ARTIFACT EVALUATED
USENIX
AVAILABLE

Designing Wallet-Based User Intervention for Approval Phishing Mitigation

Maggie Yongqi Guan[1] Yuqi Xu[2] Yunlong Mao[2,✉] Wei Tong[2] Xiaobo Zhou[1] Kanye Ye Wang[1,✉]

[1]*University of Macau* [2]*Nanjing University*

✉ *Corresponding authors*

Abstract

Approval phishing is a form of Web3 phishing that exploits token approval mechanisms to trick users into granting attackers spending authority over their tokens. As attackers increasingly hijack legitimate websites, URL-based detection alone becomes insufficient, leaving crypto wallets as the last line of defense. Based on the characteristics of approval mechanisms, we propose four wallet-based interventions for mitigating approval phishing: Spending Cap Suggestion, Active Spender Warning, Passive Spender Warning, and Delayed Confirmation. We evaluate the interventions through a between-subjects experiment (n = 364) and semi-structured interviews (n = 23). Compared with the control group, the Spending Cap Suggestion condition significantly increases the likelihood that users set spending caps. The Active Spender Warning, Passive Spender Warning, and Delayed Confirmation conditions all increase cancellation rates of phishing tasks, although the increases are statistically significant only for Active Spender Warning and Delayed Confirmation conditions. The effectiveness of the interventions varies across users, as users may struggle to interpret suspicious cues and focus on transaction outcomes while overlooking approval details. Our findings highlight the need to strengthen defenses against such attacks by increasing users' awareness of post-approval consequences and supporting approval-parameter verification at the moment of authorization.

1 Introduction

Phishing has emerged as the most damaging threat in Web3. In particular, **approval phishing** is among the most harmful forms, causing nearly $873 million in losses over the past three years [40] and attracting recent law enforcement attention [47]. It exploits a core mechanism in Web3 interactions: the token approval function, which allows users to grant third parties permission to spend tokens on their behalf [11].

Approval phishing is a form of phishing in which victims are tricked into signing a malicious approval transaction that grants a malicious spender the ability to access and drain their cryptocurrencies [6, 16, 39]. As shown in Figure 1, attackers hijack a legitimate decentralized application (DApp) website that users trust and visit for token trading [36]. When the user connects their wallet to the website and initiates a transaction, the compromised website constructs the malicious approval transaction and sends it to the user's wallet for further confirmation [53]. The malicious approval transaction looks like a benign transaction; however, the key difference is that the spender is an attacker-controlled account and is granted an unlimited allowance. Once the user signs the approval transaction, the attacker-controlled account is authorized to transfer an unlimited amount of the user's approved tokens[1].

User-oriented interventions are an important component of phishing defense [12]. Nudges are interventions that guide users toward more secure behaviors in anti-phishing contexts [12, 34], with common designs including URL highlighting [30, 51], forced-attention warnings [38], and time delays [37, 50]. However, these interventions are insufficient for defending against approval phishing. First, approval phishing can be delivered through attacker-hijacked legitimate websites rather than solely through obviously malicious URLs [36, 67]. Second, the critical user action shifts from entering credentials on the spoofed website to signing a transaction in the trusted wallet. This change in execution context and users' habitual signing routines can weaken perceived risk and reduce vigilance during the signing process [49].

Therefore, there is a need to explore interventions that specifically target approval phishing. Because the wallet serves as the final interface where users review and sign transactions regardless of how the deceptive context is introduced, wallet-centered interventions are uniquely positioned to provide a second layer of defense. Our goal is to understand how

[1]Another delivery approach resembles traditional phishing: attackers create a spoofed DApp site that imitates a legitimate service and distribute its URL through social channels to lure users. Once users land on the spoofed site, the workflow aligns with the steps described above: they connect their wallets and sign an approval transaction that is presented as legitimate but authorizes an attacker-controlled spender.

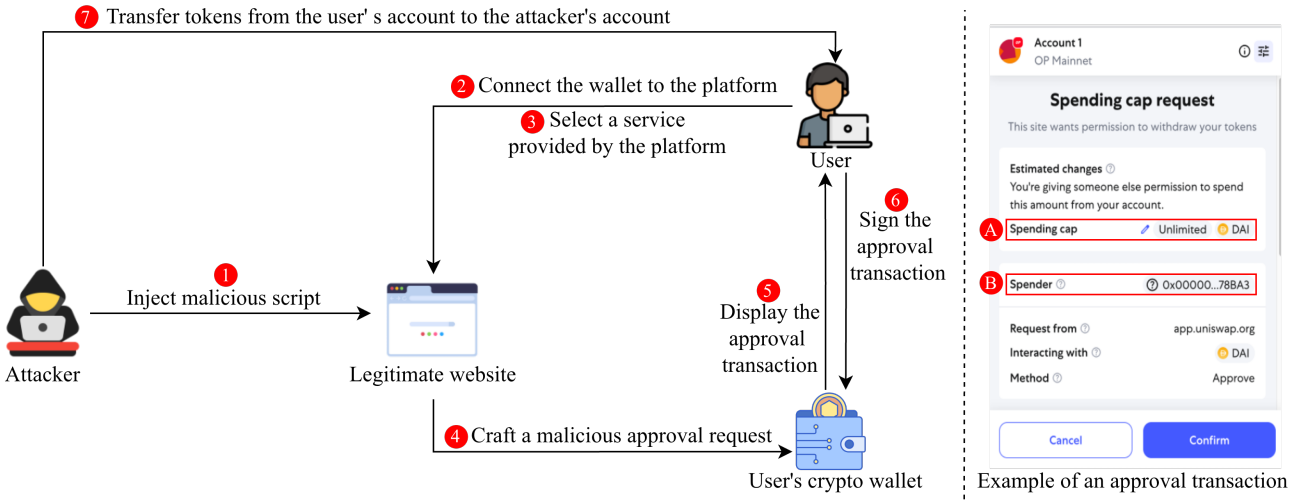


Figure 1: Approval phishing process. Step 1: The attacker hijacks a legitimate DApp website by injecting a malicious script. Steps 2–3: The user visits the compromised site, connects their wallet, and selects a service. Step 4: The compromised site constructs an approval transaction specifying (i) the token contract, (ii) the spender address, and (iii) the allowance. Steps 5–6: The wallet displays the approval request for the user to review and sign. Step 7: Once signed, the attacker can transfer the user's tokens up to the approved allowance. The right panel shows an example of an approval transaction. In a malicious approval transaction, the spender address is replaced with an attacker-controlled address, typically with an unlimited allowance.

to encourage users to scrutinize approval details and avoid signing malicious approval transactions through wallet-based interventions.

In this study, we adopt a mixed-methods approach to evaluate the effectiveness of wallet-based interventions. We situate our design within self-custodial software wallets, the primary interface driving Web3 interactions [64]. To systematically counter the insidious threat of approval phishing, we draw on nudge theory [4] to design a suite of interventions across the approval process. Specifically, to combat the default bias in which users tend to accept unlimited token allowances without scrutiny, we introduce *Spending Cap Suggestion*, which structures choice alternatives and prompts users to set explicit approval limits. To address risks associated with interacting with a malicious spender, we employ *Active Spender Warning*, a dynamic warning mechanism that surfaces salient risk cues at the point of interaction, helping users recognize suspicious spender requests that might otherwise be overlooked. Finally, we implement *Delayed Confirmation*, which introduces temporal friction to slow down the decision process and encourage a more deliberate review before transaction execution. Following an expert workshop, we refined the wording of warning messages to improve clarity and further introduced *Passive Spender Warning* as a comparative condition to examine how different warning strategies align with Web3 transaction workflows. We then conducted a between-subjects controlled experiment (n = 364) via Prolific with five conditions, including four intervention groups and a control group without intervention. Participants completed six token-approval tasks and a follow-up survey. In addition, we conducted semi-structured interviews (n = 23) to gain insights into participants' perceptions of the nudge interventions.

In our experimental analysis, we observed that the *Spending Cap Suggestion* condition increased the proportion of participants who set a spending cap compared to the *Control Group* condition, but the effect was limited because many participants prioritized convenience over repeatedly configuring caps. The *Active Spender Warning* and *Delayed Confirmation* interventions significantly reduced the likelihood of signing malicious approval transactions in phishing tasks relative to the *Control Group* condition, while the *Passive Spender Warning* condition showed a descriptively higher cancellation rate, although the increase did not reach statistical significance. However, behavioral analyses showed that participants often failed to move from reviewing the information to identifying the transaction as suspicious, which limited the overall effectiveness of the interventions. For the spender warnings, this limitation stemmed from strong completion goals, the cognitive effort required to verify spender legitimacy, and the perception of the passive warning as a routine notification rather than a meaningful alert. In the *Delayed Confirmation* condition, the intervention was constrained by users' uncertainty about what information should be reviewed and by misplaced trust in interface consistency. We further discuss how differences between phishing types reshape intervention design, identify the challenges faced by wallet-based interventions for approval phishing, and propose future designs to strengthen defenses against such attacks.

Our study makes two key contributions. First, we broaden the design space for phishing interventions by providing in-

sights into interventions at the stage where users execute actions in phishing contexts. Unlike prior link-based warning interventions that primarily aim to interrupt interactions, our work shifts the focus to execution-stage decision points, revealing new opportunities and constraints for behavioral defenses in transaction workflows. Second, we provide empirical evidence on which wallet-based intervention formats are more effective for mitigating approval phishing. Our study informs future design by emphasizing the need to reduce users' recognition burden and embed actionable cues within the interface.

2 Background of Approval Phishing

In this section, we first provide an overview of the key elements involved in an approval transaction. We then describe the approval process and explain how approval phishing unfolds through interactions between the attacker and the victim.

2.1 Tokens and Wallets

Tokens. Assets recorded on the blockchain are commonly referred to as *tokens*, which can be broadly categorized into two types. The first category includes *native tokens* such as BTC on Bitcoin and ETH on Ethereum. These tokens are integral to the underlying protocol and are used to pay transaction fees or reward validators [3, 14, 54].

The second category comprises *smart contract-based tokens*, issued and managed through programmable contracts deployed on the blockchain. These tokens, including stablecoins (e.g., USDT, DAI) and meme coins (e.g., SHIB), represent the majority of digital assets in Web3 today. On Ethereum, most tokens follow the ERC-20 standard [11]. The standard specifies a set of required functions that facilitate the transfer and management of tokens, along with events that record these operations on the blockchain. Users can transfer tokens by calling functions embedded in the token contracts.

Crypto Wallets. A crypto wallet is a software application or hardware device that allows users to manage their assets and conduct transactions on the blockchain [64]. While commonly referred to as *"wallets"*, they do not store the cryptocurrencies themselves. Instead, they store and manage users' private keys, which are required to access and control their associated accounts on the blockchain [65]. Crypto wallets can be categorized as custodial or self-custodial. In Web3 contexts, self-custodial wallets are widely used because they allow users to directly interact with DApps and sign transactions.

2.2 Approval Process

Approval Transaction. When users trade ERC-20 tokens on DApps, they may first need to sign an approval transaction authorizing the platform to spend tokens on their behalf [53]. This interaction follows the ERC-20 standard [11]. Specifically, the `approve` function allows account *User A* to authorize account *User B* (the spender) to spend up to a specified amount of tokens (e.g., *c*) from *User A*'s balance. Once the approval is executed, *User B* can call `transferFrom` to transfer up to *c* tokens from *User A* to a third account. As shown in Figure 1, an approval transaction specifies a spender address and a spending cap (allowance).

Approval Phishing. Attackers exploit approval transactions by manipulating the spender address and the approval amount to launch phishing attacks. One strategy is to hijack a legitimate website and modify its front end so that user interactions trigger approval requests for an attacker-controlled address, delegating spending rights over the users' tokens [36]. Another way is to create deceptive websites that closely mimic well-known legitimate platforms and distribute links to these websites through social media. In both cases, the approval transaction appears structurally valid, but the spender address is malicious. Combined with the common practice of requesting unlimited approval amounts by default, signing such malicious transactions grants the attacker unrestricted access to the user's tokens [6, 16].

Because the attack model of approval phishing is not limited to malicious websites but also includes the compromise of otherwise legitimate websites, defenses that focus solely on blocking suspicious URLs or preventing users from visiting malicious websites are insufficient. This characteristic motivates a shift toward wallet-centered defenses, where approval parameters can be scrutinized at the point of signing, regardless of how the user arrived at the transaction.

3 Related Work

In this section, we organize related work into two main areas. We first review prior research on phishing in the Web3 ecosystem, and then examine human-centered interventions for phishing prevention to ground our research goals.

3.1 Phishing Scams in Web3

Numerous research efforts have been dedicated to categorizing phishing scams, including rug pulls, fake tokens, honeypots, and giveaway scams [21, 29]. More recently, several studies have focused on phishing transactions. He et al. [16] introduced the concept of transaction phishing (TxPhish), where attackers craft phishing websites to deceive users into initiating malicious transactions that can fully drain their assets. Chen et al. [6] further revealed that attackers exploit complex on-chain semantics by crafting transactions or messages with malicious payloads that manipulate smart contract interactions. In addition, He et al. [17] observed that attackers increasingly deploy phishing contracts to evade detection,

tricking users into either directly transferring tokens or granting token approvals to malicious contracts. In this study, we use the term *approval phishing* to refer to phishing attacks that exploit token approval mechanisms, regardless of their specific delivery channel or scam narrative.

Building upon these understandings of phishing attacks, researchers have further developed various techniques to detect phishing activities. Because phishing-related transactions naturally form networked structures [59], many detection methods leverage graph representations to extract distinctive patterns. These approaches generally fall into three categories: feature engineering-based methods [5, 52], random walk-based methods [60, 62], and GNN-based methods [20, 41]. Industry efforts complement academic work through large-scale threat intelligence and early blocking. MetaMask and Etherscan maintain blocklists to warn users about malicious domains and flagged phishing addresses [2, 48], while Chainabuse enables user reports of phishing scams [1].

Prior research has identified the existence of approval phishing and examined its variations across contexts. However, limited attention has been paid to how to effectively defend against such attacks, particularly from a human-centered perspective. Such an approach is important because phishing is inherently a socio-technical problem, and purely technical defenses cannot fully mitigate users' vulnerability to deceptive interactions [12]. Building on the motivation of prior studies that emphasize the importance of mitigating approval phishing, our work focuses on developing human-centered defense strategies, aiming to support users in recognizing deceptive cues and avoiding signing malicious approval transactions.

3.2 Human-Centered Interventions for Phishing Prevention

Design interventions are approaches to mitigating phishing by influencing users' decision-making processes. One approach is to present warnings in situ to help users recognize phishing attempts [12]. Studies comparing passive and active warnings have shown that passive approaches, such as browser toolbars, are often ignored by users [8, 61]. In contrast, active warnings that demand user attention or require explicit action have been found to significantly reduce susceptibility to phishing attacks [8, 51]. Other research has explored the use of fear appeals and link-focused warnings to increase user compliance and to direct attention toward high-risk elements [24, 38, 42].

Another line of work examines how visual design supports secure decision-making [12]. Visual cues, such as dynamic security skins [7] and individualized site appearances [23], have been proposed to help users verify website authenticity. Color-coding and highlighting techniques have also been used to draw attention to critical information, including domain names and sender details [30, 34]. Although these visual cues can enhance users' security awareness, their overall effectiveness has shown mixed results across studies [26, 58].

Prior work has developed link-based human-centered interventions to discourage users from clicking suspicious links. These insights also inform efforts to mitigate approval phishing by preventing users from interacting with suspicious URLs. However, these early-stage defenses are undermined by attackers' hijacking of legitimate websites. This motivates a second layer of protection at signing time: because wallets are the mandatory interface where approvals are reviewed and signed, they remain a final checkpoint regardless of whether users arrive via a malicious URL or a compromised legitimate website. Accordingly, we position wallets as an intervention surface and design in-wallet nudges that prompt users to scrutinize approval parameters and avoid signing malicious approval transactions.

4 Methods

We adopted a mixed-methods approach to investigate nudge interventions for mitigating approval phishing. We developed and refined initial designs through an expert workshop, then built an experimental platform and conducted a controlled user study to evaluate the effectiveness of the proposed interventions. We subsequently conducted follow-up interviews to examine participants' perceptions of the interventions.

4.1 Initial Nudge Intervention Designs

To avoid confusion between our preliminary concepts and the finalized interventions that emerged after the expert workshop, this section summarizes the initial designs. The final versions are presented in Section 4.3.

We designed interventions targeting two dimensions of the approval function: *how much to approve* and *who to approve*. For the first dimension, we offered three suggested spending cap options during the approval process to draw users' attention to the allowance setting and encourage them to choose a safer limit. For the second dimension, we implemented an active warning that notified users when the spender information appeared suspicious. For the final submission stage, we introduced a short delay to prompt users to review the approval details before confirming.

4.2 Expert Workshop

Expert Recruitment. We defined experts in our study as individuals with more than three years of professional Web3 experience and familiarity with approval phishing. We recruited them through personal networks and Web3 discussion groups on Discord, and screened them based on their years of Web3 experience and whether their work involved Web3 security or wallets. In total, we recruited five blockchain experts. Demographic information is summarized in Table 1. Among them, E1 and E4 had expertise in wallet UI/UX design, and E5 was a researcher with expertise in Web3 phishing detection.

Table 1: Demographics of expert participants. The last column indicates years of experience in the Web3 industry.

ID	Age	Education	Occupation	Years
E1	25–34	Master's	Developer	3–5
E2	35–44	Bachelor's	Developer	> 5
E3	25–34	Bachelor's	Smart contract auditor	> 5
E4	35–44	Master's	Wallet developer	3–5
E5	25–34	Ph.D.	Postdoctoral researcher	> 5

Procedure. The workshop consisted of an introduction and two brainstorming sessions. The introduction established common ground by briefing all participants on the study's procedures, the motivation for addressing approval phishing, and the current prototype design. Participants also introduced their backgrounds and expertise to facilitate group collaboration. In the first brainstorming session, participants shared individual perspectives on the proposed design. The authors then facilitated a collaborative discussion in the second session, which built upon the initial feedback. Experts jointly explored alternative design solutions, debated trade-offs (e.g., friction vs. usability), and proposed new ideas.

Design Feedback from Workshop. The expert workshop yielded two key recommendations: (1) for the *Spending Cap Suggestion* condition, experts noted that showing preset buttons might not sufficiently encourage deliberate choices. They advised adding a brief caption to provide contextual guidance and including a visible *"Max"* option. Although functionally equivalent to unlimited approval, making it explicit among other predefined limits could prompt users to notice and compare alternatives; (2) for the *Active Spender Warning* condition, experts cautioned that a dynamic pop-up might disrupt users' transaction flow and cause unnecessary friction. They proposed adding a static visual cue as a less intrusive option. While none objected to keeping the pop-up, they suggested testing both static and dynamic forms to assess which better captures user attention.

4.3 Overview of Wallet-Based Interventions

After the expert workshop, we refined the initial designs and added an additional intervention targeting spender information. The following subsections describe each finalized intervention in detail. The figures in this section are cropped versions that highlight the most important elements, and the complete designs are available in Figure 8 in Appendix A.

Nudge 1: Spending Cap Suggestion for Approval Amount Configuration. This design is motivated by a common behavioral gap in which users overlook the need to set an approval limit because the default configuration grants unlimited access [53]. The goal is to increase users' awareness of the need to configure spending caps so that they can construct safer approval transactions and reduce potential financial losses in the event of approval phishing.

Guided by choice-architecture principles, this nudge follows three design intents. First, it reduces default-driven overexposure by steering users away from unlimited approvals. Second, it lowers configuration burden by providing a small set of selectable caps as decision shortcuts instead of requiring manual input. Third, it makes the security trade-off legible by framing each option as a bounded exposure level. Accordingly, we present four predefined options (25%, 50%, 75%, and Max) below the spending-cap input field, accompanied by a brief caption to provide contextual guidance (cf. Figure 2).

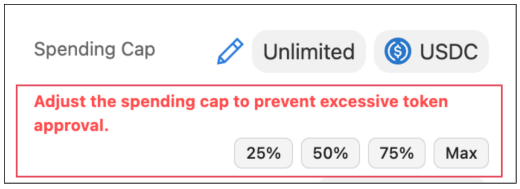


Figure 2: Spending cap suggestion. The wallet presents four spending-cap options accompanied by a reminder that prompts users to make an explicit choice. This intervention is intended to draw attention to the approval amount and reduce the habitual acceptance of unlimited approvals.

Nudge 2: Active Spender Warning for Spender Identification. This nudge targets spender identification. If the spender address is attacker-controlled, signing the approval grants the attacker an allowance to spend the user's tokens, which can enable unauthorized transfers and lead to asset loss.

We introduced the *Active Spender Warning* nudge to raise users' awareness of potential hazards and encourage more deliberate decision-making. Prior studies in usable privacy and security have shown that active warnings are effective when they explicitly interrupt user actions and draw attention to critical information before a decision is made [4, 8, 38]. In our design, the warning appears when the user attempts to submit an approval transaction and clearly indicates that the spender information may be suspicious. The warning presents two options, allowing the user to either continue or cancel the transaction (cf. Figure 3).

Nudge 3: Passive Spender Warning for Spender Identification. The third nudge also targets spender information but adopts a passive form. This design was informed by expert feedback and was intended to explore whether passive warnings suit the Web3 transaction context, where reducing intrusiveness helps maintain a smooth user experience.

Prior work shows that static warnings are generally less effective than active warnings in browser-based phishing contexts [8], yet this pattern does not consistently hold in email-

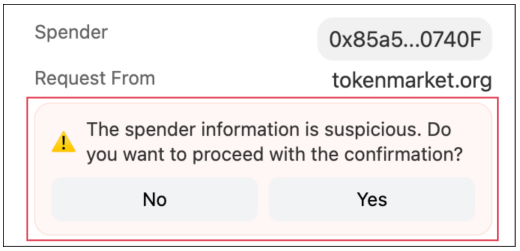


Figure 3: Active spender warning. The wallet presents a dynamic warning that explicitly flags the spender address as suspicious. Users can either proceed with the approval or cancel. This intervention aims to provide a clear, low-effort cue that prompts users to reassess the spender before signing.

based phishing scenarios [38]. These findings suggest that effectiveness may depend on context. Building on this insight, we designed a *Passive Spender Warning*: a static message placed directly beneath the "Request From" field to ensure users encounter it at the moment they evaluate the spender (cf. Figure 4).

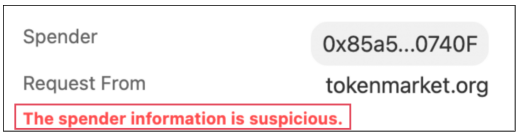


Figure 4: Passive spender warning. The wallet displays a static warning that highlights the spender address without interrupting the approval flow. This intervention provides a clear, low-effort cue that prompts users to reassess the spender before signing in a non-intrusive manner.

Nudge 4: Delayed Confirmation Before Transaction Submission. The fourth nudge focuses on the final checkpoint, where users review and sign an approval transaction in the wallet before it is submitted to the blockchain. Once the transaction is confirmed on-chain, a malicious spender gains an allowance to spend the approved tokens.

Building on prior work showing that short delays can encourage more deliberate review and disrupt habitual actions [37, 50, 55], we designed the *Delayed Confirmation* nudge to introduce a brief pause at confirmation time (cf. Figure 5). Specifically, when users click the "Confirm" button, a warning dialog appears and cannot be dismissed for five seconds. The dialog reminds users to review the spender information carefully and advises users to cancel the transaction if they are unsure about its legitimacy. This brief pause provides a moment for reflection and prompts users to reconsider the approval details, thereby reducing impulsive confirmations that could lead to token loss.

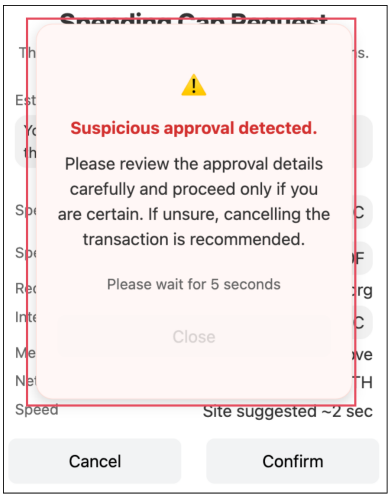


Figure 5: Delayed confirmation. The wallet enforces a five-second waiting period before users can confirm the approval, accompanied by a reminder to review the approval details. This intervention aims to reduce rushed, habitual signing by introducing a brief pause for deliberation.

4.4 User Study

In this section, we describe our user study in detail. We conducted a controlled experiment on Prolific to evaluate the effectiveness of the wallet-based interventions. We also conducted follow-up semi-structured interviews to further understand how participants interpreted the interventions.

4.4.1 Experiment

We conducted a controlled between-subjects experiment. We built a simulated Web3 trading platform named TokenMarket. This platform implements common functions, including token swapping and liquidity provision. Using these functions requires users to first approve TokenMarket as a spender so that the platform can execute token transfers on their behalf. To avoid priming participants about approval phishing, we employed deception and presented the study as an evaluation of TokenMarket for routine transaction activities. Participants interacted with the platform through simulated wallets and were assigned to one of the intervention conditions. Before starting the tasks, the platform homepage displayed the platform's official contract address, which served as the legitimate spender address in benign scenarios.

Each participant then completed six tasks on the platform. Four tasks were benign transactions that used the legitimate spender address, and two tasks were phishing transactions, one embedded in a token swap scenario and the other in a liquidity provision scenario. In the phishing transactions, the approval request specified a deceptive spender address

that differed from the platform's legitimate contract address[2]. All transactions were simulated and involved no real assets. Given that recent work on phishing detection systems has achieved high detection accuracy [6], our study treats the detection result as a relatively reliable input and focuses on users' responses to the intervention designs. Accordingly, we abstract away from detection errors and do not evaluate false positive rates (FPRs) or false negative rates (FNRs).

Pilot Study. Before the formal experiment, we conducted a pilot study with ten participants who were randomly assigned to one of the five experimental conditions. The pilot aimed to assess the feasibility of the study design and gather feedback for refinement. Two main issues emerged. First, participants unfamiliar with token swapping or liquidity provision were unsure how to complete these tasks. Second, the task instructions were accessible only from the home page, disrupting the task flow. To address these issues, we added brief introductions to token swapping and liquidity provision and made task details persistently visible at the top of the task list to improve navigation and continuity.

User Recruitment. We recruited participants through Prolific. Each experimental condition was released separately, and Prolific's *"exclude previous participants"* rule was applied to prevent repeated participation across conditions. In total, 387 responses were collected, exceeding the minimum required sample size estimated using G*Power[3]. Participants received $10 upon completing the study, and this honorarium was above Prolific's recommended minimum hourly rate.

Data Collection. Each participant was required to complete six tasks and a post-task survey. Participants were assigned to one of the five experimental conditions through separate Prolific postings. To minimize potential order effects, the sequence of tasks was randomized. After completing the tasks, participants filled out a post-task survey consisting of three sections: (1) reflections on their task experience and evaluation of the nudge intervention, (2) prior Web3 experience, and (3) demographic information. When participants interacted with the simulated trading website, the system captured their behavioral responses at each decision point. These included whether they set a spending cap, the proportion selected for the cap, whether they proceeded or canceled after receiving an intervention, and their final decision for each transaction.

Data Cleaning and Analysis. We collected 387 responses and excluded 23 with incomplete behavioral logs, as these participants did not complete all tasks and their button-click behaviors could not be verified. The final sample consisted of 364 participants, and participant demographics are presented in Table 2.

Table 2: Demographics of experimental participants.

Gender		**Age**	
Female	132 (36.3%)	18–24	36 (9.9%)
Male	231 (63.5%)	25–34	126 (34.6%)
Prefer not to say	1 (0.3%)	35–44	96 (26.4%)
		45–54	72 (19.8%)
		55 and above	34 (9.3%)
Education		**Web3 Experience**	
High school or below	81 (22.3%)	No experience	180 (49.5%)
Bachelor's degree	160 (44.0%)	Less than 1 year	65 (17.9%)
Master's degree	92 (25.3%)	1–3 years	90 (24.7%)
Doctorate	19 (5.2%)	Over 3 years	29 (8.0%)
Other	11 (3.0%)		
Prefer not to say	1 (0.3%)		
Location			
Africa	37 (10.2%)	Europe	249 (68.4%)
Asia	21 (5.8%)	North America	53 (14.6%)
South America	1 (0.3%)	Oceania	2 (0.5%)
Prefer not to say	1 (0.3%)		

Quantitative Data. The quantitative analysis focused on users' observable behaviors during approval transactions. These behavioral indicators included whether participants set a spending cap, whether they canceled the task, and other interaction actions captured during the tasks.

Given that the outcome measures did not follow a normal distribution, we used Kruskal–Wallis tests to examine whether configuration rates and cancellation rates differed across the five experimental conditions. When the test indicated a significant effect, we conducted post-hoc pairwise comparisons to examine differences between conditions. We conducted mixed-effects logistic regressions on configuration and cancellation actions to examine heterogeneity in how interventions and personal characteristics affected participants' behavior.

We further examined participants' decision-making processes. Based on the questionnaires, we decomposed the process into specific behaviors, including noticing, reviewing, and identifying. We used Firth logistic regression to examine which behavioral steps predicted cancellation and used Fisher's exact tests to examine whether the binary behavior differed across conditions.

We used a threshold ($p < .05$) to determine statistical significance. In Section 5, we report whether each result is statistically significant without listing the exact p-value for each test. Full statistical outputs are provided in Appendix B.

Qualitative Data. We employed thematic analysis [10, 22] to code the open-ended responses. The coding was jointly conducted by two researchers. Because the responses were generally clear and concise, making them straightforward to interpret, no major disagreements emerged during coding, and

[2]In the benign task, the spender address is the official TokenMarket address (0x7a25...2488E), whereas in the phishing task, the spender address is an attacker-controlled malicious address (0x85a5...0740F).

[3]The sample size was estimated using a repeated-measures ANOVA with a within–between interaction as an approximation, assuming an effect size of $f = 0.25$ and $\alpha = 0.05$.

we therefore did not conduct an inter-rater reliability (IRR) test [33]. We reference participants' open-ended responses using the notation (e.g., S0X) to indicate individual survey responses in Section 5.

4.4.2 Follow-up Interviews

After obtaining the preliminary quantitative results from the experiment, we selected 23 participants from the survey pool for semi-structured interviews. The interview data provided richer explanations and addressed the limitations of the quantitative results [15,43]. Our goal was to interpret the underlying behavioral mechanisms by eliciting deeper qualitative insights into participants' perceptions of the nudge interventions and how these interventions influenced their behavior.

Participant Recruitment. Participants were asked in the post-task survey whether they were willing to take part in a follow-up interview. Respondents who expressed interest in participating in the follow-up interview were shortlisted. Among those who consented, we purposively selected participants to ensure diversity and depth in qualitative insights. The selection primarily considered their task performance and prior Web3 experience. The detailed demographics of our interview participants are presented in Table 3.

Table 3: Demographics of interview participants

ID	Gender	Years in Web3	Condition
P01	Female	1–3	Control Group
P02	Male	1–3	Control Group
P03	Male	< 1	Spending Cap Suggestion
P04	Female	1–3	Spending Cap Suggestion
P05	Male	> 3	Active Spender Warning
P06	Female	> 3	Passive Spender Warning
P07	Female	1–3	Delayed Confirmation
P08	Male	No experience	Spending Cap Suggestion
P09	Male	< 1	Delayed Confirmation
P10	Male	1–3	Active Spender Warning
P11	Female	1–3	Active Spender Warning
P12	Female	< 1	Passive Spender Warning
P13	Female	1–3	Spending Cap Suggestion
P14	Male	> 3	Passive Spender Warning
P15	Male	1–3	Delayed Confirmation
P16	Female	< 1	Passive Spender Warning
P17	Male	1–3	Delayed Confirmation
P18	Female	No experience	Delayed Confirmation
P19	Male	> 3	Passive Spender Warning
P20	Female	1–3	Active Spender Warning
P21	Male	No experience	Spending Cap Suggestion
P22	Female	No experience	Passive Spender Warning
P23	Female	> 3	Delayed Confirmation

Interview Procedure. Each interview lasted approximately 45 minutes, and participants received a $10 honorarium. The semi-structured interviews comprised three parts: (1) participants' background and prior Web3 experience, (2) their decision-making processes during the experiment, and (3) their suggestions for improving the interventions.

Data Collection and Analysis. All interviews were audio-recorded and transcribed for analysis. We analyzed the qualitative data using thematic analysis [10]. Two researchers independently coded 20% of the transcripts and then met to reconcile the results and refine the codebook until reaching a high IRR (Cohen's $\kappa = 0.82$). One researcher then used the finalized codebook to code all transcripts.

We refrain from showing the exact frequency of codes or quantifying patterns, as doing so might imply our findings are generalizable. Instead, we present the general themes that emerged. Similar to prior work [9,13], we adopted the following terminology to provide a relative sense of the frequency of major themes: a few (0–20%), some (20–40%), about half (40–60%), many (60–80%), and almost all (80–100%). In Section 5, we reference interviewees' responses using the notation (e.g., P0X) to indicate individual responses.

4.5 Limitations

Our study design has several limitations. First, it raises ecological validity concerns, which are common in experimental research [8,61]. To mitigate this issue, we used deception by not disclosing the study's phishing focus in order to reduce expectancy effects [46], and we approximated real-world interaction contexts [25] by implementing transaction flows that closely mirror real-world processes. These design choices are consistent with prior phishing simulation studies and have been shown to improve behavioral realism [25,46].

Second, our sample includes non-Web3 users. However, the heterogeneity analyses in Section 5.1 show that prior Web3 experience is not significantly associated with either spending-cap configuration behavior or transaction cancellation behavior, suggesting that participants' outcomes were not systematically driven by familiarity with Web3 technologies. In addition, participants received a tutorial on the approval process before completing the tasks, which helped establish a shared baseline understanding and reduced the potential influence of non-Web3 users' unfamiliarity.

Third, our study does not directly test the behavioral effects of FNRs and FPRs. Repeated FP may undermine the perceived credibility of interventions and lead to warning fatigue or habituation. In contrast, FN may create a false sense of safety when no warning is shown, encouraging users to overrely on the absence of warnings. Although we did not manipulate FPR or FNR, our results in Section 5.2 provide evidence of their potential behavioral implications. Specifically, participants became overly cautious and canceled benign transactions, suggesting the type of behavioral cost that false-positive warnings could induce. Meanwhile, participants' difficulty in identifying phishing transactions without intervention highlights the residual risk that may arise when phishing transactions are not flagged.

5 Findings

Because our interventions aim to prompt users to scrutinize approval parameters during execution, we assess effectiveness along two complementary behavioral outcomes. First, we examine the configuration rate, which captures whether users actively modify the approval allowance instead of accepting the default unlimited amount. Second, we examine the cancellation rate, which indicates whether users detect risk cues at the final checkpoint and abort the approval transaction. We then analyze users' in-task behaviors to understand how they interpret and respond to each intervention.

5.1 Overall Effectiveness of Interventions

To evaluate the overall effectiveness of the four nudges, we analyzed two key indicators. The first is the *configuration rate (COR)*, defined as the proportion of tasks in which participants set a bounded approval cap. The second is the *cancellation rate (CAR)*, defined as the proportion of tasks in which participants canceled the approval transaction before execution.

5.1.1 Configuration Rate

A key parameter in an approval transaction is the allowance amount, which determines how much the spender is permitted to spend. The *Spending Cap Suggestion* nudge is designed to serve as a damage-limitation measure: it encourages users to configure the approval allowance to reduce potential losses if the approval turns out to be malicious. We use the COR to assess whether the *Spending Cap Suggestion* condition shifts users away from unlimited approvals.

As shown in Table 4, we found that the *Spending Cap Suggestion* condition had the highest phishing COR (31.25%), whereas the other conditions were approximately 1–2%. A Kruskal–Wallis test confirmed significant differences in phishing COR across conditions ($\chi^2(4) = 93.16$, $p < .001$, $\varepsilon^2 = .26$). Post-hoc pairwise comparisons showed that the *Spending Cap Suggestion* condition yielded a significantly higher COR than each of the other four conditions. No significant differences were observed among the remaining conditions, suggesting that the *Spending Cap Suggestion* condition was particularly effective in prompting users to configure a spending cap (cf. Table 6 in Appendix B).

Our mixed-effects logistic regression showed that configuration behavior was not significantly driven by task context or task sequence (cf. Table 7 in Appendix B). Furthermore, the heterogeneity analysis showed that Web3 experience did not significantly predict spending-cap configuration. Compared with participants with no Web3 experience, those with varying levels of Web3 experience (from less than one year to over three years) were not significantly more likely to set a spending cap (cf. Table 8 in Appendix B). These results suggest that neither task-specific triggers nor accumulated experience alone is sufficient to foster secure configuration habits.

Table 4: Configuration rates (COR) across the five conditions

Condition	Phishing Task COR (%)	Benign Task COR (%)
Control Group	1.97	2.63
Spending Cap Suggestion	31.25	6.60
Active Spender Warning	1.39	1.04
Passive Spender Warning	1.41	0.70
Delayed Confirmation	2.05	1.37

5.1.2 Cancellation Rate

The *Active Spender Warning*, *Passive Spender Warning*, and *Delayed Confirmation* interventions were designed to prompt users to assess potential risks and avoid signing malicious approval transactions. Therefore, we analyzed the CAR in these three conditions compared with the *Control Group*.

As shown in Table 5, the highest phishing CAR was observed in the *Active Spender Warning* condition (48.61%), followed by *Delayed Confirmation* (32.88%) and *Passive Spender Warning* (19.72%). A Kruskal–Wallis test revealed a significant difference in phishing cancellation rates across conditions ($\chi^2(4) = 79.84$, $p < .001$, $\varepsilon^2 = .21$). Post-hoc analysis showed that *Active Spender Warning* and *Delayed Confirmation* achieved significantly higher CAR than the *Control Group*, whereas *Passive Spender Warning* showed a higher descriptive CAR but did not reach significance relative to the *Control Group* after correction. *Active Spender Warning* significantly outperformed *Passive Spender Warning*, but its difference from *Delayed Confirmation* was not statistically significant after correction, suggesting that actively emphasizing spender risks can help users avoid signing malicious approvals (cf. Table 9 in Appendix B).

Our mixed-effects logistic regression also showed that cancellation behavior was not significantly driven by specific task context or task sequence, indicating that differences between phishing task types did not systematically influence participants' likelihood of canceling transactions (cf. Table 7 in Appendix B). Furthermore, the heterogeneity analysis showed that prior Web3 experience was not significantly associated with transaction cancellation in the overall sample. Compared with participants with no Web3 experience, those with varying levels of Web3 experience, ranging from less than one year to over three years, did not show a significantly higher propensity to cancel suspicious transactions (cf. Table 10 in Appendix B). Condition-specific analyses showed some significant experience-related differences in the *Active Spender Warning* and *Delayed Confirmation* conditions. However, these effects did not form a consistent monotonic pattern across experience levels. These results suggest that accumulated experience alone is generally insufficient to trigger

proactive defense, highlighting the importance of interface-level risk communication in supporting secure decisions.

Table 5: Cancellation rates (CAR) across the five conditions

Condition	Phishing Task CAR (%)	Benign Task CAR (%)
Control Group	1.32	0.00
Spending Cap Suggestion	0.00	0.00
Active Spender Warning	48.61	3.82
Passive Spender Warning	19.72	2.46
Delayed Confirmation	32.88	2.05

5.2 Analysis of Behavioral Mechanisms of the Interventions

Our nudge interventions were designed to target three key stages of the approval transaction process: (1) *parameter configuration*: how users set spending limits and determine how much an external contract is allowed to spend; (2) *information recognition*: how users identify whether the spender address is legitimate; and (3) *transaction confirmation*: how users review and confirm the approval transaction within the wallet interface. We analyzed the effectiveness of the interventions across these three aspects and examined how user behaviors in each dimension were influenced by the corresponding nudges.

5.2.1 Parameter Configuration

In many DApp workflows, approval requests use an unlimited allowance by default [53]. If users retain this default setting, the spender may transfer an unrestricted amount of the approved tokens. This exposes users to the risk of having their entire approved token balance transferred to the attacker's address. Therefore, the first intervention, *Spending Cap Suggestion*, is designed to encourage users to configure a suitable cap to minimize the risk of full token transfer.

To examine users' behavioral mechanisms, we analyzed their interpretation of and response to the intervention through a sequence of actions. We broke down their decision-making process into two key steps: first, whether they noticed the suggestive alternatives, and second, whether they configured the approval amount accordingly.

First, the appearance of alternatives raised participants' awareness of the need to set a cap. All participants reported noticing the suggestive alternatives, and 40.3% of them chose to set a spending cap. Some participants reported that presenting multiple predefined alternatives piqued their curiosity and prompted them to make a selection (e.g., *"The red characters reminded me to select one, so I did"* (S80)). This perception was echoed by P21, who had no prior Web3 experience. *"In two tasks, the interface offered options for me to choose from. To be cautious, I still made a selection, and I roughly understood that setting an approval cap limits the allowance"* (P21). Moreover, a few participants reported that the intervention had an educational effect that influenced their behavior in later tasks. P04, an experienced Web3 user, noted: *"I have used DApps for many years . . . and had never set an approval cap before. When I received a reminder . . . it caught my attention . . . and made me realize this is something worth considering, since unlimited approval could be risky"* (P04).

We further analyzed the specific alternatives selected by participants. Among the 45 phishing tasks in which participants set a spending cap, a custom value was selected in 2 tasks, while the 25%, 50%, 75%, and Max options were selected in 29, 9, 3, and 2 tasks, respectively. The distribution indicates a clear tendency toward lower-percentage limits rather than higher or unrestricted amounts. This suggests that the *Spending Cap Suggestion* not only increased the likelihood of setting a cap but also guided users toward safer configuration behaviors.

Despite all participants noticing the nudge, 59.7% still chose not to configure a spending cap, prompting us to examine the reasons behind their reluctance to set a cap. Qualitative feedback revealed two primary reasons: (1) a misunderstanding of the spending cap's function in approval transactions, and (2) a tendency to prioritize convenience over security.

First, some participants interpreted the alternatives as suggesting the proportion of tokens they were about to trade (e.g., *"It was telling me that I should set a spending cap, and my way of responding was to modify the swapping amount"* (S127)). Interview data further supported this misunderstanding, as a few participants perceived the spending cap as equivalent to the current transaction amount rather than as a mechanism to control future approval risks.

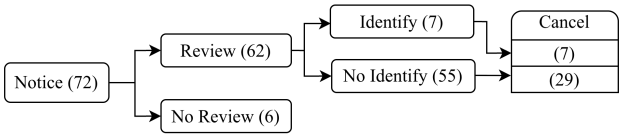
Second, prioritizing convenience over security in approval transactions was another factor contributing to users' reluctance to set a cap. Some participants who understood the potential consequences of not setting a cap expressed concern that doing so could complicate future transactions, particularly when larger transfers might exceed the preset limit. *"The cap is not specific to this transaction but applies to subsequent transactions on the same platform. If I set a limit now . . . and want to make a larger transaction later that exceeds the cap, I'll have to revoke the approval and set it again, which is quite troublesome"* (P13). A few participants also perceived setting an approval amount as burdensome, as it required an additional step in the transaction process.

5.2.2 Information Recognition

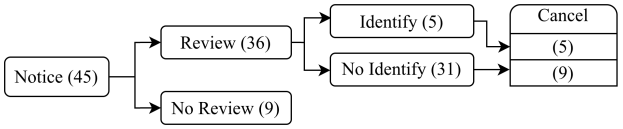
Verifying whether the spender is legitimate is a key step in the approval process. If users determine that the spender is malicious, they may avoid signing a malicious approval transaction. We designed two interventions to support spender recognition: *Active Spender Warning* (a warning that interrupts the user before confirmation) and *Passive Spender Warning* (a warning displayed concurrently with the transaction

details).

To examine how users interpreted these warnings and evaluated spender legitimacy, we decomposed behavior into four steps, each captured as a binary variable: noticing the warning (`SpenderNotice`), inspecting spender information (`SpenderReview`), judging the spender as suspicious (`SpenderIdentify`), and canceling the phishing transaction (`SpenderCancel`). Figure 6 reports the number of participants reaching each step, conditional on completing the preceding step, in the two spender-warning conditions.



(a) Active Spender Warning condition.



(b) Passive Spender Warning condition.

Figure 6: Number of participants at each of the four behavioral steps in the two spender warning conditions. Values in parentheses indicate participant counts.

We first assessed noticeability. In the *Active Spender Warning* condition, all participants noticed the warning, compared to 63.38% in the *Passive Spender Warning* condition. A Fisher's exact test showed a significant difference, suggesting that the interruptive warning was more attention-grabbing during signing. Some participants missed the passive warning because they signed out of habit and did not attend to the transaction content. For example, P06 noted: *"All I needed to do was to sign . . . it simply means I'm proceeding with the transaction . . . so I signed without much thought"* (P06).

Among those who noticed the warning, most participants chose to review spender information in both conditions (86.11% in *Active*; 80.00% in *Passive*). However, few participants identified the spender as suspicious (11.29% in *Active*; 13.88% in *Passive*). Even when participants did not conclude that the spender was malicious, some still canceled after reviewing because they perceived that the warning itself served as a risk cue (e.g., *"Just the pop-up warning made me not trust the transaction"* (S203); *"The red warning . . . I wouldn't take the risk to sign it"* (S286)). Survey responses further suggest that this caution spilled over to benign transactions. A few participants reported that after seeing and canceling tasks with warnings, they became suspicious of subsequent tasks even when no warning appeared. For example, one participant stated, *"I canceled the two tasks where warnings appeared. Then, when no warning appeared in the later task, I felt it might be a trap, so I canceled it as well"* (S265). This suggests that warnings may heighten users' general risk perception, but when users cannot confidently distinguish benign from malicious approval requests, such heightened caution may also produce a behavioral cost similar to false-positive warnings in deployment, leading users to unnecessarily cancel benign transactions.

Our Firth logistic regression analysis showed that both `SpenderReview` and `SpenderIdentify` were significantly associated with canceling phishing tasks in the spender-warning conditions (cf. Table 11 in Appendix B). In the *Passive Spender Warning* condition, `SpenderIdentify` showed the strongest association with cancellation, whereas in the *Active Spender Warning* condition, both reviewing and identifying spender information were strong predictors of cancellation. This suggests that users who reviewed spender information or judged a spender as suspicious were more likely to cancel. However, many participants struggled to judge whether the spender was malicious.

One reason was shallow inspection. Some participants reported that they glanced at the spender information but did not find a clear anomaly, and some also viewed the spender verification as costly because it required third-party tools. Their goal was to complete the transaction, and they focused on whether the transaction outcome matched expectations, treating approval information as secondary. As P10 explained: *"I just wanted to complete the transaction quickly and didn't pay attention to other details. Generally, I only check the final step, especially whether the expected change matches the transaction I intended"* (P10).

In addition, participants in the *Passive Spender Warning* condition interpreted the non-interruptive format as a routine notice rather than a risk alert. Some suggested that if the spender were truly dangerous, the wallet would block the approval instead of labeling it as suspicious. P14 described similar experiences: *"I don't think this is a dangerous signal, because MetaMask often gives me similar reminders. It usually says it's the first time I'm interacting with this address. Even if the site is legitimate, this message still appears, so I don't think there's a problem"* (P14).

5.2.3 Transaction Confirmation

Once an approval transaction is signed, the attacker gains the ability to transfer the approved tokens without further user interaction. To reinforce user review at the final checkpoint before transaction submission, we introduced the *Delayed Confirmation* intervention. It imposes a mandatory waiting period before the transaction is submitted, during which a brief reminder prompts users to review approval details.

We decomposed user behavior into three sequential steps, each represented by a binary variable: whether participants reviewed the approval details after receiving the delay

notification (DelayReview), identified suspicious elements (DelayIdentify), and ultimately canceled the malicious task to avoid phishing (DelayCancel). Figure 7 illustrates the number of participants at each step relative to the preceding behavior within the *Delayed Confirmation* condition.

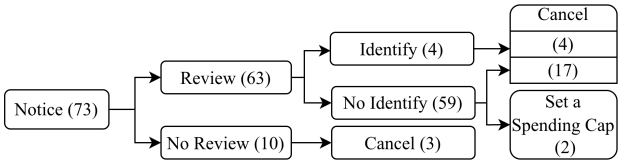


Figure 7: Number of participants at each behavioral step in the Delayed Confirmation condition. The values in parentheses represent the number of participants.

First, we examined users' actions after receiving the delay notification. Almost all participants re-examined the approval information, while a few proceeded to submit the transaction without review. Those who did not re-examine focused on the transaction rather than the approval, echoing the behavioral pattern discussed in Section 5.2.2. A few participants complained that the delay was intrusive and disrupted their transaction flow (e.g., *"I have never encountered this kind of pop-up warning in my everyday transactions. Since I could not close it immediately, it felt like a waste of time, although the delay was only 5 seconds"* (S320)).

Among participants who reviewed the approval transaction, only 6.34% identified suspicious elements and then canceled. Among those who failed to identify any suspicious elements, 28.81% still canceled. They perceived the delayed nudge as a helpful prompt to act cautiously, even without recognizing explicit danger. Similar to the responses observed in the *Active Spender Warning* and *Passive Spender Warning* conditions, benign cancellations in the *Delayed Confirmation* condition also reflected overcautious responses triggered by repeated exposure to warning-like friction. This finding further suggests that overly frequent interventions may impose usability costs by prompting users to abandon legitimate transactions. Experienced users also appreciated the delay. As P17 mentioned, it disrupted habitual confirmation and encouraged rechecking. *"The delay gave me a moment to think . . . I usually just glanced at the expected changes, but now it made me review details I would never have thought to check"* (P17).

A few participants adjusted the spending cap during the delay, even without perceiving explicit risk. For example, P07 said the waiting time led her to explore the interface and notice that the approval amount was editable: *"I clicked the small pencil icon during the delay and realized the amount could be customized. I hadn't noticed that before"* (P07). This suggests that *Delayed Confirmation* provided an opportunity for reflective interaction, prompting deeper engagement and awareness of security features previously overlooked.

Although some participants canceled out of caution rather than direct recognition, the identification of suspicious information still played a crucial role in preventing phishing. Firth logistic regression analysis confirmed that DelayIdentify was a significant positive predictor of cancellation (cf. Table 11 in Appendix B). However, users struggled to identify phishing cues in practice, as reflected in the low transition rate from reviewing to successful identification. This gap suggests that while most users engaged in reviewing, few could translate it into accurate detection.

Qualitative evidence suggests two main reasons: (1) uncertainty about what information to verify, and (2) misplaced trust arising from perceived consistency. First, many participants felt overwhelmed by the transaction information and were uncertain about where to focus their attention (e.g., *"I didn't know what I was looking for, so the prompt wasn't helpful since it didn't specify what to check"* (S321)). Second, some participants relied heavily on visual or contextual consistency, assuming that familiar layouts implied legitimacy (e.g., *"The layout looked consistent with the rest of the site, so I thought everything was fine"* (S351)). Similarly, some participants stated they verified details such as wallet address, network fee, and approval amount, yet still trusted the transaction because all information appeared normal. *"I looked at the wallet info, network fee, approval amount, and everything looked normal, so I thought it was safe"* (P15). Overall, participants tended to equate familiarity and consistency with safety rather than authenticity, revealing a key limitation of current transaction review practices.

6 Discussion

As outlined at the beginning of this study, although approval phishing is a subtype of phishing, it differs from traditional phishing. We reflect on our key findings in light of these differences to discuss how they reshape nudge design, including what new insights they offer for phishing research and what challenges remain. We then propose design implications for strengthening defenses against approval phishing.

6.1 How Phishing Differences Reshape Nudge Interventions

Compared with traditional phishing, approval phishing differs in (1) what the attacker ultimately gains, (2) when and where the user makes the critical decision, and (3) the type and amount of information the user must evaluate. These differences reshape the intervention design space and motivate defenses tailored to the approval workflow. In what follows, we compare traditional phishing and approval phishing in detail and further discuss how their differences reshape the design of interventions.

Attack mechanism: credential theft vs. permission abuse. Traditional phishing typically relies on spoofed login interfaces to trick users into disclosing sensitive credentials (e.g., passwords) [57]. Accordingly, many existing interventions prioritize blocking entry, aiming to interrupt attacks before users click suspicious links or land on fraudulent sites [12, 37, 50]. By contrast, interventions that support users at the moment of interaction (e.g., during form filling or decision-making on a phishing site) remain comparatively underexplored, although emerging work has begun to characterize the interaction patterns elicited by phishing websites [44].

We acknowledge that link- and site-based defenses can still partially mitigate approval phishing, as blocking known malicious domains remains an effective defensive path [5]. However, when attackers hijack the front ends of legitimate DApp websites and inject malicious interactions into an otherwise familiar workflow, the primary risk cue is no longer an obviously suspicious URL. More importantly, the security-critical information that users must evaluate (e.g., spender identity and permission scope) is surfaced most explicitly at the wallet's execution interface. Therefore, we extend interventions from initial exposure and entry to the execution stage by designing wallet-side interventions centered on the semantic information of approval transactions.

However, our findings indicate that users' awareness of permission abuse is limited (Section 5.2.2). Although warnings could attract attention, they often elicited only fast, automatic reactions and did not consistently prompt users to interpret approval semantics or anticipate downstream consequences. Consequently, warnings alone were insufficient to reliably prevent users from signing malicious approval transactions. We therefore identify a central design challenge: wallet-based interventions should help users shift from automated execution to reflective judgment while preserving transactional flow, enabling them to recognize that a website's appearance does not necessarily indicate its true security properties and to treat each signature as a security-relevant decision.

Interaction context: passive evaluation vs. goal-driven execution. In traditional phishing, users often encounter phishing emails in a relatively passive context [34, 56]. Their attention is typically directed toward evaluating surface cues such as sender identity, message wording, and contextual consistency [27, 28]. In such settings, nudges can be effective by drawing users' attention to suspicious elements and encouraging verification of key information. Prior work further suggests that when users adopt an evaluative mindset—reflecting on their goals and prior experience—they may become more skeptical and engage in deeper cognitive processing, increasing their sensitivity to deception cues [35, 56].

In contrast, approval phishing unfolds in a goal-driven execution context. Users are not simply assessing information but actively authorizing transactions to accomplish concrete objectives, often in the absence of salient link-based or content-based phishing cues. Intervening in this execution-focused setting, therefore, requires different design strategies. Our findings show that dynamic interventions (*Active Spender Warning* and *Delayed Confirmation*) tended to promote safer behavior than the static intervention (*Passive Spender Warning*). This result aligns with prior research indicating that active formats tend to be more effective than passive formats [8].

Our findings also highlight the heightened importance of personalization in the goal-driven execution context. Because users allocate attention primarily toward task completion, the same wallet-based intervention can be perceived either as valuable decision support or as disruptive friction. In our study, some participants welcomed warnings as an opportunity to slow down and inspect approval details, whereas others dismissed them to complete transactions quickly. A similar divergence emerged in responses to spending-cap configurations: while some users viewed configurable caps as meaningful safeguards, others prioritized convenience and perceived additional options as unnecessary overhead. This suggests that adaptive or configurable wallet interventions that account for users' goals and preferences may better balance usability and security while preserving transactional efficiency.

Information burden: recognizable cues vs. technical transaction data. Traditional phishing leverages stimuli such as emails, websites, or social media messages, which contain semantic and visual cues that users can perceive and evaluate. Accordingly, interventions can be designed to support users in identifying deceptive cues. In contrast, a key stimulus in approval phishing is technical transaction data. These parameters are presented as random-looking strings (e.g., spender addresses) with little intuitive semantic structure, making them difficult to judge based on appearance alone. Unlike URLs, which admit structured parsing rules and user education, developing transferable heuristics for interpreting on-chain parameters is substantially more challenging in practice.

When the cues themselves are not readily interpretable, interventions that rely on users independently identifying suspicious content are less likely to succeed. Informed by this constraint, our interventions do not ask users to authenticate addresses on their own; instead, we surface potentially abnormal spender information within the wallet and guide users to inspect approval parameters. Our results suggest that, when stimuli are difficult to interpret, explicitly informing users which cues are suspicious provides better support for safer decision-making than giving only a general warning about potential risk. This finding also resonates with prior observations that users are generally reluctant to invest substantial time and effort in complex security behaviors [19].

Moreover, our results highlight persistent challenges in helping users interpret technical transaction information [63]. The readability of transaction details has long been a bottle-

neck in Web3 interactions, and assessing whether an approval spender is risky often requires switching to external services such as blockchain explorers or third-party security tools. This semantic and workflow gap motivates future work to integrate more direct verification and decision support within wallet environments. Such integration could enable necessary double-checking without imposing excessive cognitive or interaction burden, thereby more effectively mitigating approval phishing risks.

6.2 Design Implications

Building on insights from Section 5 and Section 6.1, we demonstrate that wallet-based nudging interventions can support users in assessing the legitimacy of approval transactions, but their effectiveness is constrained by several challenges. We therefore propose a set of design implications to strengthen users' defenses against approval phishing.

Hybrid nudges to support reflective thinking. As discussed in Section 6.1, our findings indicate that warning-only interventions are insufficient to help users translate warnings into informed protective actions. Prior work on information nudges suggests that educational elements, such as explanatory text, can foster active cognitive engagement and reflective reasoning by clarifying why an action may be risky [4].

Building on these insights, we identify hybrid nudges as a promising design direction for wallet-side interventions. By combining warning signals with concise, decision-relevant information, hybrid nudges explicitly communicate why a warning is triggered and summarize the potential consequences of granted permissions, helping users move beyond passive risk acknowledgment toward more deliberate approval decisions.

At the same time, this approach introduces an inherent trade-off between informational sufficiency and interface burden. Wallet interfaces already present dense transaction and state information, and excessive disclosure risks increasing cognitive load and undermining effectiveness. Future designs should therefore balance transparency with usability, enabling access to additional details when needed while preserving efficiency in routine interactions.

Configurable nudges as personalized security measures. It must be acknowledged that nudging is not always appropriate or desirable, particularly given the complexity and uncertainty of users' real-world contexts [45]. This concern is especially salient in Web3 transactions, where users operate in strongly goal-driven execution contexts [66]. As shown in Section 5, participants expressed mixed reactions to wallet-based interventions, with some perceiving them as helpful safeguards and others viewing them as disruptive.

Wallets should treat nudging mechanisms as configurable security measures and allow users to customize interventions within a bounded range so that they better align with individual preferences and usage routines. Importantly, configurability should extend beyond a simple on/off switch and can be expressed along dimensions such as presentation form, triggering conditions, and intervention strength. For example, some users may prefer brief risk cues over explicit decision confirmations, while others may favor delayed confirmations to avoid committing to actions immediately. By granting users limited control over these dimensions while retaining core protections, wallets can improve intervention acceptability and compliance while preserving task efficiency.

Integrated auxiliary functions to bridge the semantic gap. As discussed in Section 6.1, the semantic gap in non-intuitive stimuli prevents reflective judgment from translating into concrete protective actions. Wallet-side interventions should not stop at issuing warnings; they should also provide auxiliary functions that help users understand and verify the legitimacy of the authorization target. On the one hand, auxiliary functions can consolidate authoritative information sources and verification signals, minimizing cross-platform switching and making decision-relevant evidence available at the signing moment, thereby reducing the perceived effort and inconvenience for users. For example, wallets could surface contract-level provenance, such as verified labels and audit references, alongside lightweight risk indicators derived from on-chain behavior. On the other hand, wallets may incorporate social or peer-level reference cues (e.g., signals derived from community experience or collective usage traces) as complementary context to support judgment. Such integrated assistance does not replace user agency; rather, it helps transform opaque parameters into interpretable evidence, enabling users to bridge the semantic gap and convert warnings into effective verification and mitigation behaviors. Because such collective signals may involve sensitive wallet-interaction traces, future wallet-side defenses should consider privacy-preserving collaborative learning or aggregation mechanisms when deriving shared risk indicators [18, 31, 32].

7 Conclusion

We designed four wallet-based interventions to mitigate approval phishing. We conducted a between-subjects experiment and follow-up interviews to evaluate their effectiveness. The results revealed substantial variation in how the interventions affected users' responses to approval details. We discussed how differences between traditional phishing and approval phishing shape intervention design and highlighted the challenges of applying nudges to approval phishing. We provided design implications for interventions in contexts where the volume and complexity of approval information require redirecting users' attention to reduce phishing risk.

Acknowledgments

The authors would like to thank the reviewers for the time and effort they devoted to reviewing this paper. This work was supported in part by the National Natural Science Foundation of China under Grant No. 62272222, the Science and Technology Development Fund, Macao S.A.R (FDCT), under File Nos. 0056/2025/RIB2, 0078/2023/AMJ, and 001/2024/SKL; and the University of Macau under Grant Nos. MYRG-GRG2024-00052-FST-UMDF and MYRG-GRG2025-00111-FST. We would like to thank the Faculty of Science and Technology, University of Macau, for providing a travel grant to support the presentation of this work.

Ethical Considerations

Our study was reviewed and approved by our university's Institutional Review Board (IRB). Because approval phishing involves potential financial harm, a central ethical priority was ensuring that participants were not exposed to real asset loss. All approval transactions were simulated in a controlled environment and were not connected to any blockchain. No real wallets or on-chain assets were involved.

The study involved deception, as the full purpose of the experiment was withheld until completion. During consent, participants were informed that some study details could not be disclosed in advance. After the study, all participants were debriefed about the true research goals and how approval phishing operates in real-world settings. The debriefing helped restore transparency, reduce potential confusion, and raise participants' awareness of similar risks outside the study.

We also protected participants' privacy. Survey responses and interview transcripts were anonymized and coded, and no identifiable information appears in the paper or shared materials. All data were stored securely and accessed only by the research team. Participants were compensated for their time in accordance with institutional guidelines.

Beyond participant protection, we considered the broader implications of this work for Web3 users, wallet developers, Web3 platform designers, and the security research community. For users, our findings highlight how habitual signing behavior may lead them to overlook or misunderstand important approval details. For designers, the results identify approval-interface elements that users commonly miss or misinterpret, suggesting opportunities to make high-risk information more salient, interpretable, and actionable. For security researchers, our work contributes a user-centered perspective to Web3 phishing defense by examining risks that arise within trusted wallet interfaces, rather than only through suspicious links.

We also considered potential publication risks. Attackers could, in principle, use our findings to infer which approval-related cues users tend to overlook. However, this risk is limited because the paper does not provide operational guidance for conducting attacks. Instead, it offers defensive insights for educating users and informing wallet and DApp interface design. Overall, we believe the defensive benefits of publication outweigh the associated risks.

Open Science

In the spirit of transparency and reproducibility, we make the following artifacts publicly available: (1) the interview protocol, (2) the codebooks for the interviews and survey, and (3) the experimental code. The appendix provides additional results from the statistical analyses. To protect participants' privacy and confidentiality, we do not release raw interview recordings or full transcripts. This decision is consistent with our IRB-approved protocol and established ethical research practices. The artifacts are available at https://doi.org/10.5281/zenodo.20414886.

References

[1] Chainabuse. Accessed: 2025-07-22. https://www.chainabuse.com/.

[2] Etherscan. Accessed: 2025-07-22. https://etherscan.io/.

[3] Agostino Capponi, Ruizhe Jia, and Kanye Ye Wang. Maximal extractable value and allocative inefficiencies in public blockchains. *Journal of Financial Economics*, 172:104132, 2025.

[4] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–15, 2019.

[5] Weili Chen, Xiongfeng Guo, Zhiguang Chen, Zibin Zheng, and Yutong Lu. Phishing scam detection on ethereum: Towards financial security for blockchain ecosystem. In *IJCAI*, volume 7, pages 4456–4462, 2020.

[6] Zhuo Chen, Yufeng Hu, Bowen He, Dong Luo, Lei Wu, and Yajin Zhou. Dissecting payload-based transaction phishing on ethereum. In *Network and Distributed Systems Security (NDSS) Symposium*, 2025.

[7] Rachna Dhamija and J. D. Tygar. The battle against phishing: Dynamic security skins. In *Proceedings of the 2005 Symposium on Usable Privacy and Security*, SOUPS '05, page 77–88, New York, NY, USA, 2005. Association for Computing Machinery.

[8] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1065–1074, 2008.

[9] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. Exploring how privacy and security factor into iot device purchase behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[10] Jennifer Fereday and Eimear Muir-Cochrane. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, 5(1):80–92, 2006.

[11] Ethereum Foundation. Erc-20 token standard. Access: 2025-07-22. https://ethereum.org/zh/developers/docs/standards/tokens/erc-20/.

[12] Anjuli Franz, Verena Zimmermann, Gregor Albrecht, Katrin Hartwig, Christian Reuter, Alexander Benlian, and Joachim Vogt. {Sok}: Still plenty of phish in the sea—a taxonomy of {user-oriented} phishing interventions and avenues for future research. In *Seventeenth symposium on usable privacy and security (SOUPS 2021)*, pages 339–358, 2021.

[13] Maggie Yongqi Guan, Yaxing Yao, Sio Hong Teng, Xiaobo Zhou, and Kanye Ye Wang. User perceptions of responsible gambling messages as nudges for gambling safety. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26, New York, NY, USA, 2026. Association for Computing Machinery.

[14] Maggie Yongqi Guan, Yaman Yu, Tanusree Sharma, Molly Zhuangtong Huang, Kaihua Qin, Yang Wang, and Kanye Ye Wang. Security perceptions of users in stablecoins: Advantages and risks within the cryptocurrency ecosystem. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 2753–2771. IEEE, 2025.

[15] Lois R Harris and Gavin TL Brown. Mixing interview and questionnaire methods: Practical problems in aligning data. 2010.

[16] Bowen He, Yuan Chen, Zhuo Chen, Xiaohui Hu, Yufeng Hu, Lei Wu, Rui Chang, Haoyu Wang, and Yajin Zhou. Txphishscope: Towards detecting and understanding transaction-based phishing on ethereum. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, pages 120–134, 2023.

[17] Bowen He, Xiaohui Hu, Yufeng Hu, Ting Yu, Rui Chang, Lei Wu, and Yajin Zhou. Phishing tactics are evolving: an empirical study of phishing contracts on ethereum. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 9(2):1–24, 2025.

[18] Ying He, Mingyang Niu, Jingyu Hua, Yunlong Mao, Xu Huang, Chen Li, and Sheng Zhong. Labobf: A label protection scheme for vertical federated learning through label obfuscation. *arXiv preprint arXiv:2405.17042*, 2024.

[19] Cormac Herley. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *Proceedings of the 2009 workshop on New security paradigms workshop*, pages 133–144, 2009.

[20] Hexiang Huang, Xuan Zhang, Jishu Wang, Chen Gao, Xue Li, Rui Zhu, and Qiuying Ma. Peae-gnn: Phishing detection on ethereum via augmentation ego-graph based on graph neural network. *IEEE Transactions on Computational Social Systems*, 11(3):4326–4339, 2024.

[21] Jintao Huang, Ningyu He, Kai Ma, Jiang Xiao, and Haoyu Wang. A deep dive into nft rug pulls. *arXiv preprint arXiv:2305.06108*, 2023.

[22] Molly Zhuangtong Huang, Rui Jiang, Tanusree Sharma, and Kanye Ye Wang. Exploring user perceptions of security auditing in the web3 ecosystem. *Cryptology ePrint Archive*, 2024.

[23] Luigi Lo Iacono, Hoai Viet Nguyen, Tobias Hirsch, Maurice Baiers, and Sebastian Möller. Ui-dressing to detect phishing. In *2014 IEEE Intl Conf on High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICESS)*, pages 747–754. IEEE, 2014.

[24] Jurjen Jansen and Paul Van Schaik. The design and evaluation of a theory-based intervention to promote security behaviour against phishing. *International Journal of Human-Computer Studies*, 123:40–55, 2019.

[25] Chris Karlof, J Doug Tygar, and David A Wagner. Conditioned-safe ceremonies and a user study of an application to web authentication. In *NDSS*, 2009.

[26] Iacovos Kirlappos and M Angela Sasse. Security education against phishing: A modest proposal for a major rethink. *IEEE Security & Privacy*, 10(2):24–32, 2011.

[27] Daniele Lain, Tarek Jost, Sinisa Matetic, Kari Kostiainen, and Srdjan Capkun. Content, nudges and incentives: A study on the effectiveness and perception of embedded phishing training. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pages 4182–4196, 2024.

[28] Daniele Lain, Kari Kostiainen, and Srdjan Čapkun. Phishing in organizations: Findings from a large-scale and long-term study. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 842–859. IEEE, 2022.

[29] Xigao Li, Anurag Yepuri, and Nick Nikiforakis. Double and nothing: Understanding and detecting cryptocurrency giveaway scams. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2023.

[30] Eric Lin, Saul Greenberg, Eileah Trotter, David Ma, and John Aycock. Does domain highlighting help people identify phishing sites? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2075–2084, 2011.

[31] Yunlong Mao, Mingyang Niu, Ziqin Dang, Chengxi Li, Hanning Xia, Yuejuan Zhu, Haoyu Bian, Yuan Zhang, Jingyu Hua, and Sheng Zhong. Ofl: Opportunistic federated learning for resource-heterogeneous and privacy-aware devices. *arXiv preprint arXiv:2503.15015*, 2025.

[32] Yunlong Mao, Zexi Xin, Zhenyu Li, Jue Hong, Qingyou Yang, and Sheng Zhong. Secure split learning against property inference, data reconstruction, and feature space hijacking attacks. In *European Symposium on Research in Computer Security*, pages 23–43. Springer, 2023.

[33] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for cscw and hci practice. *Proceedings of the ACM on human-computer interaction*, 3(CSCW):1–23, 2019.

[34] James Nicholson, Lynne Coventry, and Pam Briggs. Can we fight social engineering attacks by social means? assessing social salience as a means to improve phish detection. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 285–298, 2017.

[35] Norbert Nthala and Rick Wash. How non-experts try to detect phishing scam emails. In *Workshop on Consumer Protection*, 2021.

[36] Pierluigi Paganini. Threat actors stole $120 m in crypto from badgerdao defi platform, December 2021.

[37] Justin Petelka, Benjamin Berens, Carlo Sugatan, Melanie Volkamer, and Florian Schaub. Restricting the link: Effects of focused attention and time delay on phishing warning effectiveness. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 1–19. IEEE, 2025.

[38] Justin Petelka, Yixin Zou, and Florian Schaub. Put your warning where your link is: Improving and evaluating email phishing warnings. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–15, 2019.

[39] Nicola Ruaro, Fabio Gritti, Dongyu Meng, Robert McLaughlin, Ilya Grishchenko, Christopher Kruegel, and Giovanni Vigna. Approve once, regret forever: On the exploitation of ethereum's {Approve-TransferFrom} ecosystem. In *34th USENIX Security Symposium (USENIX Security 25)*, pages 1281–1298, 2025.

[40] Scam Sniffer Research. Scam Sniffer 2025: Crypto Phishing Losses Fall 83% to $84 Million, January 2026. Accessed: 2026-05-18. https://drops.scamsniffer.io/scam-sniffer-2025-crypto-phishing-losses-fall\-83-to-84-million/.

[41] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.

[42] Sebastian W Schuetz, Paul Benjamin Lowry, Daniel A Pienta, and Jason Bennett Thatcher. The effectiveness of abstract versus concrete fear appeals in information security. *Journal of Management Information Systems*, 37(3):723–757, 2020.

[43] Janice Jianing Si, Tanusree Sharma, and Kanye Ye Wang. Understanding user-perceived security risks and mitigation strategies in the web3 ecosystem. CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.

[44] Karthika Subramani, William Melicher, Oleksii Starov, Phani Vadrevu, and Roberto Perdisci. Phishinpatterns: measuring elicited user interactions at scale on phishing websites. In *Proceedings of the 22nd ACM Internet Measurement Conference*, pages 589–604, 2022.

[45] CASS R. SUNSTEIN. Nudges that fail. *Behavioural Public Policy*, 1(1):4–25, 2017.

[46] Jan Tolsdorf, David Langer, and Luigi Lo Iacono. Phishing susceptibility and the (in-) effectiveness of common anti-phishing interventions in a large university hospital. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security*, pages 4334–4348, 2025.

[47] United States Secret Service. Multi-National Operation Atlantic Targets Approval Phishing, April 2026. Accessed: 2026-05-18. https://www.secretservice.gov/newsroom/behind-the-shades/2026/04/multi-national-operation-atlantic-targets-approval-phishing.

[48] Nikita Varabei. Metamask and chainpatrol protect users with phishing warnings, 2024. Access: 2025-07-22. https://metamask.io/news/metamask-chainpatrol-protect-users-with-phishing-warnings.

[49] Arun Vishwanath, Brynne Harrison, and Yu Jie Ng. Suspicion, cognition, and automaticity model of phishing susceptibility. *Communication research*, 45(8):1146–1166, 2018.

[50] Melanie Volkamer, Karen Renaud, and Benjamin Reinheimer. Torpedo: tooltip-powered phishing email detection. In *ICT Systems Security and Privacy Protection: 31st IFIP TC 11 International Conference, SEC 2016, Ghent, Belgium, May 30-June 1, 2016, Proceedings 31*, pages 161–175. Springer, 2016.

[51] Melanie Volkamer, Karen Renaud, Benjamin Reinheimer, and Alexandra Kunz. User experiences of torpedo: Tooltip-powered phishing email detection. *Computers & Security*, 71:100–113, 2017.

[52] Yun Wan, Feng Xiao, and Dapeng Zhang. Early-stage phishing detection on the ethereum transaction network. *Soft Computing*, 27(7):3707–3719, 2023.

[53] Dabao Wang, Hang Feng, Siwei Wu, Yajin Zhou, Lei Wu, and Xingliang Yuan. Penny wise and pound foolish: quantifying the risk of unlimited approval of erc20 tokens on ethereum. In *Proceedings of the 25th International Symposium on Research in Attacks, Intrusions and Defenses*, pages 99–114, 2022.

[54] Kaiqi WANG, Ruiqi HONG, Yunlong MAO, and Sheng ZHONG. Secure solution for decentralized federated learning with blockchain. *SCIENTIA SINICA Informationis*, 54(2):316, 2024.

[55] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. A field trial of privacy nudges for facebook. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2367–2376, 2014.

[56] Rick Wash. How experts detect phishing scam emails. 4(CSCW2), October 2020.

[57] Rick Wash and Molly M Cooper. Who provides phishing training? facts, stories, and people like me. In *Proceedings of the 2018 chi conference on human factors in computing systems*, pages 1–12, 2018.

[58] Oliver Wiese, Joscha Lausch, Jakob Bode, and Volker Roth. Beware the downgrading of secure electronic mail. In *Proceedings of the 8th Workshop on Socio-Technical Aspects in Security and Trust*, pages 1–9, 2018.

[59] Jiajing Wu, Jieli Liu, Yijing Zhao, and Zibin Zheng. Analysis of cryptocurrency transactions from a network perspective: An overview. *Journal of Network and Computer Applications*, 190:103139, 2021.

[60] Jiajing Wu, Qi Yuan, Dan Lin, Wei You, Weili Chen, Chuan Chen, and Zibin Zheng. Who are the phishers? phishing scam detection on ethereum via network embedding. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(2):1156–1166, 2020.

[61] Min Wu, Robert C Miller, and Simson L Garfinkel. Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 601–610, 2006.

[62] Yijun Xia, Jieli Liu, and Jiajing Wu. Phishing detection on ethereum via attributed ego-graph embedding. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 69(5):2538–2542, 2022.

[63] Kailun Yan, Xiaokuan Zhang, and Wenrui Diao. Stealing trust: Unraveling blind message attacks in web3 authentication. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pages 555–569, 2024.

[64] Yaman Yu, Tanusree Sharma, Sauvik Das, and Yang Wang. " don't put all your eggs in one basket": How cryptocurrency users choose and secure their wallets. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2024.

[65] Liyi Zhou, Xihan Xiong, Jens Ernstberger, Stefanos Chaliasos, Zhipeng Wang, Ye Wang, Kaihua Qin, Roger Wattenhofer, Dawn Song, and Arthur Gervais. Sok: Decentralized finance (defi) attacks. In *2023 IEEE Symposium on Security and Privacy (SP)*, pages 2444–2461. IEEE, 2023.

[66] Vivian Ziemke, Benjamin Estermann, Roger Wattenhofer, and Ye Wang. What determines the price of nfts? In *2023 IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS)*, pages 1562–1568. IEEE, 2023.

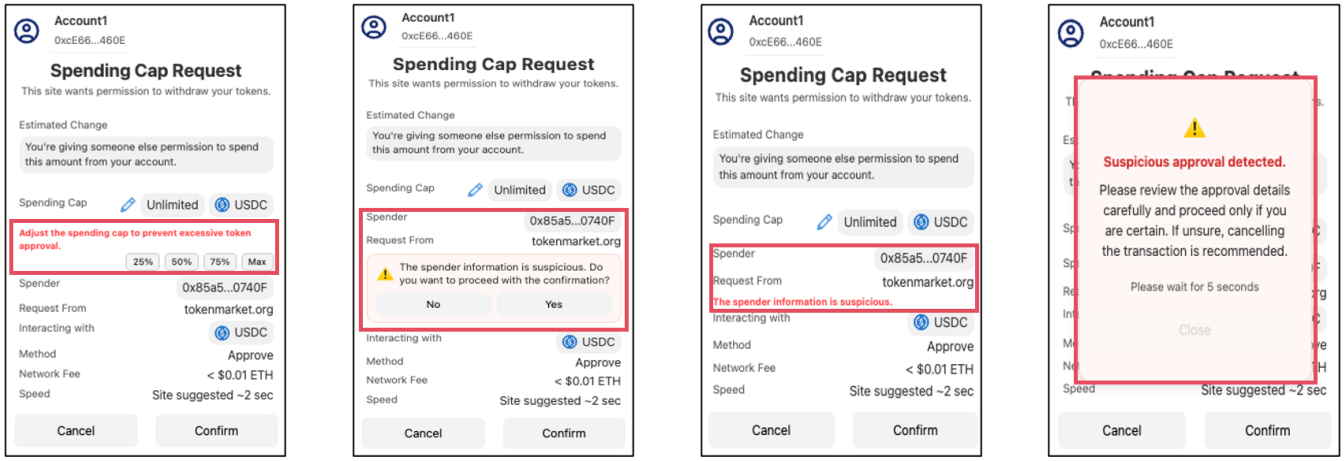
[67] Zeljka Zorz. Coinmarketcap, cointelegraph compromised to serve pop-ups to drain crypto wallets, June 2025.

A Full Views of the Four Wallet-Based Interventions

Figure 8 presents the complete design of our four wallet-based interventions.

B Statistical Results

Tables 6–11 provide detailed statistical results for the analyses reported in this paper. For readability, the main text reports statistical significance rather than full numerical outputs. Complete statistics are presented in the following tables.



(a) Spending Cap Suggestion (b) Active Spender Warning (c) Passive Spender Warning (d) Delayed Confirmation

Figure 8: Full views of the four wallet-based interventions evaluated in our study.

Table 6: Dunn's post-hoc test results for configuration rates

Comparison	***Z*-statistic**	**Adjusted *p*-value**	**Significance**
Spending Cap Suggestion vs. Control Group	7.67	< .001	***
Spending Cap Suggestion vs. Active Spender Warning	7.56	< .001	***
Spending Cap Suggestion vs. Passive Spender Warning	7.77	< .001	***
Spending Cap Suggestion vs. Delayed Confirmation	7.57	< .001	***
Control Group vs. Active Spender Warning	-0.00	1.000	ns
Control Group vs. Passive Spender Warning	0.23	1.000	ns
Control Group vs. Delayed Confirmation	-0.02	1.000	ns

Note. *** $p < .001$.

Table 7: Mixed-effects logistic regression results for cancellation and configuration behaviors

	Cancellation rate		**Configuration rate**	
Predictor	**Estimate (β)**	***p*-value**	**Estimate (β)**	***p*-value**
(Intercept)	-15.42	.715	-11.71	< .001***
Task ID	0.72	.719	0.67	.341
Task Sequence	-0.06	.958	0.11	.702

Note. *** $p < .001$.

Table 8: Effects of Web3 experience on transaction configuration behavior

Scope	**Experience Level**	**OR**	**95% CI**	***p***
Overall	Less than 1 year vs. No experience	2.09	[0.84, 5.01]	0.111
	1–3 years vs. No experience	1.62	[0.68, 3.77]	0.272
	Over 3 years vs. No experience	2.19	[0.63, 6.50]	0.204
Control Group	Less than 1 year vs. No experience	0.54	[0.00, 7.24]	0.683
	1–3 years vs. No experience	0.54	[0.00, 7.24]	0.683
	Over 3 years vs. No experience	1.05	[0.01, 14.80]	0.974
Spending Cap Suggestion	Less than 1 year vs. No experience	3.18	[0.92, 11.54]	0.067
	1–3 years vs. No experience	1.92	[0.57, 6.43]	0.288
	Over 3 years vs. No experience	3.12	[0.65, 16.33]	0.154
Active Spender Warning	Less than 1 year vs. No experience	2.63	[0.01, 497.33]	0.639
	1–3 years vs. No experience	10.76	[0.82, 1516.90]	0.073
	Over 3 years vs. No experience	5.46	[0.03, 1054.31]	0.429
Passive Spender Warning	Less than 1 year vs. No experience	0.73	[0.00, 14.77]	0.849
	1–3 years vs. No experience	0.35	[0.00, 6.78]	0.493
	Over 3 years vs. No experience	1.67	[0.01, 35.95]	0.772
Delayed Confirmation	Less than 1 year vs. No experience	10.83	[0.54, 1624.28]	0.117
	1–3 years vs. No experience	8.03	[0.41, 1198.26]	0.168
	Over 3 years vs. No experience	9.22	[0.05, 1810.85]	0.317

Table 9: Dunn's post-hoc test results for cancellation rates

Comparison	***Z*-statistic**	**Adjusted *p*-value**	**Significance**
Control Group vs. Active Spender Warning	-7.26	< .001	***
Spending Cap Suggestion vs. Active Spender Warning	-7.36	< .001	***
Active Spender Warning vs. Passive Spender Warning	4.41	< .001	***
Control Group vs. Delayed Confirmation	-4.78	< .001	***
Spending Cap Suggestion vs. Delayed Confirmation	-4.91	< .001	***
Spending Cap Suggestion vs. Passive Spender Warning	-2.92	.034	*
Control Group vs. Passive Spender Warning	-2.77	.057	ns
Control Group vs. Spending Cap Suggestion	0.20	1.000	ns
Active Spender Warning vs. Delayed Confirmation	2.47	.134	ns
Passive Spender Warning vs. Delayed Confirmation	-1.96	.502	ns

Note. *** $p < .001$, * $p < .05$.

Table 10: Effects of Web3 experience on transaction cancellation behavior

Scope	**Experience Level**	**OR**	**95% CI**	*p*
Overall	Less than 1 year vs. No experience	1.03	[0.49, 2.07]	0.939
	1–3 years vs. No experience	1.36	[0.73, 2.50]	0.326
	Over 3 years vs. No experience	2.04	[0.84, 4.71]	0.113
Control Group	Less than 1 year vs. No experience	9.22	[0.46, 1379.12]	0.142
	1–3 years vs. No experience	2.86	[0.02, 540.05]	0.611
	Over 3 years vs. No experience	5.53	[0.03, 1061.43]	0.424
Spending Cap Suggestion	All participants exhibited identical behavior; regression not estimable			
Active Spender Warning	Less than 1 year vs. No experience	**0.16**	[0.03, 0.66]	**.010****
	1–3 years vs. No experience	0.61	[0.20, 1.87]	0.388
	Over 3 years vs. No experience	**9.83**	[1.03, 1318.30]	**.047***
Passive Spender Warning	Less than 1 year vs. No experience	3.99	[0.91, 18.86]	0.067
	1–3 years vs. No experience	1.09	[0.25, 4.71]	0.907
	Over 3 years vs. No experience	1.81	[0.16, 13.57]	0.590
Delayed Confirmation	Less than 1 year vs. No experience	1.81	[0.44, 6.92]	0.396
	1–3 years vs. No experience	**4.33**	[1.33, 14.97]	**.015***
	Over 3 years vs. No experience	3.42	[0.47, 25.15]	0.211

Note. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table 11: Firth logistic regression analyses of behaviors predicting transaction cancellation across three interventions

Condition	**Predictor**	β	**SE**	**OR**	**95% CI for OR**	*p*
Active Spender Warning	(Intercept)	-3.71	1.43	0.02	[0.00, 0.18]	$< .001^{***}$
	Review	3.75	1.44	42.50	[5.60, 5453.72]	$< .001^{***}$
	Identify	3.33	1.45	27.97	[3.58, 3606.90]	$< .001^{***}$
Passive Spender Warning	(Intercept)	-4.65	1.42	0.01	[0.00, 0.07]	$< .001^{***}$
	Notice	1.04	2.02	2.84	[0.02, 532.81]	.613
	Review	2.73	1.46	15.38	[1.91, 1995.93]	$.005^{**}$
	Identify	3.92	1.47	50.51	[5.96, 6621.46]	$< .001^{***}$
Delayed Confirmation	(Intercept)	-0.80	0.47	0.45	[0.17, 1.09]	.077
	Review	-0.09	0.51	0.91	[0.34, 2.65]	.856
	Identify	3.73	1.47	41.64	[4.96, 5438.61]	$< .001^{***}$

Note. $^{***}p < .001$, $^{**}p < .01$.