Would you buy this bicycle?
“Users spend most of their time on other sites. This means that users prefer your site to work the same way as all the other sites they already know.”

–Jakob’s Law
Increasing Adoption of 2FA

Data source: https://github.com/2factorauth/twofactorauth
How consistent is the 2FA user experience across different websites?
What are the factors to compare the 2FA user journeys of different websites?
**Our Methodology**

- Literature review
- Basic structure of user journeys
- Recording of user journeys
- Open and axial coding of the user journeys
- Consistency analysis
Our Methodology

Data collection

- Literature review
- Basic structure of user journeys
- Recording of user journeys
- Open and axial coding of the user journeys
- Consistency analysis

Recording of user journeys on websites
Coding of comparison factor and user journeys
Agreement between researchers
Website comparison and consistency analysis code

Data Collection
Identifying Comparison Factors
Analysis
• Researchers independently explored and screen-recorded the 2FA user journeys

• Basic structure of exploration consists of 5 steps:
  – Discovery
  – Education
  – Setup
  – Usage
  – Deactivation
Data Collection: Data Set

• Websites chosen from the 2fa.directory data set
  – Websites ranked by Tranco data set
  – Top-ranked websites for each 2fa.directory category

• Final data set 85 websites

https://tranco-list.eu/
https://2fa.directory/
Our Methodology

Data collection

- Literature review
- Basic structure of user journeys
- Recording of user journeys

Identifying Comparison Factors

- Open and axial coding of the user journeys

Consistency analysis
Identifying Comparison Factors

• Emergent coding to identify factors: open and axial coding
  – Segment the user journeys into meaningful parts and assign codes ("concepts")
  – Combine codes via induction and deduction into categories (=factors)
Identifying Comparison Factors

• Emergent coding to identify factors: open and axial coding
  – Segment the user journeys into meaningful parts and assign codes ("concepts")
  – Combine codes via induction and deduction into categories (=factors)

Axial coding

Factor: “Promotion of 2FA”

Code: “2FA advertised prior to account creation”

Code: “2FA advertised during/after account creation”
**Set of Comparison Factors**

- 22 factors that either match, quasi-match, or do not match a website
  - 8 conditional factors that might not be applicable to a website

| Factors for Discovery                  |   |   |   |
|---------------------------------------|---|---|---|
| D1 Promotion                          | ● | ● | ● |
| D2 Non-Optional                       | ● |   | ● |
| D3 Common-Naming-and-Location         | ● |   | ● |

| Factors for Education                 |   |   |   |
|---------------------------------------|---|---|---|
| E1 Descriptive-Notification           | ● | ● | ● |
| E2 Additional-Information             |   | ● |   |

| Factors for Setup                     |   |   |   |
|---------------------------------------|---|---|---|
| S1 Option-Specific-Information       | ● |   | ● |
| S2 Step-Wise-Instructions            | ● |   | ● |
| S3 Multiselection                     | ● | ● | ● |
| S4 Grouped-Setting (S3)               | ● |   | ● |
| S5 No-Enforced-Options (S3)           | ● | ● | ● |
| S6 Selectable-Primary-Option (S3)     | ● |   | ● |

| Factors for Usage                     |   |   |   |
|---------------------------------------|---|---|---|
| U1 Device-Remembrance                 | ● | ● | ● |
| U2 No-Preselected-Option (S3,S6)      | ● | ● | ● |

| Factors for Deactivation              |   |   |   |
|---------------------------------------|---|---|---|
| R1 Informed-Deactivation              | ● |   | ● |
| R2 Deactivation-Verification (R1)     | ● | ● | ● |
| R3 Deactivation-Notification (R1)     | ● | ● | ● |
| R4 Communicate-Successful-Deactivation (R1) | ● | ● | ● |
Final Dataset

Websites (85)

Factors
- Discovery
- Education
- Setup
- Usage
- Deactivation

Matches
Quasi matches
Does not match
Does not apply
Our Methodology

Data collection

- Literature review
- Basic structure of user journeys
- Recording of user journeys

Identifying Comparison Factors

- Open and axial coding of the user journeys

Analysis

- Consistency analysis
2FA user journeys and individual factors are not very consistent across all 85 websites.
Clusters of User Journeys

**No predominant start-to-end strategy** exists that is followed by the majority of websites.

| Cluster | Description | Count |
|---------|-------------|-------|
| Cluster 1 | (n = 30) | How to inform and instruct users |
| Cluster 2 | (n = 29) | How to support multiple 2FA options |
| Cluster 3 | (n = 4) | Strategy for device remembrance |
| Cluster 4 | (n = 9) | |
| Cluster 5 | (n = 8) | |
| Cluster 6 | (n = 5) | |
Qualitative Data Analysis

Consistent Discovery for Self-Motivated Users

Two-Factor Authentication is an opt-in feature on most websites.

Consistent naming and location of the 2FA settings.

Vast majority of websites did not immediately promote 2FA before/during/after account setup.
Mixed Strategies for 2FA Setup and Configuration

Almost even split between three strategies:
1) “offering only one 2FA options”
2) “offering multiple 2FA options but only one can be active at a time”
3) “offering multiple 2FA options and supporting multiple active ones”

Half of the websites enforce a certain option (e.g., phone number) before allowing further options
Discussion

• Consistency *does not guarantee* good usability and UX
  – Example outlier in our data set: icloud.com
  – Consistent problematic design (e.g., nudges and descriptions in our data set)
  – This work: No attempt to assign a quality measurement to individual factors and overall 2FA UX
Discussion

• Consistency **does not guarantee** good usability and UX
  – Example outlier in our data set: icloud.com
  – Consistent problematic design (e.g., nudges and descriptions in our data set)
  – This work: No attempt to assign a quality measurement to individual factors and overall 2FA UX

• **Limitations** of qualitative studies and the study setup
  – Subjective bias by involved researchers
  – Skewed toward top-websites in English from certain categories
  – Only desktop client in Germany and collection between 06/21–08/21
  – Only user journey for account creation and initial 2FA setup
Conclusion

• Contributes a methodology for comparing 2FA user journeys on websites and the first systematic study of the consistency of those journeys

• No incumbent, consistent start-to-end design pattern for 2FA user journeys
  – Clusters of user journeys and individually consistent factors

• Call to action: Industry associations and the community could draft recommendations and guidelines for 2FA implementers
  – More insights needed: User and developer studies
  – Measure the impact of regulations on 2FA user journeys
  – Extending our methodology: Account recovery, other form factors, or passkeys
Backup Slides
Did users have negative experiences in transferring their 2FA knowledge?
Anectodal Evidence From Prior Work

• Did users have negative experiences in transferring their 2FA knowledge?
• Has this stopped them from enabling or using 2FA?

• Ciolino et al. ‘19 and Reynolds et al. ‘18: Evidence that users struggled with 2FA when the 2FA user journey did not match their expectations or previous experiences.

S. Ciolino et al., “Of two minds about two-factor: Understanding everyday FIDO U2F usability through device comparison and experience sampling,” in SOUPS ’19

J. Reynolds et al., “A tale of two studies: The best and worst of Yubikey usability,” in IEEE SP ’18.
Survey Among 2FA Adopters

- Did users have negative experiences in transferring their 2FA knowledge?
- Has this stopped them from enabling or using 2FA?

- Survey on Prolific with 308 participants that have 2FA experience

- Summary: 60 (19.5%) participants reported using a website less, abandoning a website, or refusing the adoption of a (specific) 2FA option due to differences in experience

S. Ghorbani Lyastani, M. Backes, and S. Bugiel, “A systematic study of the consistency of two-factor authentication user journeys on top-ranked websites (extended version),” 2022. [Online]. Available: https://arxiv.org/abs/2210.09373
FIDO UX Guidelines

• Similar steps in the user journey (promotion, invitation, registration, login)
  – Implement some best practices (“learn more,” confirm successful registration with a clear indication to users, encourage users to set up multiple keys for recovery/backup, “Security Settings”)

• Goal: Promote biometric awareness for passwordless logins or security keys for consumers on regulated industry websites (banking, healthcare)
  – Not suitable as general guidelines
  – Either no intention to cover a 2FA setting or limit themselves to security keys as a second factor
Consistency Analysis: Individual Factors

Shannon entropy $H(x)$ of non-conditional factors:

| Comparison Factor                          | $H(x)$ | $max$ |
|-------------------------------------------|--------|-------|
| Two-point scale                           |        |       |
| Non-optional                              | 0.37   | 1.0   |
| Additional-information                    | 0.90   |       |
| Option-specific-information               | 0.99   |       |
| Stepwise-instructions                      | 0.87   |       |
| Settings-changed-verification             | 0.99   |       |
| Settings-changed-notification             | 1.00   |       |
| Three-point-scale                         |        | 1.57  |
| Promotion                                 | 1.12   |       |
| Descriptive-notification                  | 1.11   |       |
| Multiselection                            | 1.57   |       |
| Confirm-successful-setup                  | 1.24   |       |
| Informed-2FA-recovery-options             | 1.26   |       |
| Informed-deactivation                     | 1.05   |       |
| Four-point-scale                          |        | 2.0   |
| Common-name-and-location                  | 1.00   |       |
| Device-remembrance                        | 1.60   |       |

$H(x) = 0$: Identical values

$H(x) = max$: Evenly split between values

Only the factors Non-optional and Common-name-and-location show high consistency across all websites.
Consistency Analysis: Pairwise Comparison

• Pairwise Hamming distance of non-conditional factors between websites
  – “Overlap without weights”

• 2FA user journeys are not very consistent across all 85 websites
  – Average website differs in 6–7 of 14 factors from the other websites
Clustering

• Are there clusters of websites with similar user journeys?

• Two-stage clustering process
  – Non-conditional factors: primary view of the websites’ strategies for 2FA UX
  – Conditional factors: Subcluster for a more differentiated view of these strategies
Clustering: Non-conditional factors

Cluster 1 ($n = 30$)
Cluster 2 ($n = 29$)
Cluster 3 ($n = 4$)
Cluster 4 ($n = 9$)
Cluster 5 ($n = 8$)
Cluster 6 ($n = 5$)

Verify 2FA settings changes and notify about them

Provide additional information about 2FA

Give step-wise setup instructions

Give option-specific information

Warn about risks of 2FA deactivation (only Cluster 4)
Clustering: Non-conditional factors

Cluster 1 (n = 30)
Cluster 2 (n = 29)
Cluster 3 (n = 4)
Cluster 4 (n = 9)
Cluster 5 (n = 8)
Cluster 6 (n = 5)

Allow multiple 2FA options to be activated simultaneously
Clustering: Non-conditional factors

Cluster 1 (n = 30)
Cluster 2 (n = 29)
Cluster 3 (n = 4)
Cluster 4 (n = 9)
Cluster 5 (n = 8)
Cluster 6 (n = 5)

Offer device remembrance
Sub-Clustering: Conditional factors

Subcluster 1 (n = 31)
No selection of multiple 2FA options or enforce a specific option

Subcluster 2 (n = 35)
Verify 2FA deactivation

Subcluster 3 (n = 19)
Do not enforce specific 2FA options
Combined Clusters

Factors
- Discovery
- Education
- Setup
- Usage
- Deactivation

Websites

Subclusters:
- Subcluster 1
- Subcluster 2
- Subcluster 3

Matches
Quasi matches
Does not match
Does not apply
Website Cluster versus Website Rank

• Divide websites by their Tranco rank in 3 equal-sized groups: Top-500, Top-4000, Long tail

• Normalized contingency table for cluster versus rank:

• Fisher’s exact test \( p = 0.04388 \) shows statistically significant association
Opinionated Separation of Comparison Factors

• Expert evaluation of our comparison factors to separate them by relevance: *Security, Usability, Both, None*

• Four disjoint sets of factors: *Non-conditional-UX* (7 factors), *Non-conditional-Security* (6 factors), *Conditional-UX* (5 factors), *Conditional-Security* (3 factors)

• Repeated consistency analysis and clustering
  – Pairwise Hamming distances: *no better consistency across all websites*
  – Clustering based on Silhouette coefficients: *more diverse strategies*
Qualitative Data Analysis

Consistent Lack of Informing and Educating Users

Only a minority of websites provided additional information ("learn more").

Most websites immediately start the 2FA setup process without informing users about the benefits/drawbacks of 2FA.

Only 1/3 of the websites provided step-by-step setup instructions but almost all confirm a successful setup.
Mixed Strategies for Device Remembrance

<50% of the websites support device remembrance

These websites describe this feature in different ways

≈2/3 offer the feature as opt-in, ≈1/5 offer as opt-out, ≈1/5 unsolicitedly places remembrance cookie (during login or even setup)