinMarketCap. 2021. Distinct Cryptocurrencies. https://coinmarketcap.com/all/views/all/. Accessed: 2020-01-09.

[12] Dead Coins. 2020. Scam Coins Overview. https://deadcoins.com//. Accessed: 2020-08-10.

[13] Tayana Conte, Jobson Massolar, Emilia Mendes, and Guilherme Horta Travassos. 2009. Web usability inspection technique based on design perspectives. *IET software* 3, 2 (2009), 106–123.

[14] Chris Elsden, Arthi Manohar, Jo Briggs, Mike Harding, Chris Speed, and John Vines. 2018. Making Sense of Blockchain Applications: A Typology for HCI. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–14.

[15] Daniel A Epstein, Nicole B Lee, Jennifer H Kang, Elena Agapie, Jessica Schroeder, Laura R Pina, James Fogarty, Julie A Kientz, and Sean Munson. 2017. Examining Menstrual Tracking to Inform the Design of Personal Informatics Tools. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* 6876–6888.

[16] Shayan Eskandari, Jeremy Clark, David Barrera, and Elizabeth Stobert. 2015. A First Look at the Usability of Bitcoin Key Management. *Proceedings 2015 Workshop on Usable Security, 2015* (2015).

[17] Adrienne Porter Felt, Erika Chin, Steve Hanna, Dawn Song, and David Wagner. 2011. Android permissions demystified. In *Proceedings of the 18th ACM conference on Computer and communications security.* 627–638.

[18] Michael Fröhlich, Felix Gutjahr, and Florian Alt. 2020. Don't lose your coin! Investigating Security Practices of Cryptocurrency Users. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference.* 1751–1763.

[19] Xianyi Gao, Gradeigh D Clark, and Janne Lindqvist. 2016. Of Two Minds, Multiple Addresses, and One Ledger: Characterizing Opinions, Knowledge, and Perceptions of Bitcoin Across Users and Non-Users. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* 1656–1668.

[20] Xiaodong Gu and Sunghun Kim. 2015. "What Parts of Your Apps are Loved by Users?". In *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE).* IEEE, 760–770.

[21] Greg Guest, Arwen Bunce, and Laura Johnson. 2006. How many interviews are enough? An experiment with data saturation and variability. *Field methods* 18, 1 (2006), 59–82.

[22] Emitza Guzman and Walid Maalej. 2014. How do users like this feature? a fine grained sentiment analysis of app reviews. In *2014 IEEE 22nd international requirements engineering conference (RE).* IEEE, 153–162.

[23] Elizabeth Ha and David Wagner. 2013. Do Android users write about electric sheep? Examining consumer reviews in Google Play. In *2013 IEEE 10th Consumer Communications and Networking Conference (CCNC).* IEEE, 149–157.

[24] James A Hanley and Barbara J McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 1 (1982), 29–36.

[25] Marc Hassenzahl and Noam Tractinsky. 2006. User experience - a research agenda. *Behaviour & information technology* 25, 2 (2006), 91–97.

[26] Haibo He and Yunqian Ma. 2013. *Imbalanced learning: foundations, algorithms, and applications.* John Wiley & Sons.

[27] Steffen Hedegaard and Jakob Grue Simonsen. 2013. Extracting usability and user experience information from online user reviews. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 2089–2098.

[28] Garrick Hileman and Michel Rauchs. 2017. Global Cryptocurrency Benchmarking Study. *Cambridge Centre for Alternative Finance* 33 (2017), 33–113.

[29] Kasper Hornbæk. 2006. Current practice in measuring usability: Challenges to usability studies and research. *International journal of human-computer studies* 64, 2 (2006), 79–102.

[30] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. 2013. *Applied logistic regression.* Vol. 398. John Wiley & Sons.

[31] Clayton J Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media.*

[32] ISO9241-210. 2019. *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems.* Standard. International Organization for Standardization.

[33] Pooya Jaferian, Kirstie Hawkey, Andreas Sotirakopoulos, Maria Velez-Rojas, and Konstantin Beznosov. 2011. Heuristics for evaluating IT security management tools. In *Proceedings of the Seventh Symposium on Usable Privacy and Security.* 1–20.

[34] Hammad Khalid. 2013. On identifying user complaints of iOS apps. In *2013 35th international conference on software engineering (ICSE).* IEEE, 1474–1476.

[35] Katharina Krombholz, Aljosha Judmayer, Matthias Gusenbauer, and Edgar Weippl. 2016. The Other Side of the Coin: User Experiences with Bitcoin Security and Privacy. In *International Conference on Financial Cryptography and Data Security.* Springer, 555–580.

[36] Bimal Aklesh Kumar and Munil Shiva Goundar. 2019. Usability heuristics for mobile learning applications. *Education and Information Technologies* 24, 2 (2019), 1819–1833.

[37] Ram L Kumar, Michael Alan Smith, and Snehamay Bannerjee. 2004. User interface features influencing overall ease of use and personalization. *Information & Management* 41, 3 (2004), 289–302.

[38] Joë l Valenzuela. Fall 2017. Coinomi Vulnerability Discovered, Developers React Harshly. https://dashnews.org/coinomi-vulnerability-discovered-developers-react-harshly/. Accessed: 2020-08-10.

[39] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[40] Walid Maalej and Hadeer Nabil. 2015. Bug report, feature request, or simply praise? on automatically classifying app reviews. In *2015 IEEE 23rd international requirements engineering conference (RE)*. IEEE, 116–125.

[41] Olibário Machado Neto and Maria Da Graça Pimentel. 2013. Heuristics for the assessment of interfaces of mobile devices. In *Proceedings of the 19th Brazilian symposium on Multimedia and the web*. 93–96.

[42] P. H. Madore. 2020. Bitcoin Nation: 22 Million US Crypto Traders Dwarf Global Rivals. https://www.ccn.com/bitcoin-nation-22-million-us-crypto-traders-dwarf-global-rivals/. Accessed: 2020-08-16.

[43] Alexandra Mai, Katharina Pfeffer, Matthias Gusenbauer, Edgar Weippl, and Katharina Krombholz. 2020. User Mental Models of Cryptocurrency Systems-A Grounded Theory Approach. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS) 2020)*. 341–358.

[44] Christopher Manning and Hinrich Schutze. 1999. *Foundations of statistical natural language processing*. MIT press.

[45] Stuart McIlroy, Nasir Ali, Hammad Khalid, and Ahmed E Hassan. 2016. Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews. *Empirical Software Engineering* 21, 3 (2016), 1067–1106.

[46] Satoshi Nakamoto. 2019. *Bitcoin: A peer-to-peer electronic cash system*. Technical Report. Manubot.

[47] Duc Cuong Nguyen, Erik Derr, Michael Backes, and Sven Bugiel. 2019. Short text, large effect: measuring the impact of user reviews on android app security & privacy. In *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 555–569.

[48] Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 249–256.

[49] Dennis Pagano and Walid Maalej. 2013. User feedback in the appstore: An empirical study. In *2013 21st IEEE international requirements engineering conference (RE)*. IEEE, 125–134.

[50] Sebastiano Panichella, Andrea Di Sorbo, Emitza Guzman, Corrado A Visaggio, Gerardo Canfora, and Harald C Gall. 2015. How Can I Improve My App? Classifying User Reviews for Software Maintenance and Evolution. In *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, 281–290.

[51] Helen Petrie and Nigel Bevan. 2009. The Evaluation of Accessibility, Usability, and User Experience. *The universal access handbook* 1 (2009), 1–16.

[52] Tomas Pranckevičius and Virginijus Marcinkevičius. 2017. Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Baltic Journal of Modern Computing* 5, 2 (2017), 221.

[53] Michel Rauchs, Apolline Blandin, Kristina Klein, Gina C. Pieters, Martino Recanatini, and Bryan Zheng Zhang. 2018. 2nd Global Cryptoasset Benchmarking Study. *Available at SSRN 3306125* (2018).

[54] Corina Sas and Irni Eliana Khairuddin. 2015. Exploring trust in Bitcoin technology: a framework for HCI research. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*. 338–342.

[55] Corina Sas and Irni Eliana Khairuddin. 2017. Design for Trust: An Exploration of the Challenges and Opportunities of Bitcoin Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6499–6510.

[56] Jonathan Sterling. Fall 2017. Coinomi – disclosure, denial, and destructive PR. https://cryptoinsider.media/coinomi-wallet-disclosure-denial-destructive-pr/. Accessed: 2020-08-10.

[57] Mike Thelwall and David Stuart. 2006. Web crawling ethics revisited: Cost, privacy, and denial of service. *Journal of the American Society for Information Science and Technology* 57, 13 (2006), 1771–1779.

[58] Muhammad Habib ur Rehman, Khaled Salah, Ernesto Damiani, and Davor Svetinovic. 2019. Trust in blockchain cryptocurrency ecosystem. *IEEE Transactions on Engineering Management* (2019).

[59] Mojtaba Vaismoradi, Hannele Turunen, and Terese Bondas. 2013. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & health sciences* 15, 3 (2013), 398–405.

[60] Sarah Theres Völkel, Ramona Schödel, Daniel Buschek, Clemens Stachl, Verena Winterhalter, Markus Bühner, and Heinrich Hussmann. 2020. Developing a Personality Model for Speech-based Conversational Agents Using the Psycholexical Approach. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[61] Artemij Voskobojnikov, Borke Obada-Obieh, Yue Huang, and Konstantin Beznosov. 2020. Surviving the Cryptojungle: Perception and Management of Risk Among North American Cryptocurrency (Non) Users. In *International Conference on Financial Cryptography and Data Security*. Springer, 595–614.

[62] Alma Whitten and J Douglas Tygar. 2003. Safe Staging for Computer Security. In *Workshop on Human-Computer Interaction and Security Systems*. Citeseer.

[63] Gavin Wood et al. 2014. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper* 151, 2014 (2014), 1–32.