ECE750: Usable Security and Privacy

Survey Design

Dr. Kami Vaniea
Electrical and Computer Engineering
kami.vaniea@uwaterloo.ca

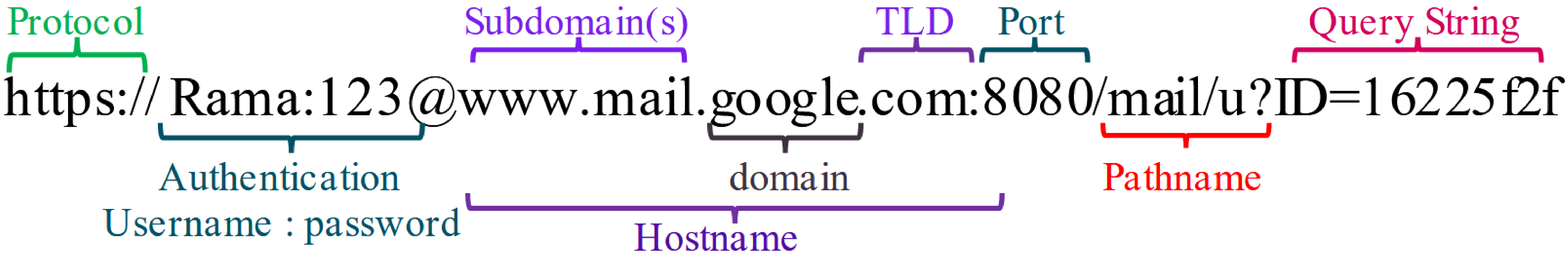
UNIVERSITY OF WATERLOO | FACULTY OF ENGINEERING

TULiPS
Technology Usability Lab in Privacy and Security

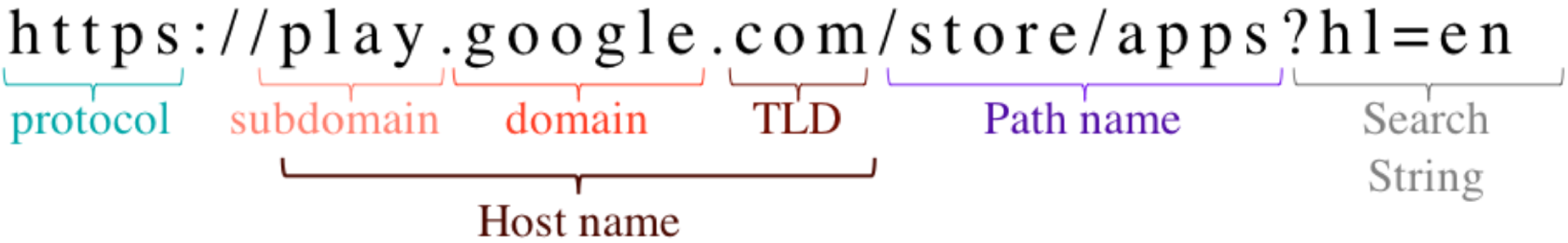
First, the news…

- First 5 minutes we talk about something interesting and recent
- You will not be tested on the news part of lecture
- You may use news as an example on tests
- Why do this?
  1. Some students show up late for various good reasons
  2. Reward students who show up on time
  3. Important to see real world examples

URLS



K. Althobaiti, G. Rummani, K. Vaniea. A Review of Human-and Computer-Facing URL Phishing Features. In IEEE European Symposium on Security and Privacy Workshops, 2019.



https://play.google.com/store/apps?hl=en

protocol

subdomain

domain

TLD

Path name

Search String

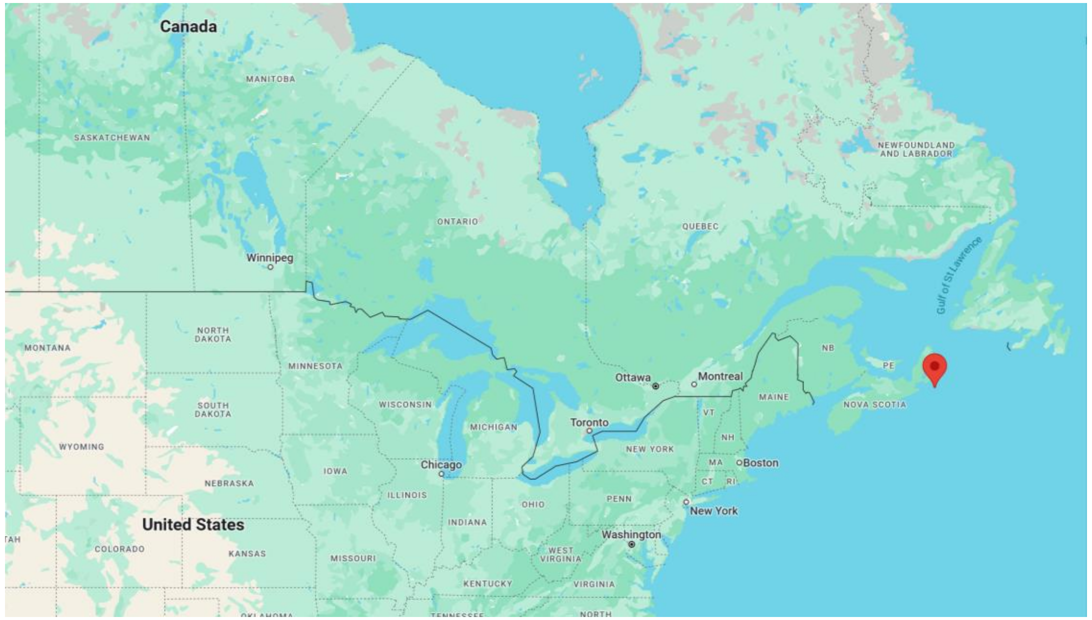
Host name

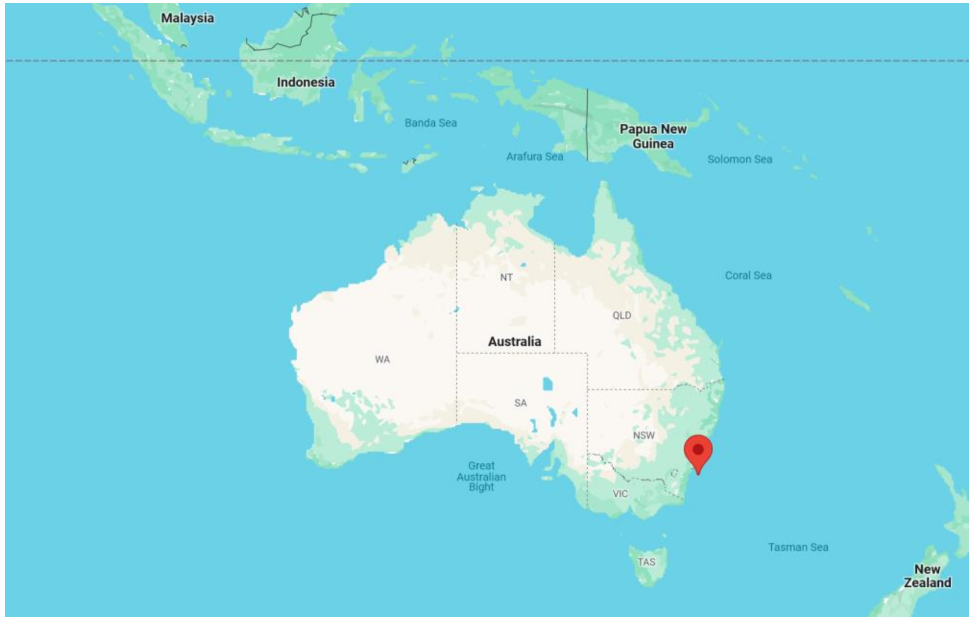
Like postal addresses, links are read right to left

https://facebook.mobile.com

Sydney, NS, Canada

Sydney, NSW, Australia





Which of these URLs goes to Facebook?

✗ https://facebook.profile.com

⬆

✓ **https://profile.facebook.com**

⬆

None of these go to Paypal

- paypal.com.login-myaccount.policy.country
- paypal.com.updates-information-accounts.ga
- paypal.com.account.update.amquipac.org
- paypal.com.login.summary-limited-account.gq
- paypal.com-websecure.limited
- paypal.com.resolution-ticket.tk
- www.update-paypal-informations-account.ga

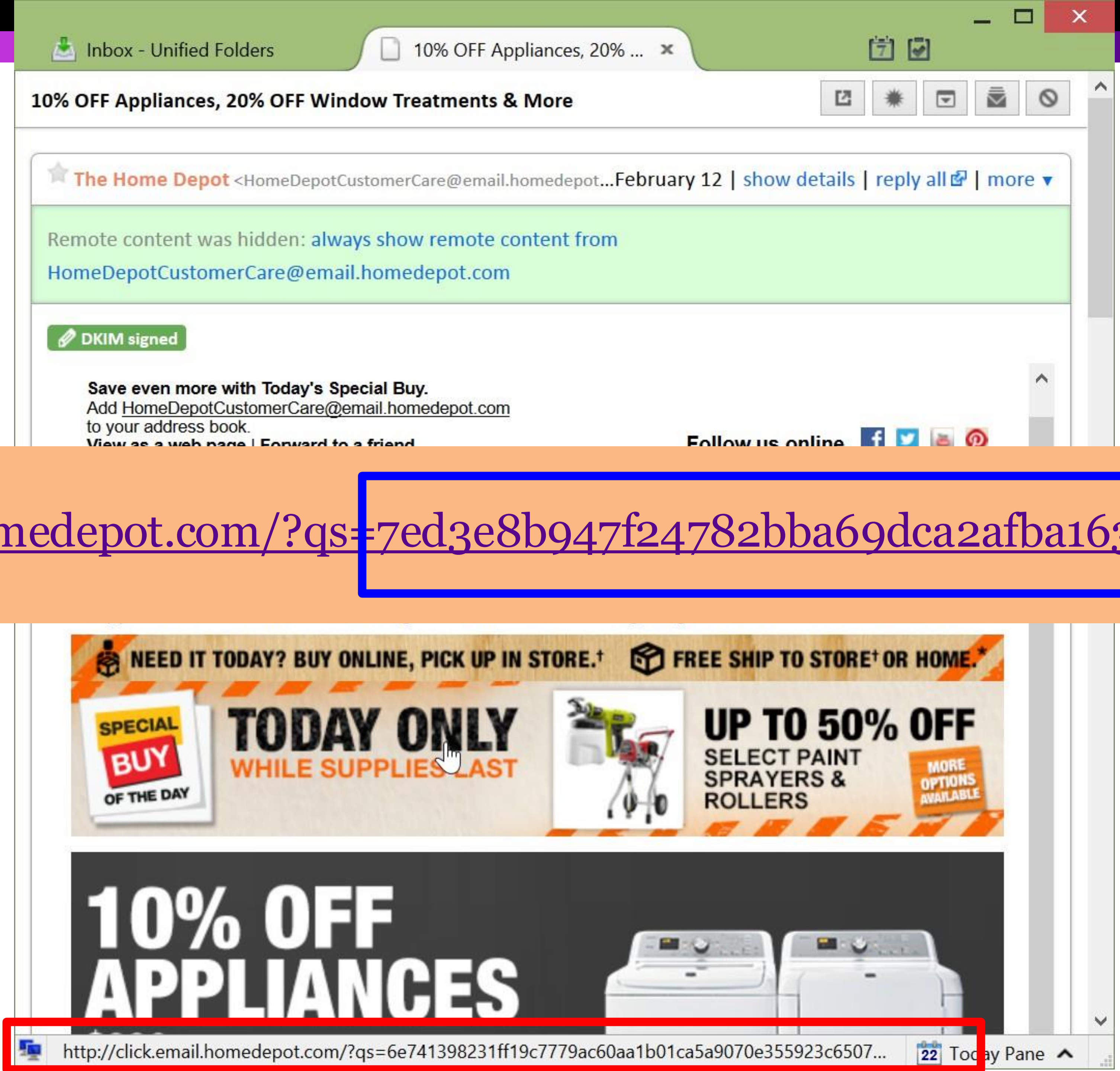
S.S. Albakry, K. Vaniea, M.K. Wolters; “What is this URL's Destination? Empirical Evaluation of Users' URL Reading”; In CHI 2020.

None of these go to Paypal

- paypal.com.login-myaccount.policy.country
- paypal.com.updates-information-accounts.ga
- paypal.com.account.update.amquipac.org
- paypal.com.login.summary-limited-account.gq
- paypal.com-websecure.limited
- paypal.com.resolution-ticket.tk
- www.update-paypal-informations-account.ga

S.S. Albakry, K. Vaniea, M.K. Wolters; “What is this URL's Destination? Empirical Evaluation of Users' URL Reading”; In CHI 2020.

Data aggregators use URLs to collect
detail
the e
clicke



Inbox - Unified Folders

10% OFF Appliances, 20% ...

10% OFF Appliances, 20% OFF Window Treatments & More

The Home Depot <HomeDepotCustomerCare@email.homedepot...February 12 | show details | reply all | more ▾

Remote content was hidden: always show remote content from HomeDepotCustomerCare@email.homedepot.com

DKIM signed

Save even more with Today's Special Buy.
Add HomeDepotCustomerCare@email.homedepot.com
to your address book.
View as a web page | Forward to a friend

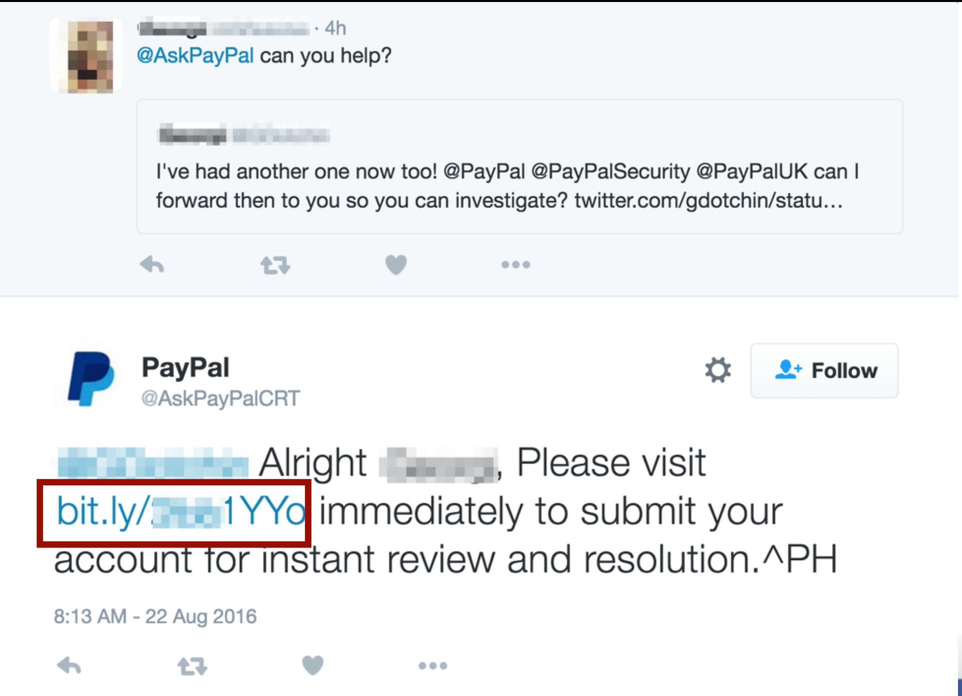
Follow us online

NEED IT TODAY? BUY ONLINE, PICK UP IN STORE.†
FREE SHIP TO STORE† OR HOME.*
SPECIAL BUY OF THE DAY
TODAY ONLY
WHILE SUPPLIES LAST
UP TO 50% OFF
SELECT PAINT SPRAYERS & ROLLERS
MORE OPTIONS AVAILABLE
10% OFF
APPLIANCES

http://click.email.homedepot.com/?qs=6e741398231ff19c7779ac60aa1b01ca5a9070e355923c6507...

Today Pane

http://click.email.homedepot.com/?qs=7ed3e8b947f24782bba69dca2afba163



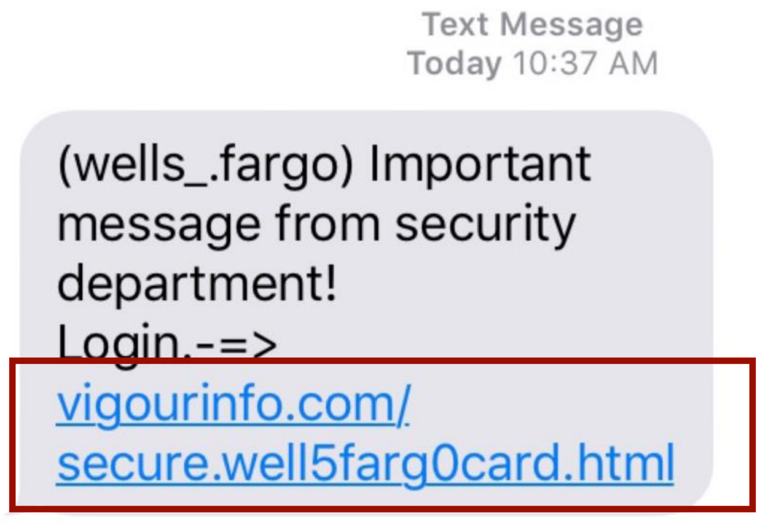
[redacted] · 4h
@AskPayPal can you help?

> [redacted]
> I've had another one now too! @PayPal @PayPalSecurity @PayPalUK can I forward then to you so you can investigate? twitter.com/gdotchin/statu…

PayPal
@AskPayPalCRT

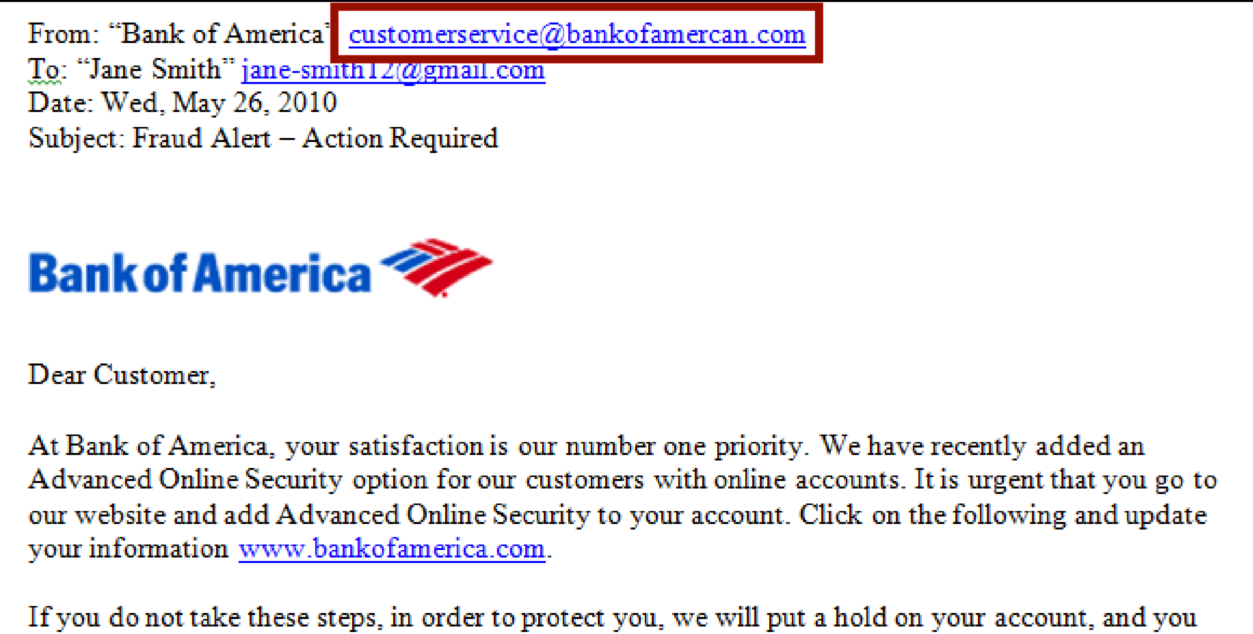
[redacted] Alright [redacted], Please visit bit.ly/[redacted]1YYo immediately to submit your account for instant review and resolution.^PH

8:13 AM - 22 Aug 2016



Text Message
Today 10:37 AM

(wells_.fargo) Important message from security department!
Login.-=>
vigourinfo.com/secure.well5farg0card.html

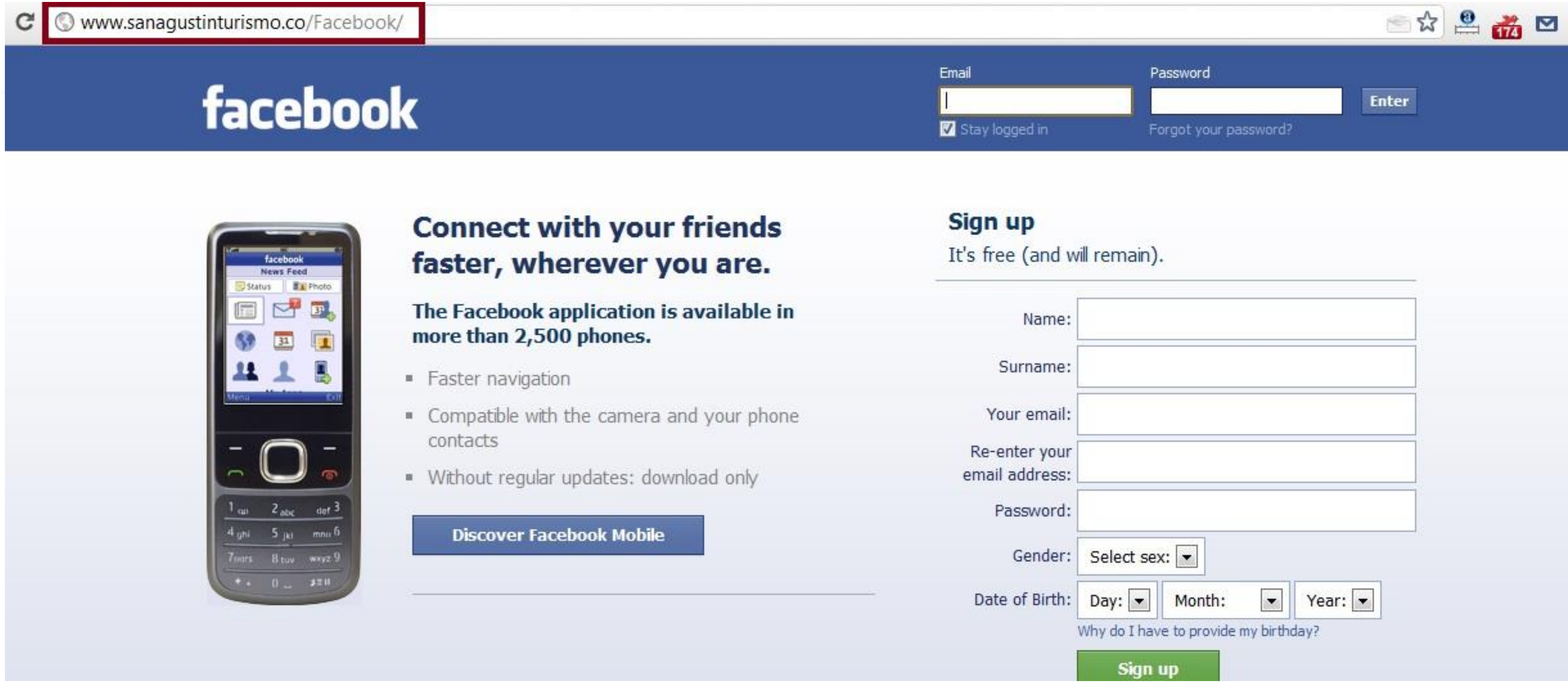


From: "Bank of America" customerservice@bankofamercan.com
To: "Jane Smith" jane-smith12@gmail.com
Date: Wed, May 26, 2010
Subject: Fraud Alert – Action Required

Dear Customer,

At Bank of America, your satisfaction is our number one priority. We have recently added an Advanced Online Security option for our customers with online accounts. It is urgent that you go to our website and add Advanced Online Security to your account. Click on the following and update your information www.bankofamerica.com.

If you do not take these steps, in order to protect you, we will put a hold on your account, and you



www.sanagustinturismo.co/Facebook/

Connect with your friends faster, wherever you are.

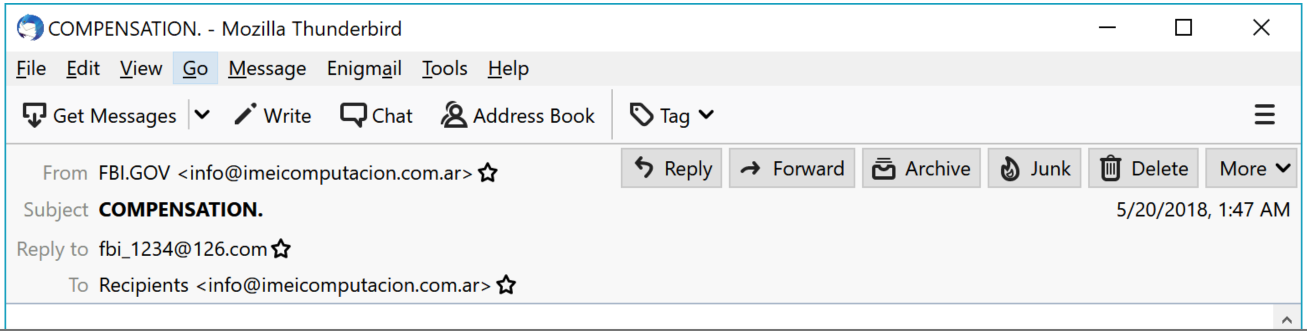
The Facebook application is available in more than 2,500 phones.

- Faster navigation
- Compatible with the camera and your phone contacts
- Without regular updates: download only

Sign up
It's free (and will remain).

Name:
Surname:
Your email:
Re-enter your email address:
Password:
Gender: Select sex:
Date of Birth: Day: Month: Year:
Why do I have to provide my birthday?

Quickly defining "phishing" so we can use it as an



From FBI.GOV <info@imeicomputacion.com.ar>

Subject **COMPENSATION.**

5/20/2018, 1:47 AM

Reply to fbi_1234@126.com

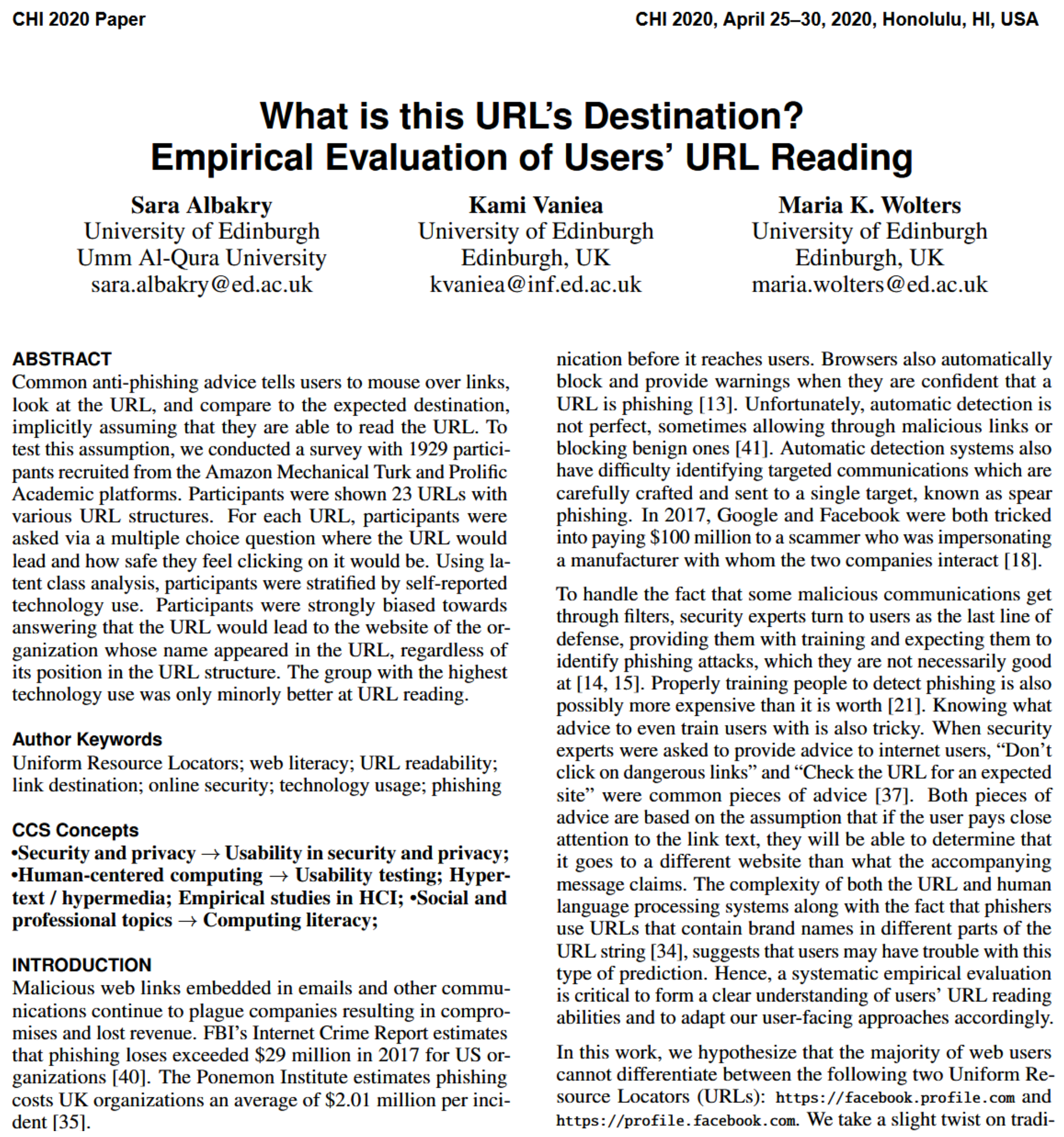
To Recipients <info@imeicomputacion.com.ar>

ATM Card: We will be issuing you a custom pin based ATM card which you will use to withdraw up to $3,000 per day from any ATM machine that has the Master Card Logo on it and the card have to be renewed in 3 years' time, which is 2021. Also with the ATM card you will be able to transfer your funds to your local bank account. The ATM card comes with a handbook or manual to enlighten you about how to use it, even if you do not have a bank account.

attempt to swindle your fund which has led to so many losses from your end and

Take note that anyone asking you for some kind of money above the usual fee is definitely a fraudster and you will have to stop communication with every other person, if you have been in contact with any. Also remember that all you will ever have to pay is $520 U.S Dollar, and we guarantee the receipt of your fund to be successfully delivered to you in four days, after the receipt of payment has been confirmed.

with the ATM card you will be able to transfer your funds to your local bank



What is this URL's Destination? Empirical Evaluation of Users' URL Reading

Sara Albakry
University of Edinburgh
Umm Al-Qura University
sara.albakry@ed.ac.uk

Kami Vaniea
University of Edinburgh
Edinburgh, UK
kvaniea@inf.ed.ac.uk

Maria K. Wolters
University of Edinburgh
Edinburgh, UK
maria.wolters@ed.ac.uk

ABSTRACT

Common anti-phishing advice tells users to mouse over links, look at the URL, and compare to the expected destination, implicitly assuming that they are able to read the URL. To test this assumption, we conducted a survey with 1929 participants recruited from the Amazon Mechanical Turk and Prolific Academic platforms. Participants were shown 23 URLs with various URL structures. For each URL, participants were asked via a multiple choice question where the URL would lead and how safe they feel clicking on it would be. Using latent class analysis, participants were stratified by self-reported technology use. Participants were strongly biased towards answering that the URL would lead to the website of the organization whose name appeared in the URL, regardless of its position in the URL structure. The group with the highest technology use was only minorly better at URL reading.

Author Keywords

Uniform Resource Locators; web literacy; URL readability; link destination; online security; technology usage; phishing

CCS Concepts

•Security and privacy → Usability in security and privacy; •Human-centered computing → Usability testing; Hypertext / hypermedia; Empirical studies in HCI; •Social and professional topics → Computing literacy;

INTRODUCTION

Malicious web links embedded in emails and other communications continue to plague companies resulting in compromises and lost revenue. FBI's Internet Crime Report estimates that phishing loses exceeded $29 million in 2017 for US organizations [40]. The Ponemon Institute estimates phishing costs UK organizations an average of $2.01 million per incident [35].

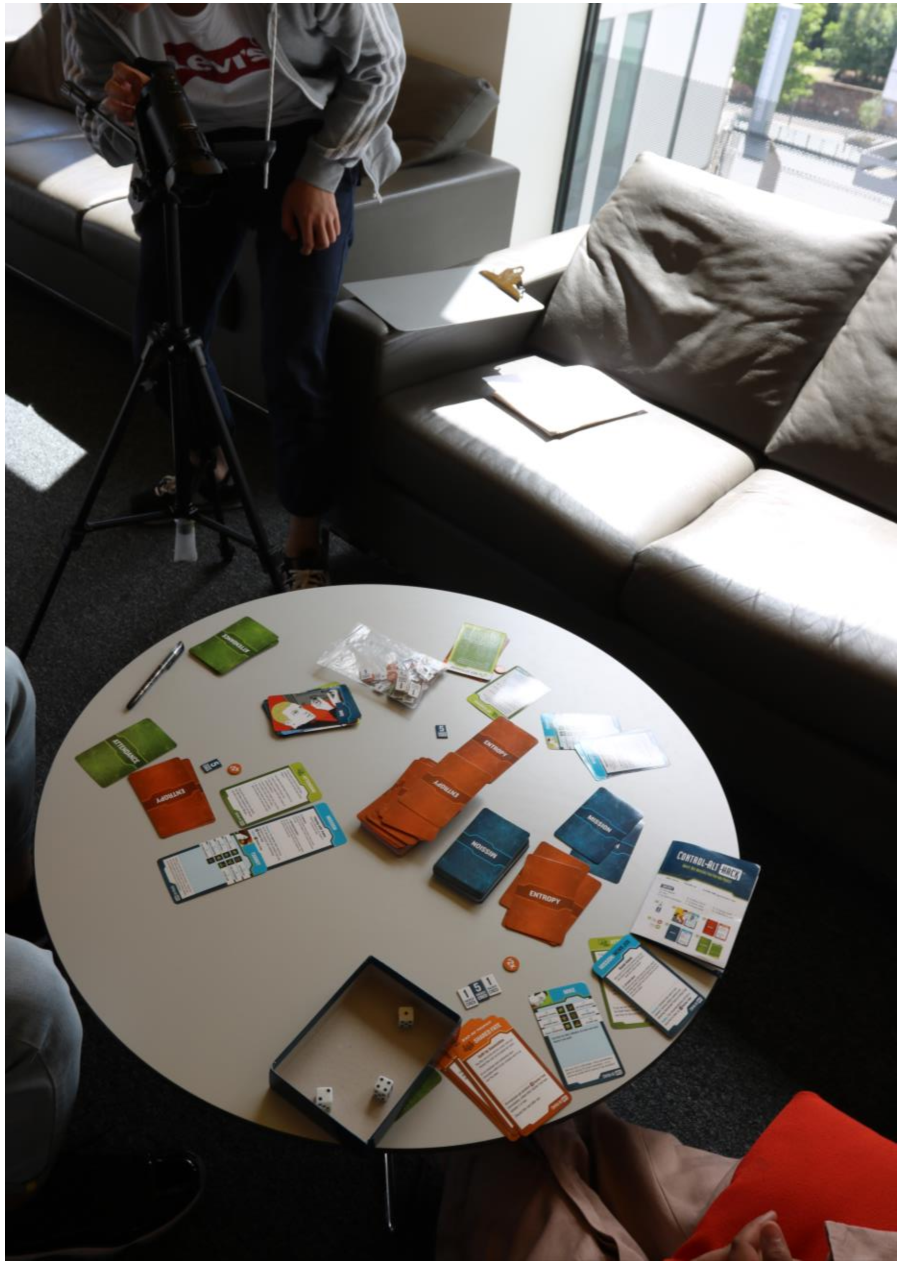
nication before it reaches users. Browsers also automatically block and provide warnings when they are confident that a URL is phishing [13]. Unfortunately, automatic detection is not perfect, sometimes allowing through malicious links or blocking benign ones [41]. Automatic detection systems also have difficulty identifying targeted communications which are carefully crafted and sent to a single target, known as spear phishing. In 2017, Google and Facebook were both tricked into paying $100 million to a scammer who was impersonating a manufacturer with whom the two companies interact [18].

To handle the fact that some malicious communications get through filters, security experts turn to users as the last line of defense, providing them with training and expecting them to identify phishing attacks, which they are not necessarily good at [14, 15]. Properly training people to detect phishing is also possibly more expensive than it is worth [21]. Knowing what advice to even train users with is also tricky. When security experts were asked to provide advice to internet users, "Don't click on dangerous links" and "Check the URL for an expected site" were common pieces of advice [37]. Both pieces of advice are based on the assumption that if the user pays close attention to the link text, they will be able to determine that it goes to a different website than what the accompanying message claims. The complexity of both the URL and human language processing systems along with the fact that phishers use URLs that contain brand names in different parts of the URL string [34], suggests that users may have trouble with this type of prediction. Hence, a systematic empirical evaluation is critical to form a clear understanding of users' URL reading abilities and to adapt our user-facing approaches accordingly.

In this work, we hypothesize that the majority of web users cannot differentiate between the following two Uniform Resource Locators (URLs): `https://facebook.profile.com` and `https://profile.facebook.com`. We take a slight twist on tradi-

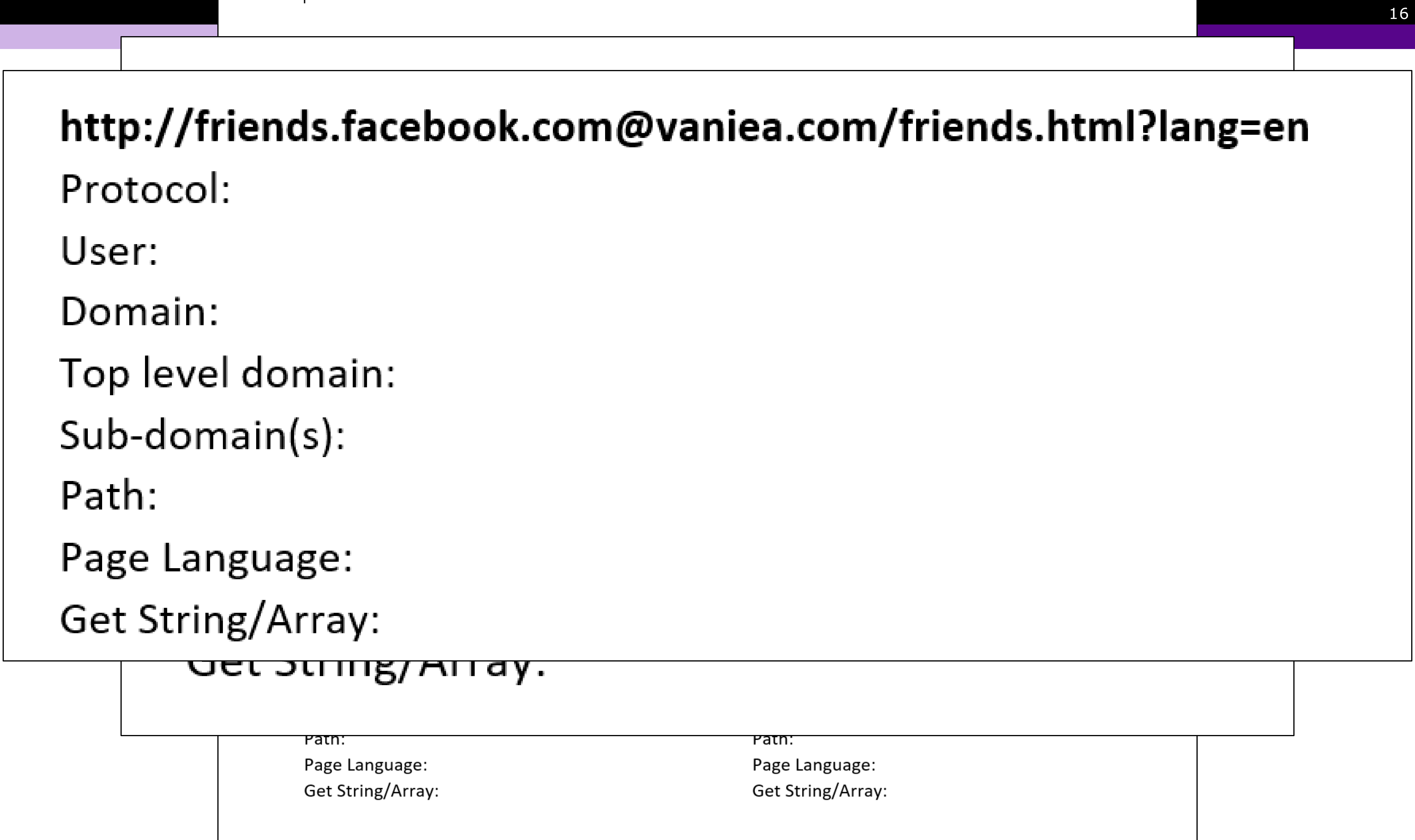
Structuring Research

- **Research question or goal**
- Literature review (what have others learned or done)
- Methods planned to answer question or achieve goal
- Evaluate outcome
- Contextualize findings
- Writeup



Inspiration

- I decided to run a “fun” worksheet on URLs with my class that could only sorta code.
- They could not even answer the first question.
- I then started noticing how often URL reading ability is assumed in safety training.



http://friends.facebook.com@vaniea.com/friends.html?lang=en

Protocol:

User:

Domain:

Top level domain:

Sub-domain(s):

Path:

Page Language:

Get String/Array:

Get String/Array:

Path:

Page Language:

Get String/Array:

Path:

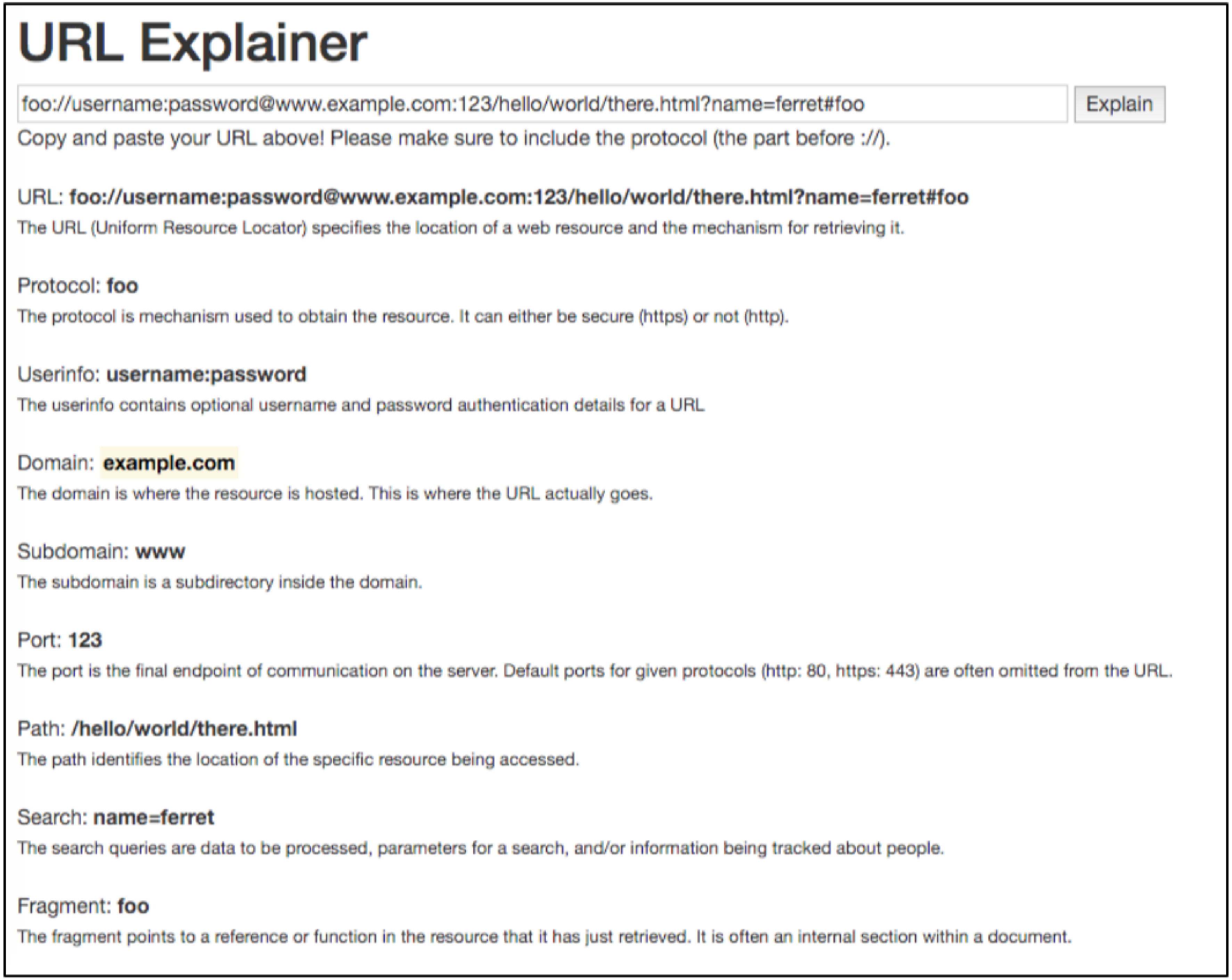
Page Language:

Get String/Array:

URL Explainer

Serena Zheng

Project by UG3 visiting student, Serena Zheng, to break up and explain a URL to someone.



URL Explainer

foo://username:password@www.example.com:123/hello/world/there.html?name=ferret#foo

Copy and paste your URL above! Please make sure to include the protocol (the part before ://).

URL: **foo://username:password@www.example.com:123/hello/world/there.html?name=ferret#foo**

The URL (Uniform Resource Locator) specifies the location of a web resource and the mechanism for retrieving it.

Protocol: **foo**

The protocol is mechanism used to obtain the resource. It can either be secure (https) or not (http).

Userinfo: **username:password**

The userinfo contains optional username and password authentication details for a URL

Domain: **example.com**

The domain is where the resource is hosted. This is where the URL actually goes.

Subdomain: **www**

The subdomain is a subdirectory inside the domain.

Port: **123**

The port is the final endpoint of communication on the server. Default ports for given protocols (http: 80, https: 443) are often omitted from the URL.

Path: **/hello/world/there.html**

The path identifies the location of the specific resource being accessed.

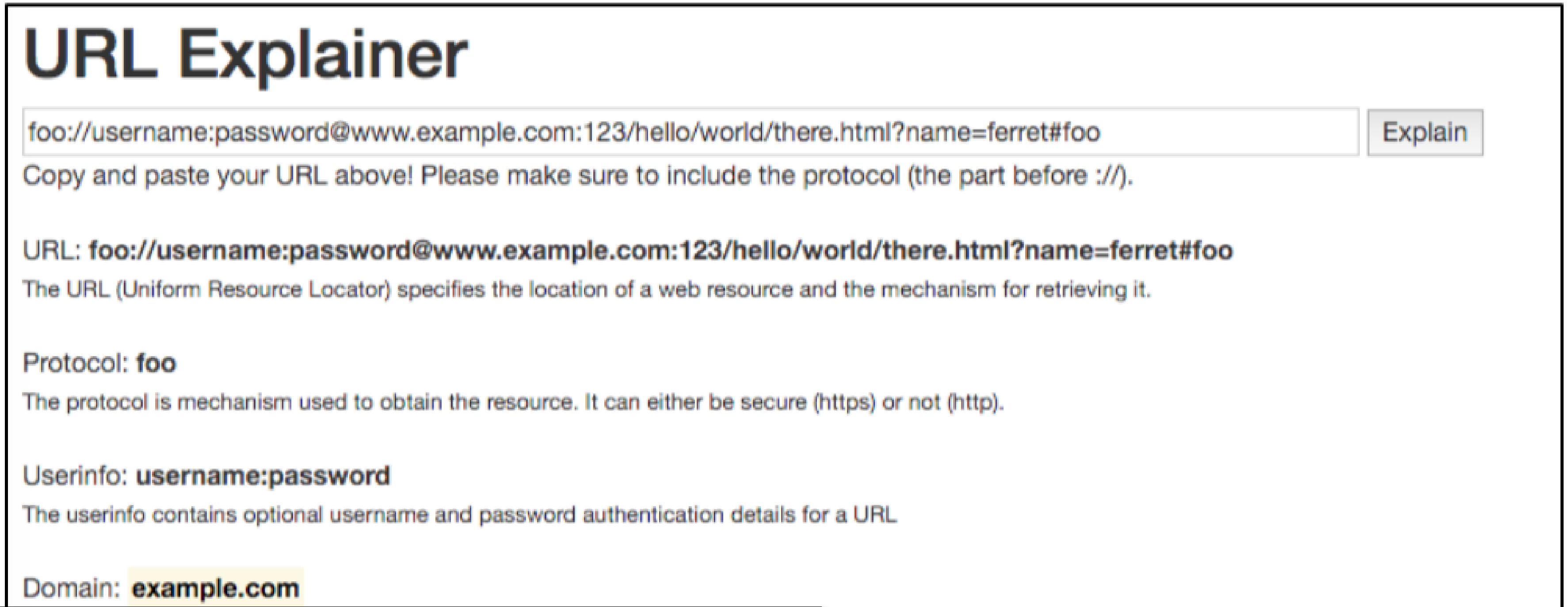
Search: **name=ferret**

The search queries are data to be processed, parameters for a search, and/or information being tracked about people.

Fragment: **foo**

The fragment points to a reference or function in the resource that it has just retrieved. It is often an internal section within a document.

Project by UG3 visiting student, Serena Zheng, to break up and explain a URL to someone.



URL Explainer

foo://username:password@www.example.com:123/hello/world/there.html?name=ferret#foo

Copy and paste your URL above! Please make sure to include the protocol (the part before ://).

URL: **foo://username:password@www.example.com:123/hello/world/there.html?name=ferret#foo**

The URL (Uniform Resource Locator) specifies the location of a web resource and the mechanism for retrieving it.

Protocol: **foo**

The protocol is mechanism used to obtain the resource. It can either be secure (https) or not (http).

Userinfo: **username:password**

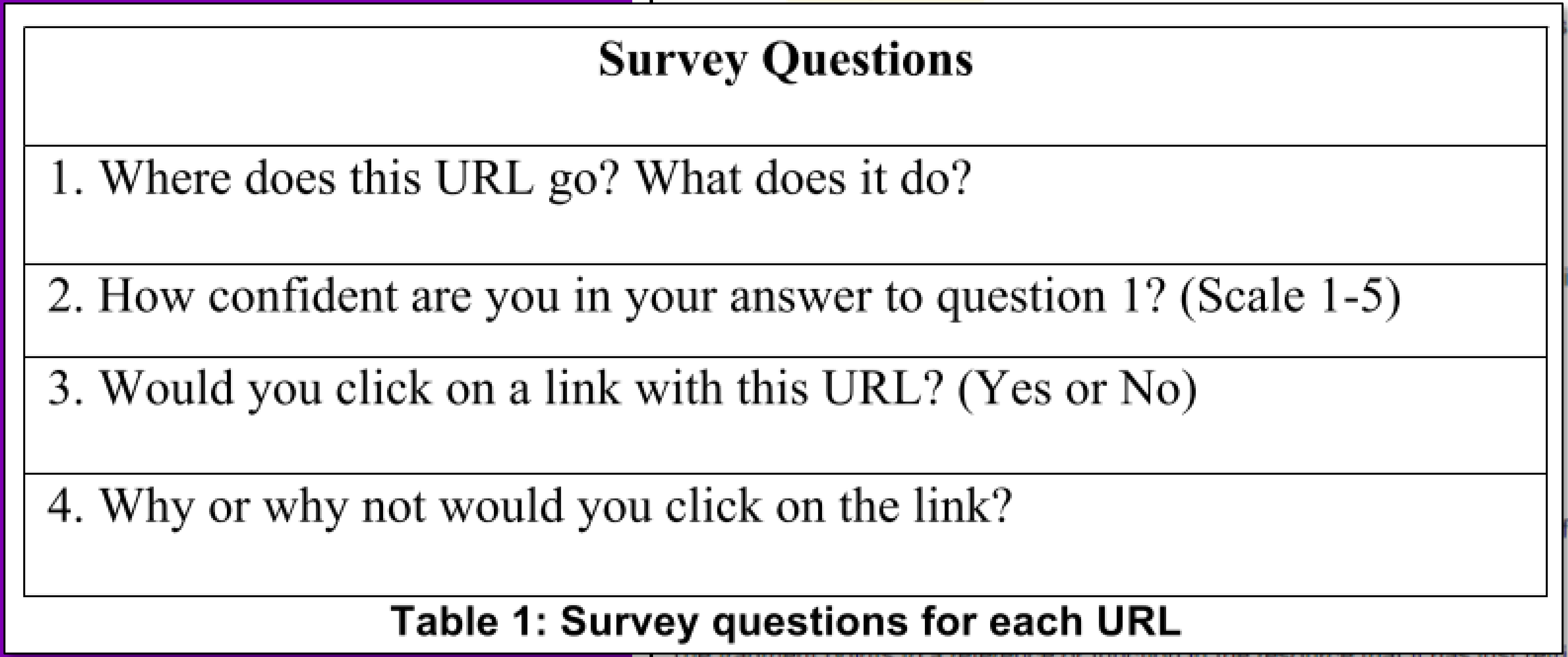
The userinfo contains optional username and password authentication details for a URL

Domain: **example.com**

iven protocols (http: 80, https: 443) are often omitted from the URL.

formation being tracked about people.

The fragment points to a reference or function in the resource that it has just retrieved. It is often an internal section within a document.



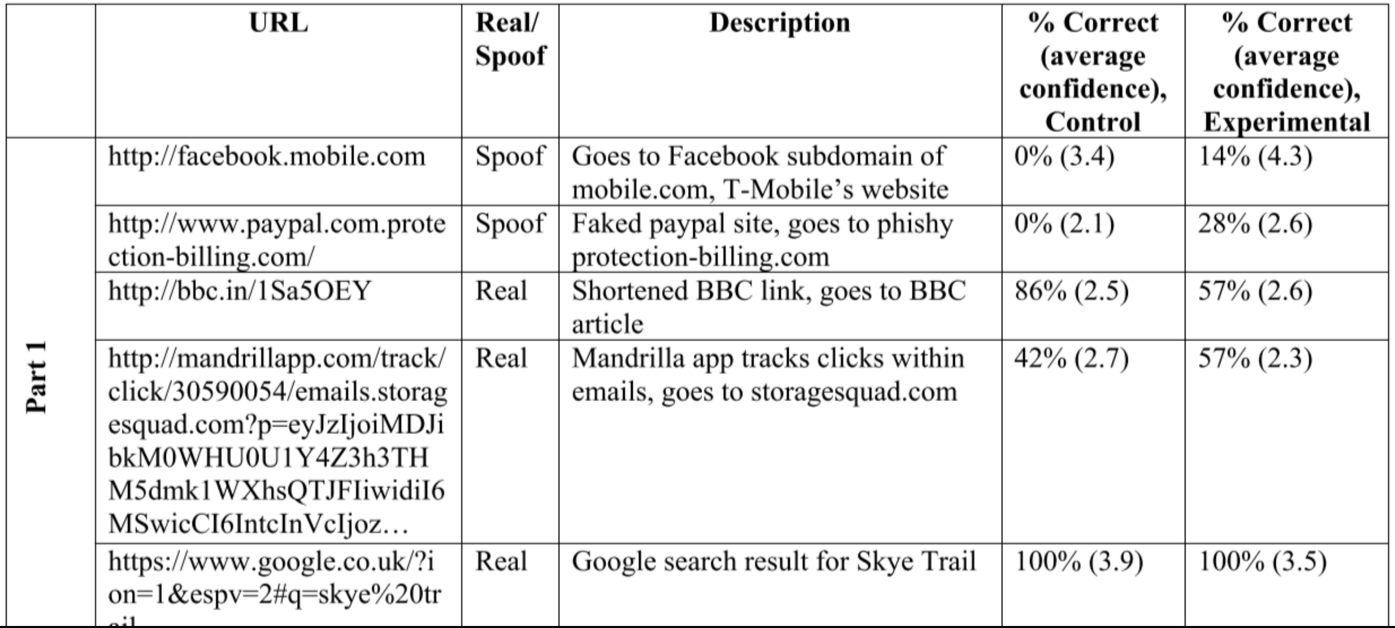
Survey Questions
1. Where does this URL go? What does it do?
2. How confident are you in your answer to question 1? (Scale 1-5)
3. Would you click on a link with this URL? (Yes or No)
4. Why or why not would you click on the link?

Table 1: Survey questions for each URL

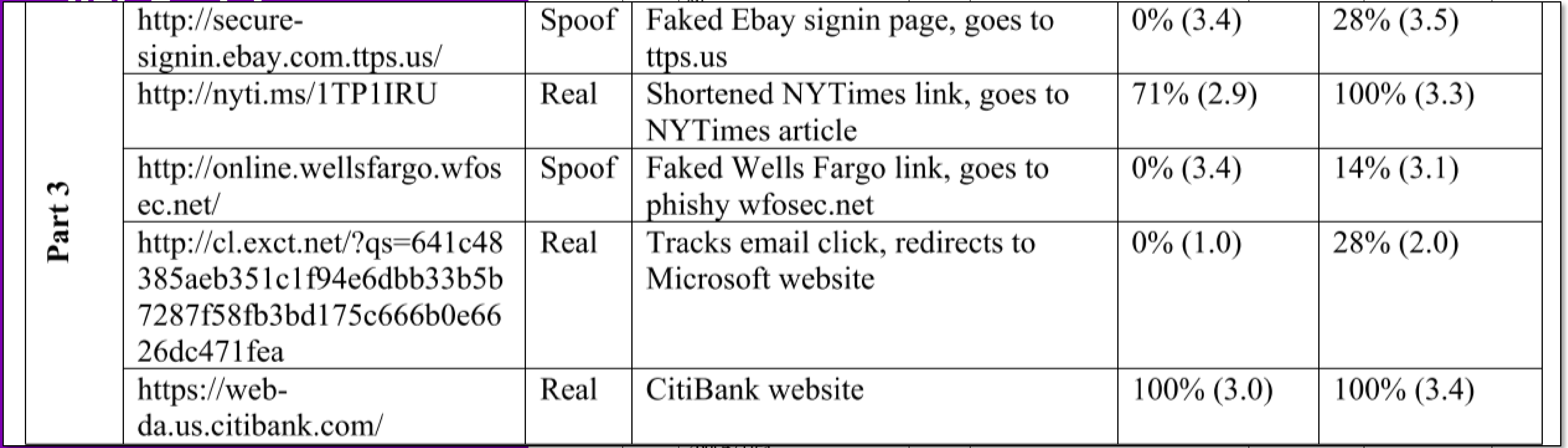
Test URL reading

Part 1 – reading without assistance

Part 2 – reading with



	URL	Real/ Spoof	Description	% Correct (average confidence), Control	% Correct (average confidence), Experimental
Part 1	http://facebook.mobile.com	Spoof	Goes to Facebook subdomain of mobile.com, T-Mobile's website	0% (3.4)	14% (4.3)
	http://www.paypal.com.protection-billing.com/	Spoof	Faked paypal site, goes to phishy protection-billing.com	0% (2.1)	28% (2.6)
	http://bbc.in/1Sa5OEY	Real	Shortened BBC link, goes to BBC article	86% (2.5)	57% (2.6)
	http://mandrillapp.com/track/click/30590054/emails.storagesquad.com?p=eyJzIjoiMDJibkM0WHU0U1Y4Z3h3THM5dmk1WXhsQTJFIiwidiI6MSwicCI6IntcInVcIjoz…	Real	Mandrilla app tracks clicks within emails, goes to storagesquad.com	42% (2.7)	57% (2.3)
	https://www.google.co.uk/?ion=1&espv=2#q=skye%20tr	Real	Google search result for Skye Trail	100% (3.9)	100% (3.5)



Part 3	http://secure-signin.ebay.com.ttps.us/	Spoof	Faked Ebay signin page, goes to ttps.us	0% (3.4)	28% (3.5)
	http://nyti.ms/1TP1IRU	Real	Shortened NYTimes link, goes to NYTimes article	71% (2.9)	100% (3.3)
	http://online.wellsfargo.wfosec.net/	Spoof	Faked Wells Fargo link, goes to phishy wfosec.net	0% (3.4)	14% (3.1)
	http://cl.exct.net/?qs=641c48385aeb351c1f94e6dbb33b5b7287f58fb3bd175c666b0e6626dc471fea	Real	Tracks email click, redirects to Microsoft website	0% (1.0)	28% (2.0)
	https://web-da.us.citibank.com/	Real	CitiBank website	100% (3.0)	100% (3.4)

26dc471fea

https://web-da.us.citibank.com/ | Real | CitiBank website | 100% (3.0) | 100% (3.4)

Test URL reading

Part 1 – reading without assistance

Part 2 – reading with URL Explainer support (experimental)

Part 3 – reading without assistance

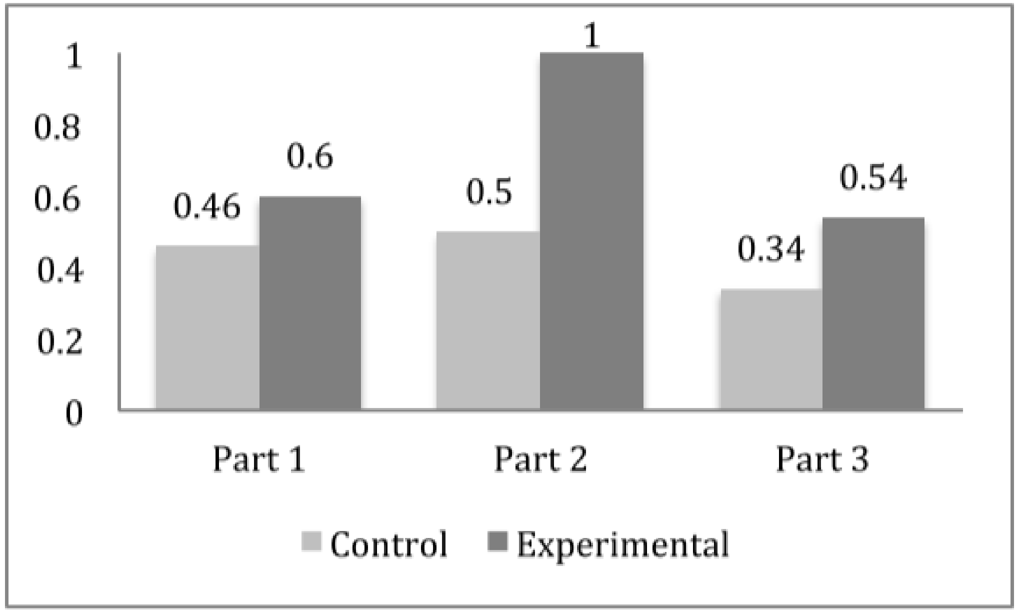


Figure 4b: Average correctness (%) for each part of the survey

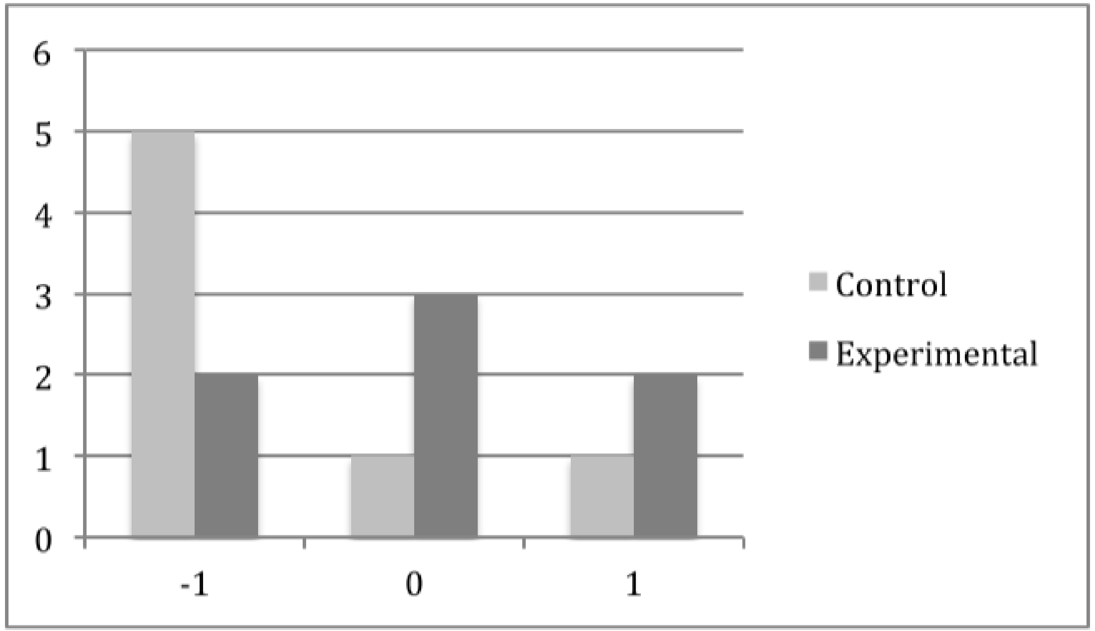


Figure 6: Change in confidence levels in reading URLs after taking the survey

Observation

Users tended to always select the recognizable organization name in the URL, even if it is in the subdomain.

Research Question

Informal RQ:

- **Can people read URLs under optimal conditions?**

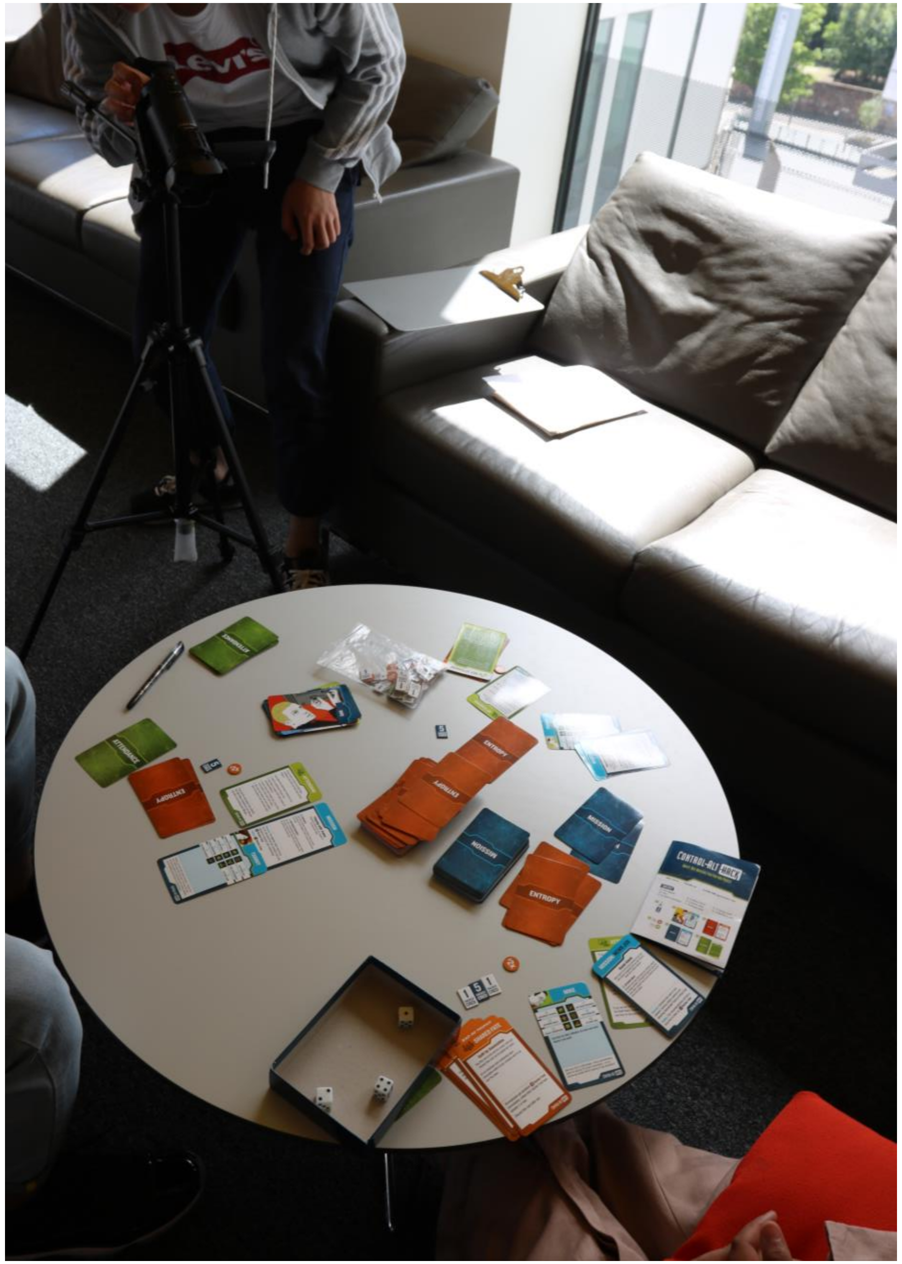
Formal RQs:

- RQ1 Can users accurately predict where a URL will go?
  - RQ1.1 Can users correctly infer from the URL that it will go to the website of the organization listed in the domain position rather than the subdomain, and what factors affect prediction accuracy?
  - RQ1.2 Can users recognize that the end destination of shortened URLs is not easy to predict?
  - RQ1.3 Can users recognize the end destination of complex URL structures?
- RQ2 What effects users' assessment of the likely safety of a URL?

S. S. Albakry, K. Vaniea, M. K. Wolters. What is this URL's Destination? Empirical Evaluation of Users' URL Reading. In CHI 2020.

Structuring Research

- Research question or goal
- **Literature review (what have others learned or done)**
- Methods planned to answer question or achieve goal
- Evaluate outcome
- Contextualize findings
- Writeup



Very little research on URL reading

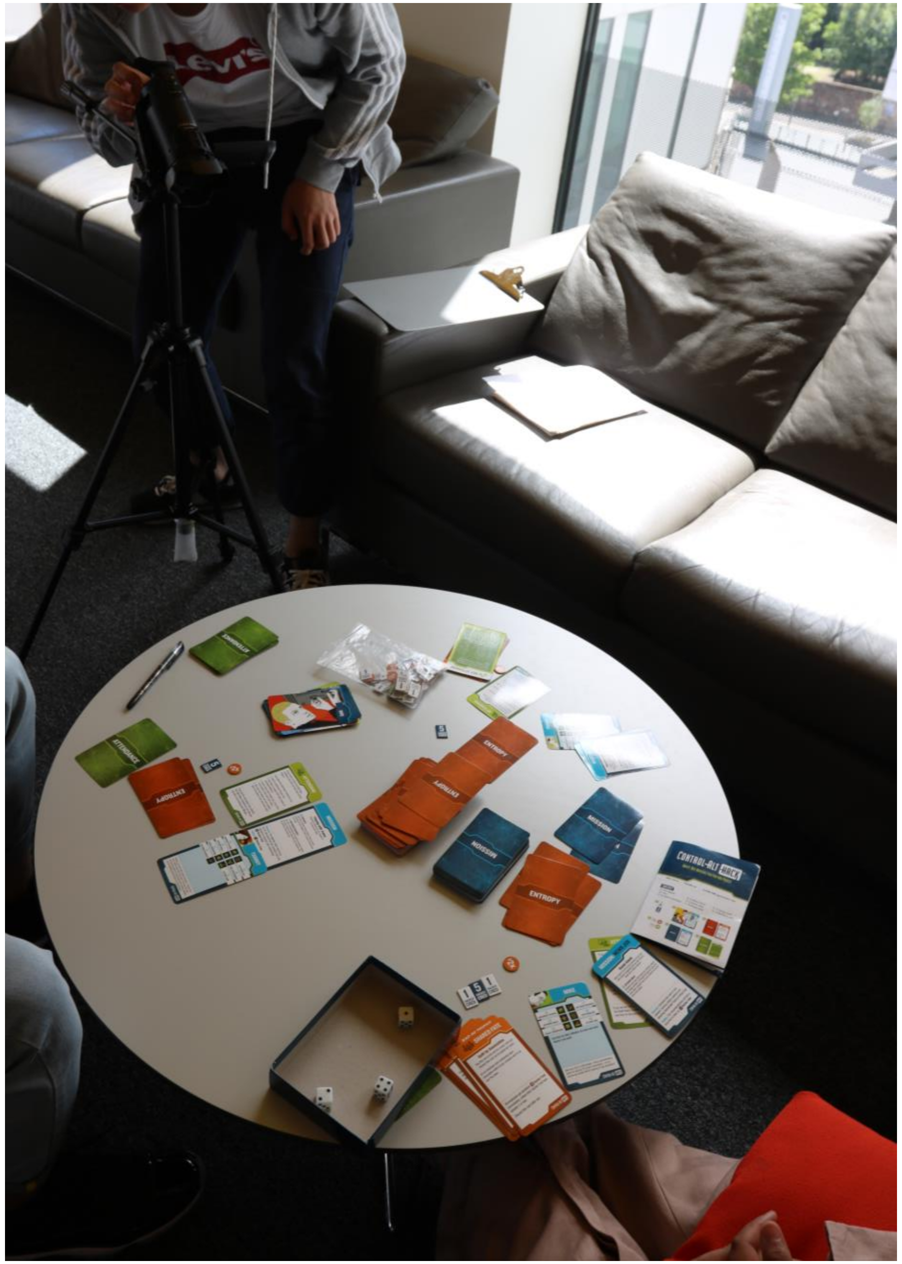
- Lots of research says users cannot detect phishing URLs.
- URLs commonly taught in larger training programs.
  - Advice like: Look at the URL to see if it is going to the correct place.
- Lots of research on common manipulation tactics.

- I also asked Ross Anderson and other senior people.
- Tim Berners-Lee (URL inventor) also had an opinion piece saying how much he regretted its structure.

S. S. Albakry, K. Vaniea, M. K. Wolters. What is this URL's Destination? Empirical Evaluation of Users' URL Reading. In CHI 2020.

Structuring Research

- Research question or goal
- Literature review (what have others learned or done)
- **Methods planned to answer question or achieve goal**
- Evaluate outcome
- Contextualize findings
- Writeup



Method

Online survey (**Total = 1929**)

- Amazon Mechanical Turk (n= 972)
- Prolific Academic (n=962)

- Advertised as "**Opinions on Web links**", advertisement and survey did not mention privacy or security.
- Survey consisted of:

1. Instructions and Training
2. Set of **23 URLs** presented in random order
3. Demographics

- Confounds considered
  - HTTPS – some users think the "s" stands for security and might erroneously use it as an indicator.
  - TLD – There are many top level domains. So .com was used to limit confounds.
  - Real URLs – Only real URLs were used as a base.
  - Recognizable names – Only URLs that contain the real company's name.

S. S. Albakry, K. Vaniea, M. K. Wolters. What is this URL's Destination? Empirical Evaluation of Users' URL Reading. In CHI 2020.

Confound

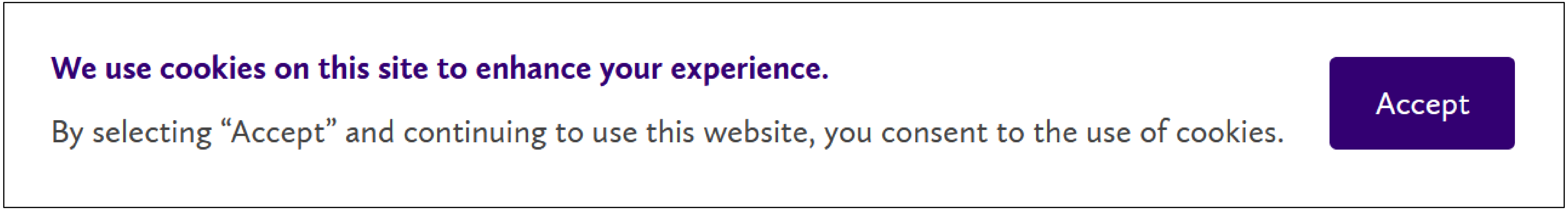
- An aspect of a study that may impact the study outcome in an unwanted way.
- A confound can mean that something other than the intended manipulation is the cause of the results.
- We control for confounds through careful study design.

Confound examples

- Saying something about security before the study – bringing security to the participant’s attention.
- Tested manipulations differ in more than the intended way.
- Something happened outside the lab, like a large data breach.

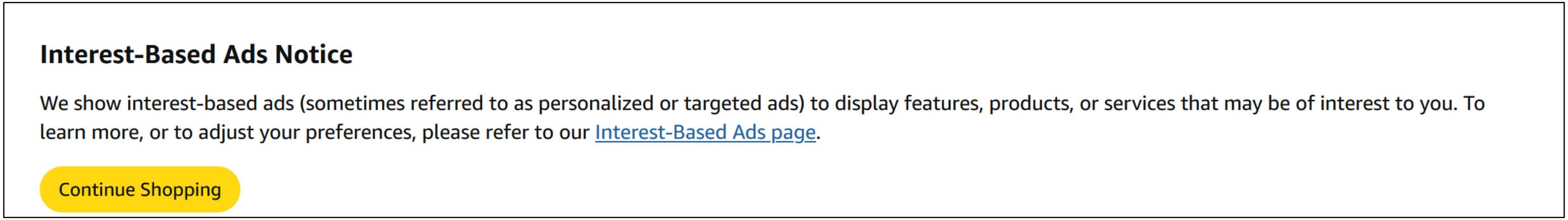
Confound

- Pretend I tested the following two cookie dialogs against each other and found that the second one leads to more people opting out.
- What likely caused the effect?



We use cookies on this site to enhance your experience.

By selecting "Accept" and continuing to use this website, you consent to the use of cookies.



Interest-Based Ads Notice

We show interest-based ads (sometimes referred to as personalized or targeted ads) to display features, products, or services that may be of interest to you. To learn more, or to adjust your preferences, please refer to our Interest-Based Ads page.

Correlation vs. Causation

- Correlation
  - Two things tend to behave in a way that seems inter-related, where if one thing changes the other thing will also change in a related way.
  - For example, if the price of rice goes up at the same time as the price for beans.
- Causation
  - When one thing changes it causes the other thing to change.
  - For example, when the weather gets cold more people wear coats. Cold weather causes more people to wear coats.

Does consuming chocolate increase the number of Nobel Laureates?

This is a correlation, not necessarily a causation.

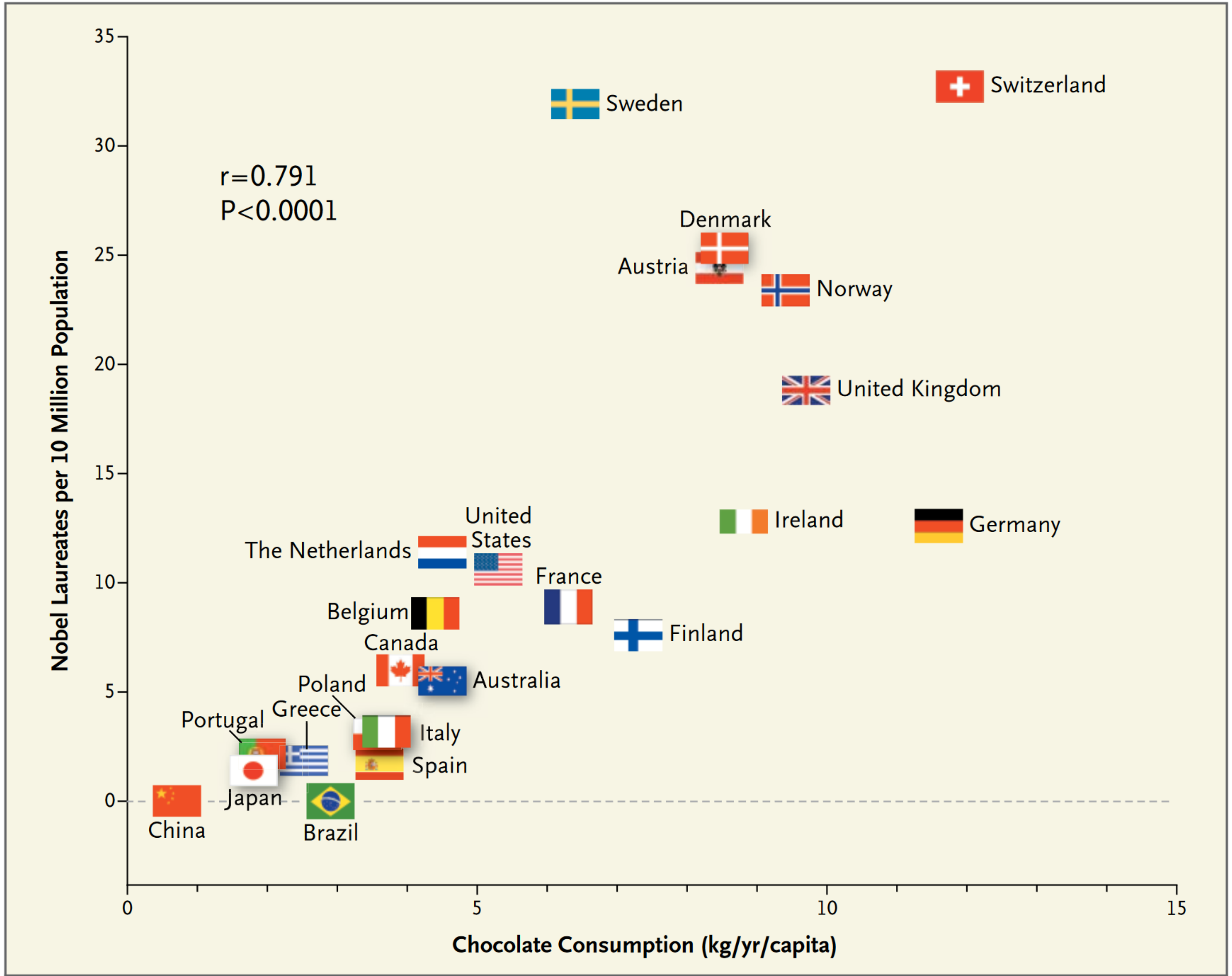
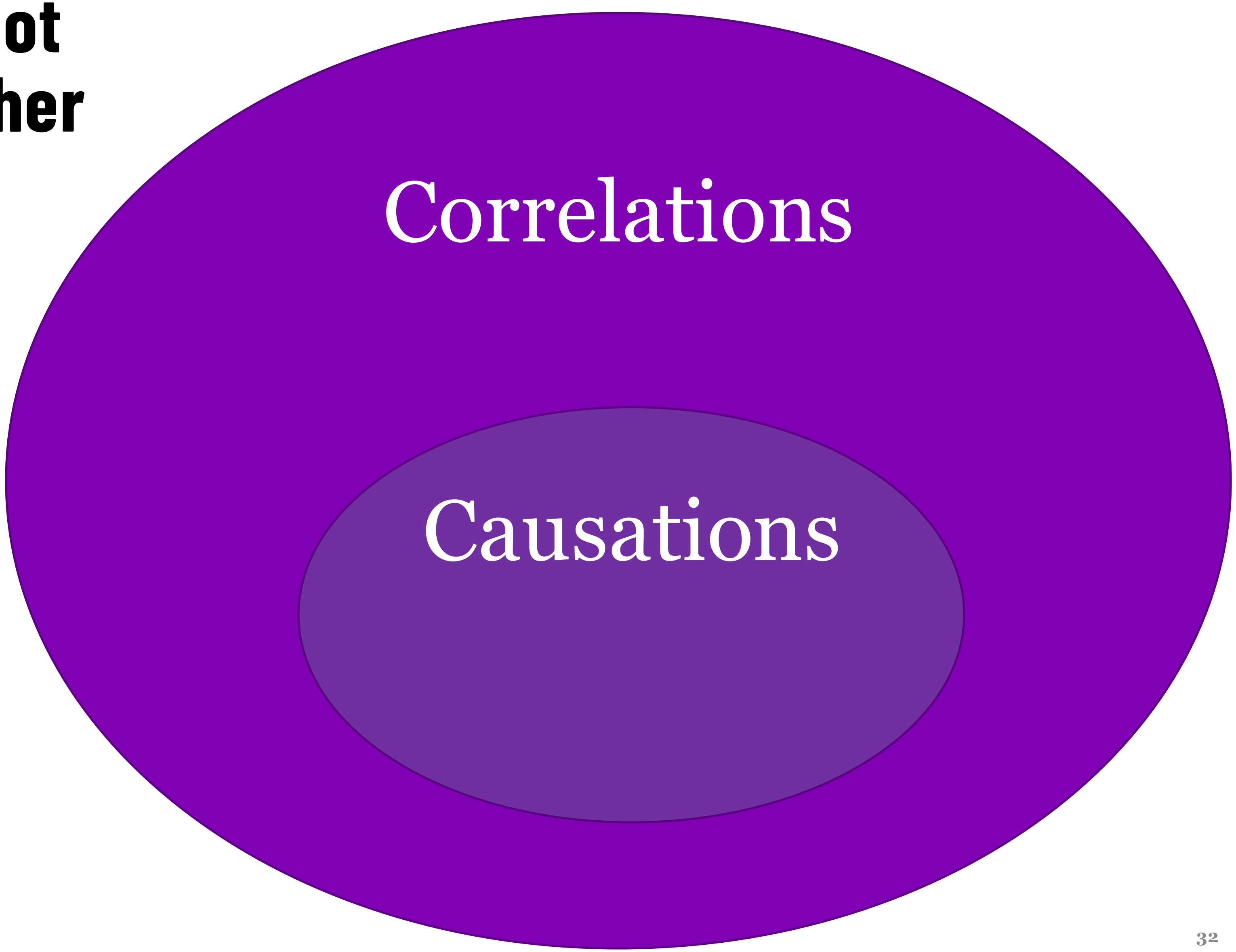


Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

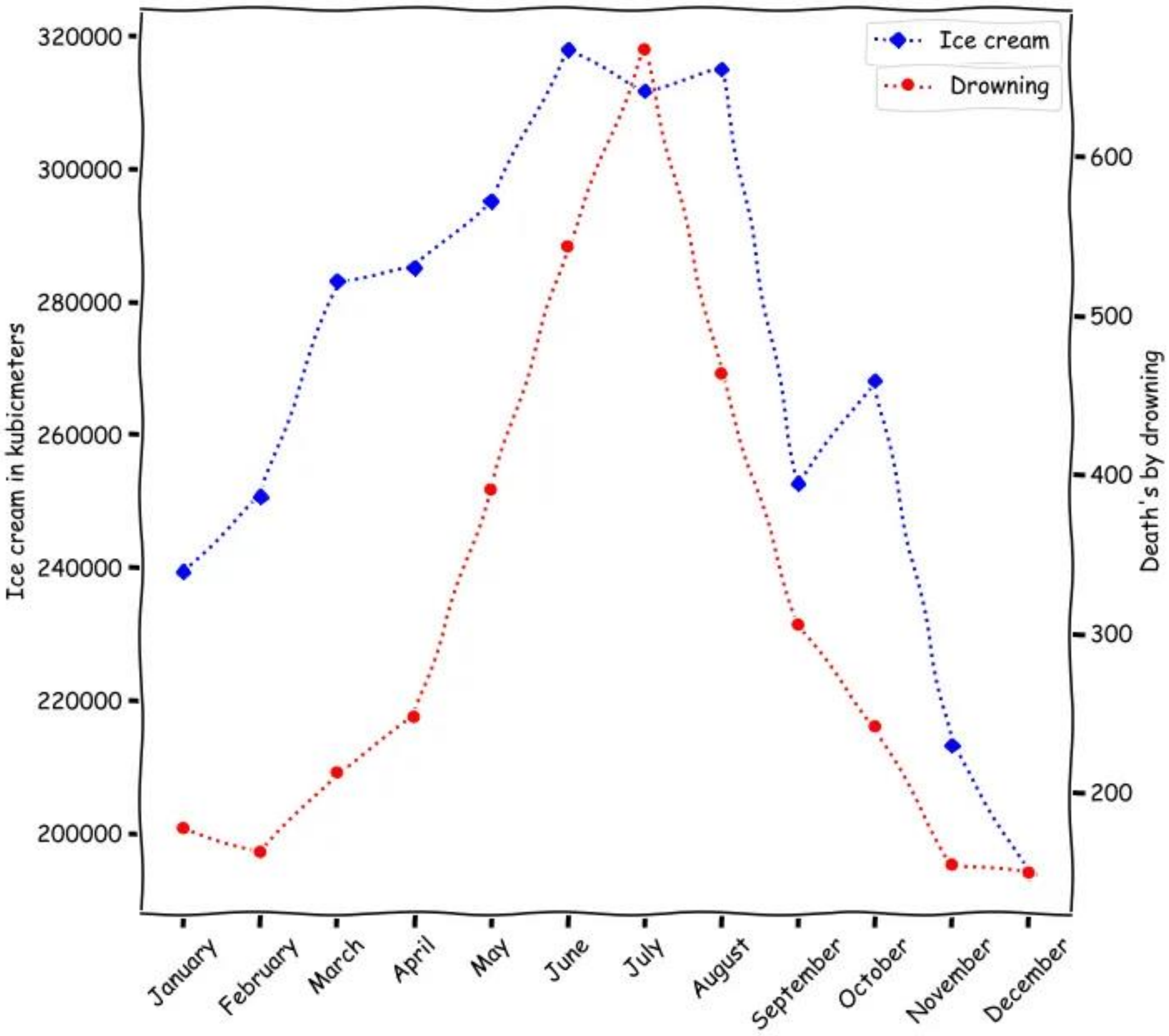
Chocolate Consumption, Cognitive Function,
and Nobel Laureates
Franz H. Messerli, M.D.

Causations can be Correlations, but not necessarily the other way round



Ice Cream vs Drowning

- Drowning deaths go up at roughly the same time of year as ice cream consumption goes up
- Does ice cream consumption cause drownings? Of course not.
- Both ice cream consumption and drowning are caused by the weather getting warmer
- They are correlated.
- There is a latent unmeasured variable (temperature) that causes them to rise and fall together



https://andreasrmadsen.medium.com/a-story-of-ice-cream-drowning-and-causal-modelling-fff3967f7671

Method

Online survey (**Total = 1929**)

- Amazon Mechanical Turk (n= 972)
- Prolific Academic (n=962)

- Advertised as “**Opinions on Web links**”, advertisement and survey did not mention privacy or security.
- Survey consisted of:

1. Instructions and Training
2. Set of **23 URLs** presented in random order
3. Demographics

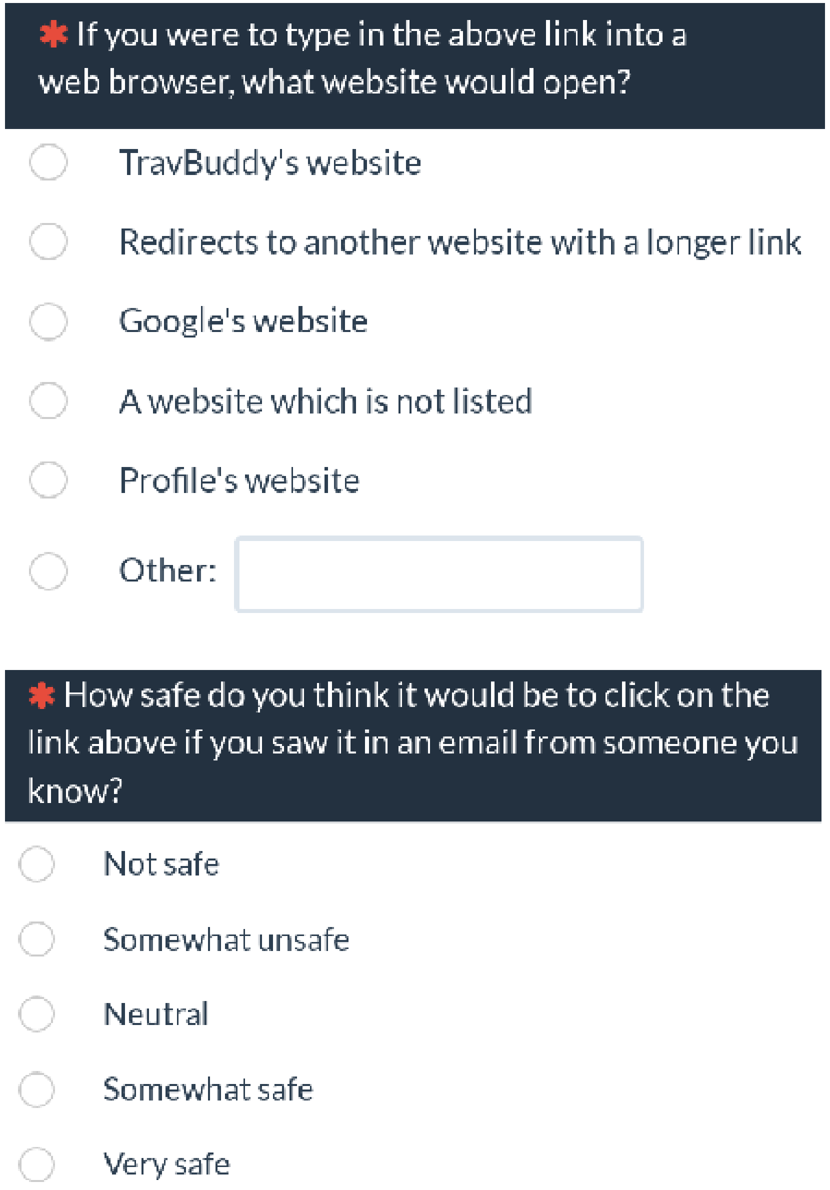
- Confounds considered
  - HTTPS – some users think the “s” stands for security and might erroneously use it as an indicator.
  - TLD – There are many top level domains. So .com was used to limit confounds.
  - Real URLs – Only real URLs were used as a base.
  - Recognizable names – Only URLs that contain the real company’s name.

Survey Overview

We tested four variations of URLs:

Name	Example
Domain only	https://microsoft.com
Subdomain	https://profile.facebook.com
Complex	https://facebook.com/picture.html?a=twitter.com
Short	https://bit.ly/1bdDlXc

https://profile.travbuddy.com



✱ If you were to type in the above link into a web browser, what website would open?

○ TravBuddy's website

○ Redirects to another website with a longer link

○ Google's website

○ A website which is not listed

○ Profile's website

○ Other:

✱ How safe do you think it would be to click on the link above if you saw it in an email from someone you know?

○ Not safe

○ Somewhat unsafe

○ Neutral

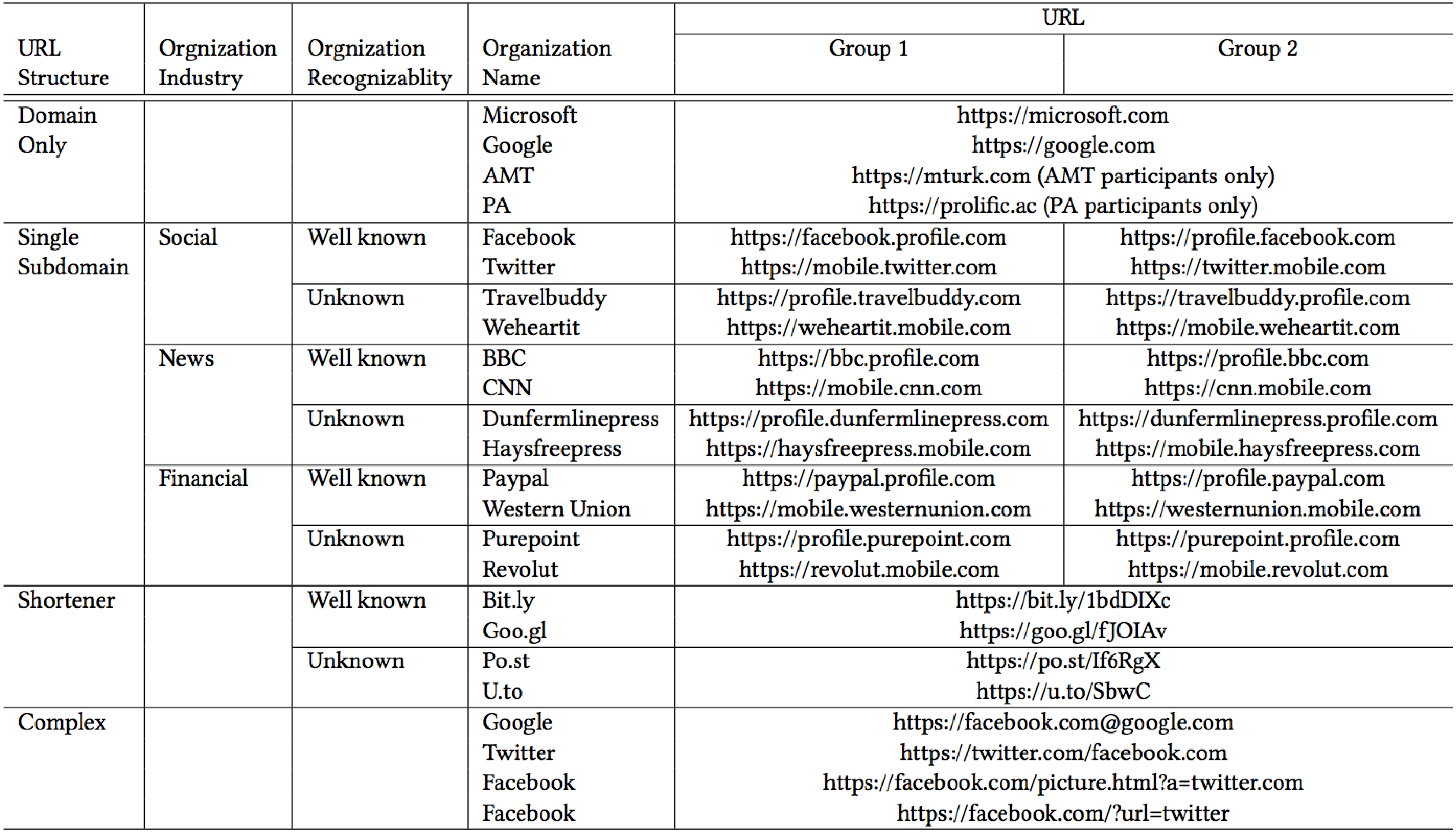
○ Somewhat safe

○ Very safe

S. S. Albakry, K. Vaniea, M. K. Wolters. What is this URL's Destination? Empirical Evaluation of Users' URL Reading. In CHI 2020.

Varied for subdomain

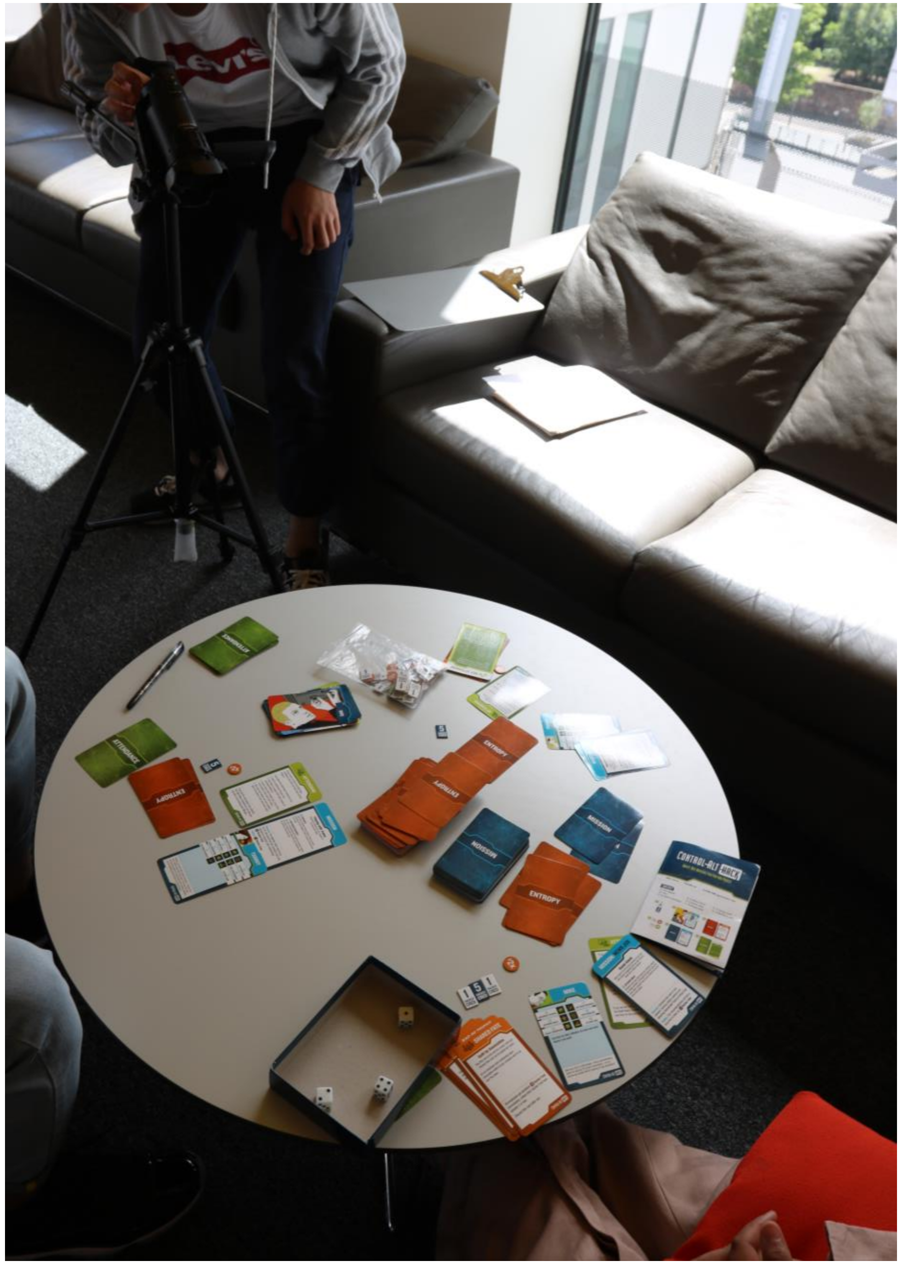
- **Sector** – social media, finance, news
- **How recognizable** – well known, relatively unknown
    - Pre-study to find known companies
    - Startups to find unknown real companies
- **Filler word** – mobile or profile



URL Structure	Orgnization Industry	Orgnization Recognizablity	Organization Name	URL	
				Group 1	Group 2
Domain Only			Microsoft Google AMT PA	https://microsoft.com https://google.com https://mturk.com (AMT participants only) https://prolific.ac (PA participants only)	
Single Subdomain	Social	Well known	Facebook Twitter	https://facebook.profile.com https://mobile.twitter.com	https://profile.facebook.com https://twitter.mobile.com
		Unknown	Travelbuddy Weheartit	https://profile.travelbuddy.com https://weheartit.mobile.com	https://travelbuddy.profile.com https://mobile.weheartit.com
	News	Well known	BBC CNN	https://bbc.profile.com https://mobile.cnn.com	https://profile.bbc.com https://cnn.mobile.com
		Unknown	Dunfermlinepress Haysfreepress	https://profile.dunfermlinepress.com https://haysfreepress.mobile.com	https://dunfermlinepress.profile.com https://mobile.haysfreepress.com
	Financial	Well known	Paypal Western Union	https://paypal.profile.com https://mobile.westernunion.com	https://profile.paypal.com https://westernunion.mobile.com
		Unknown	Purepoint Revolut	https://profile.purepoint.com https://revolut.mobile.com	https://purepoint.profile.com https://mobile.revolut.com
Shortener		Well known	Bit.ly Goo.gl	https://bit.ly/1bdDIXc https://goo.gl/fJOIAv	
		Unknown	Po.st U.to	https://po.st/If6RgX https://u.to/SbwC	
Complex			Google Twitter Facebook Facebook	https://facebook.com@google.com https://twitter.com/facebook.com https://facebook.com/picture.html?a=twitter.com https://facebook.com/?url=twitter	

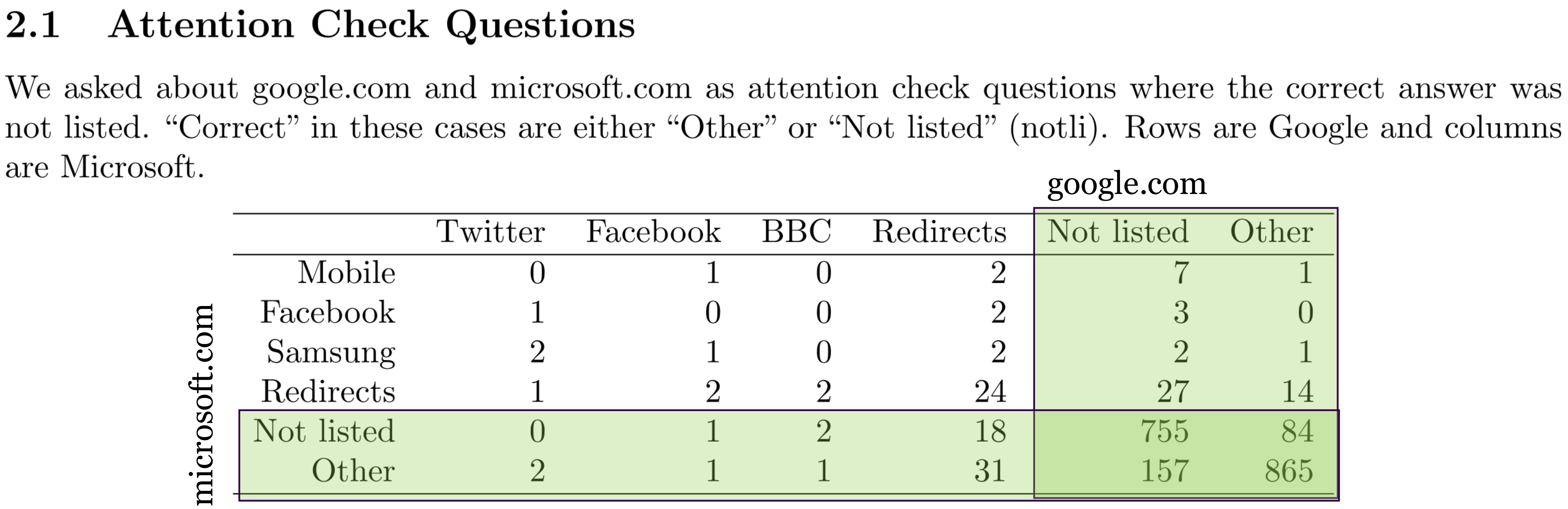
Structuring Research

- Research question or goal
- Literature review (what have others learned or done)
- Methods planned to answer question or achieve goal
- **Evaluate outcome**
- Contextualize findings
- Writeup



S. S. Albakry, K. Vaniea, M. K. Wolters. What is this URL's Destination? Empirical Evaluation of Users' URL Reading. In CHI 2020.

Raw data: attention check



2.1 Attention Check Questions

We asked about google.com and microsoft.com as attention check questions where the correct answer was not listed. “Correct” in these cases are either “Other” or “Not listed” (notli). Rows are Google and columns are Microsoft.

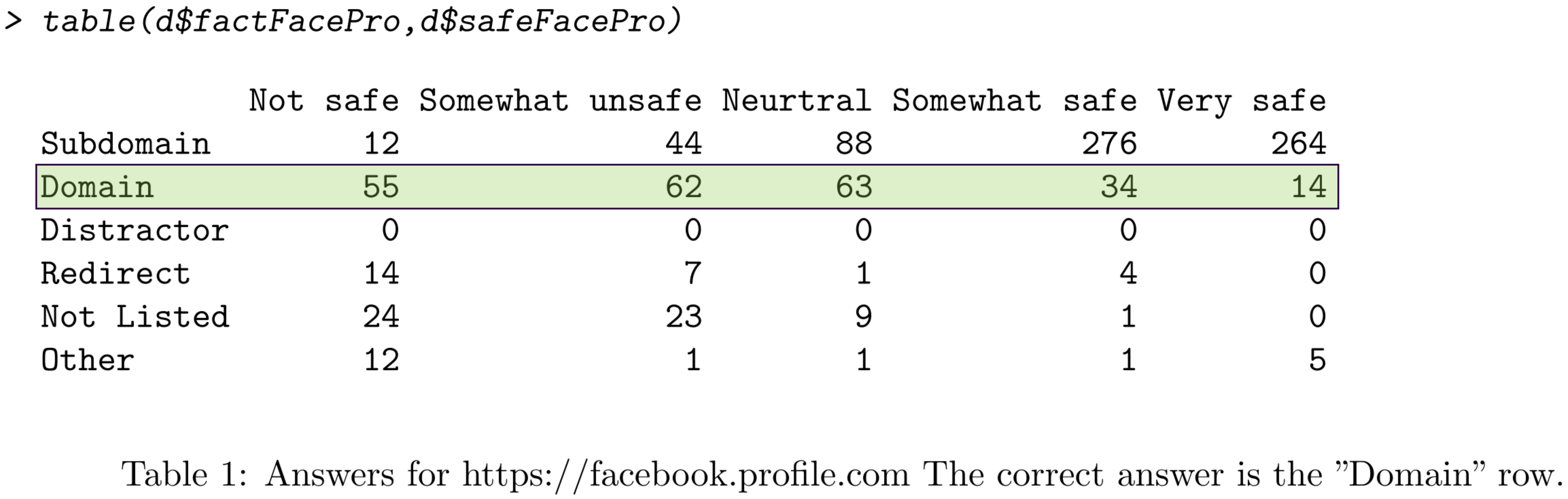
google.com

microsoft.com

	Twitter	Facebook	BBC	Redirects	Not listed	Other
Mobile	0	1	0	2	7	1
Facebook	1	0	0	2	3	0
Samsung	2	1	0	2	2	1
Redirects	1	2	2	24	27	14
Not listed	0	1	2	18	755	84
Other	2	1	1	31	157	865

S. S. Albakry, K. Vaniea, M. K. Wolters. What is this URL's Destination? Empirical Evaluation of Users' URL Reading. In CHI 2020.

Where will the link lead vs safety question facebook.profile.com



```
> table(d$factFacePro,d$safeFacePro)

           Not safe Somewhat unsafe Neurtral Somewhat safe Very safe
Subdomain        12              44       88           276       264
Domain           55              62       63            34        14
Distractor        0               0        0             0         0
Redirect         14               7        1             4         0
Not Listed       24              23        9             1         0
Other            12               1        1             1         5
```

Table 1: Answers for https://facebook.profile.com The correct answer is the ”Domain” row.

S. S. Albakry, K. Vaniea, M. K. Wolters. What is this URL's Destination? Empirical Evaluation of Users' URL Reading. In CHI 2020.

Research Question

Informal RQ:

- **Can people read URLs under optimal conditions?**

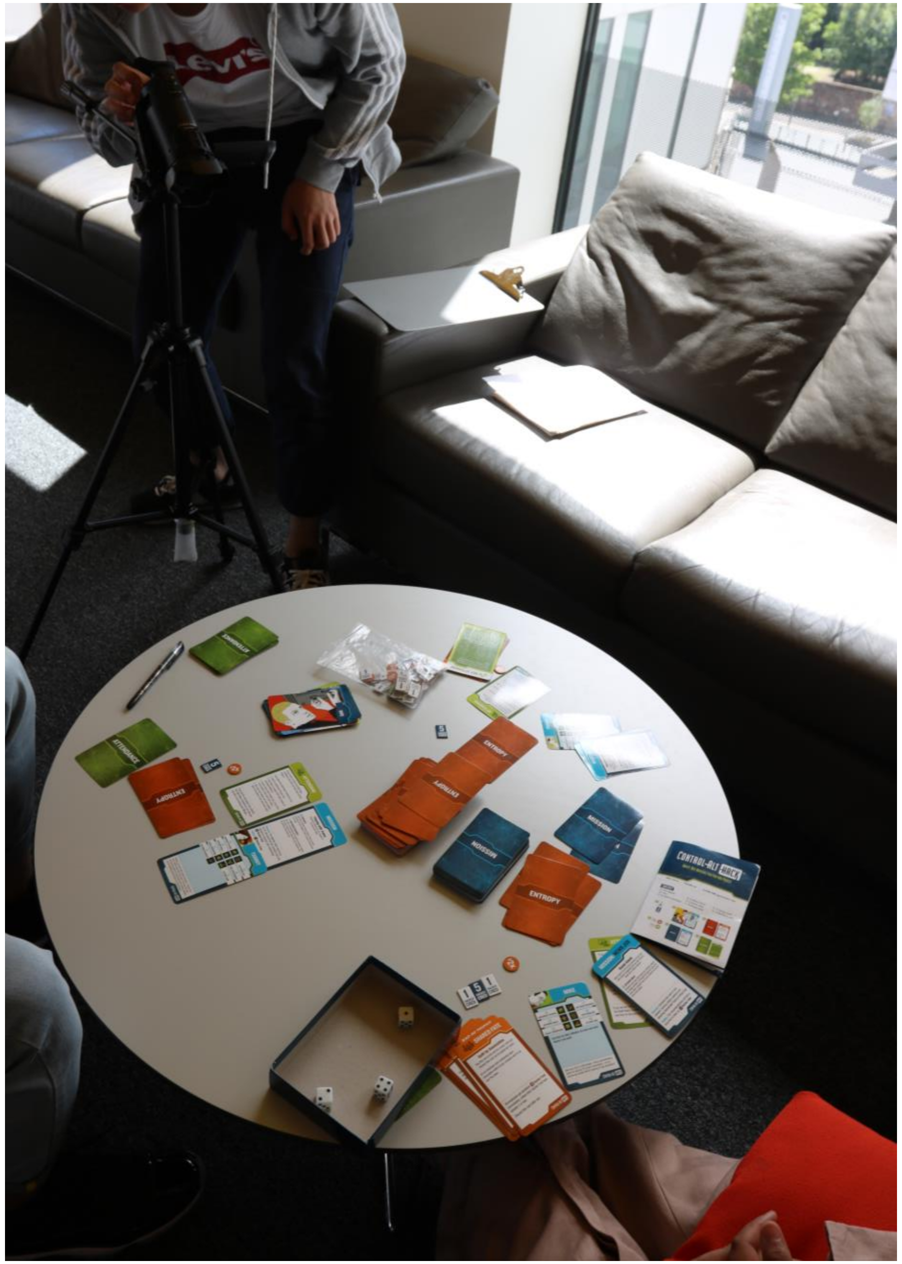
Formal RQs:

- RQ1 Can users accurately predict where a URL will go?
  - RQ1.1 Can users correctly infer from the URL that it will go to the website of the organization listed in the domain position rather than the subdomain, and what factors affect prediction accuracy?
  - RQ1.2 Can users recognize that the end destination of shortened URLs is not easy to predict?
  - RQ1.3 Can users recognize the end destination of complex URL structures?
- RQ2 What effects users’ assessment of the likely safety of a URL?

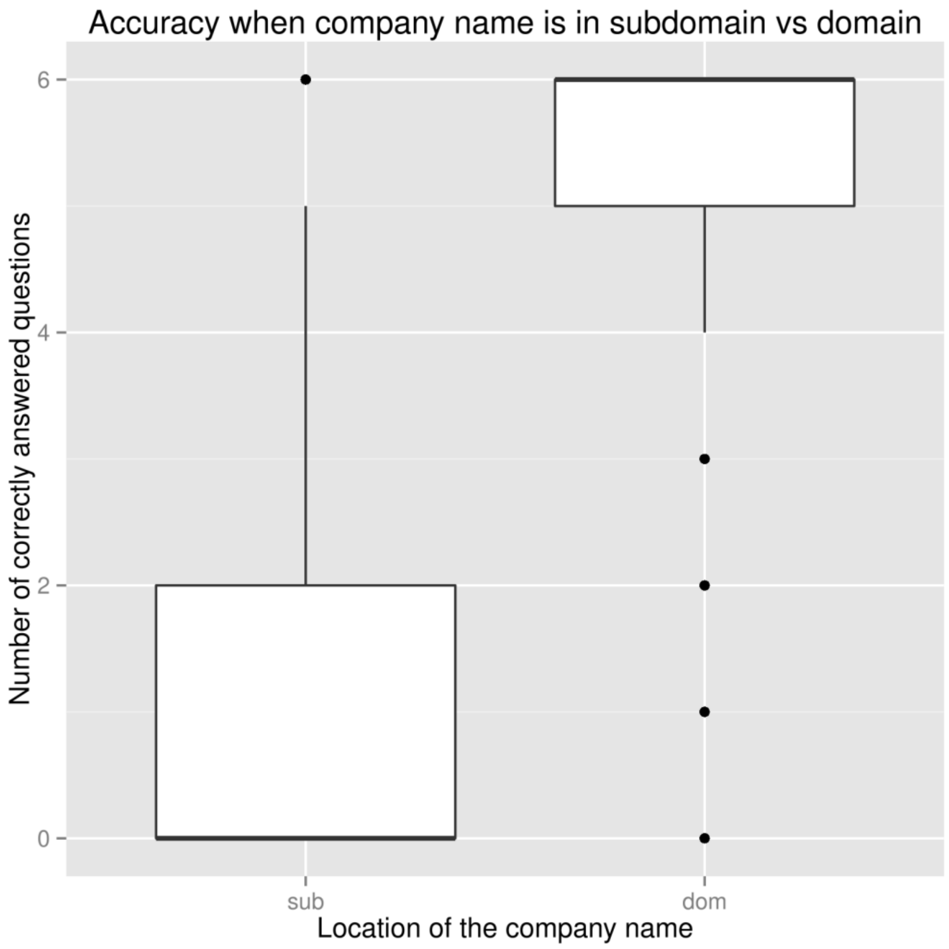
S. S. Albakry, K. Vaniea, M. K. Wolters. What is this URL's Destination? Empirical Evaluation of Users' URL Reading. In CHI 2020.

Structuring Research

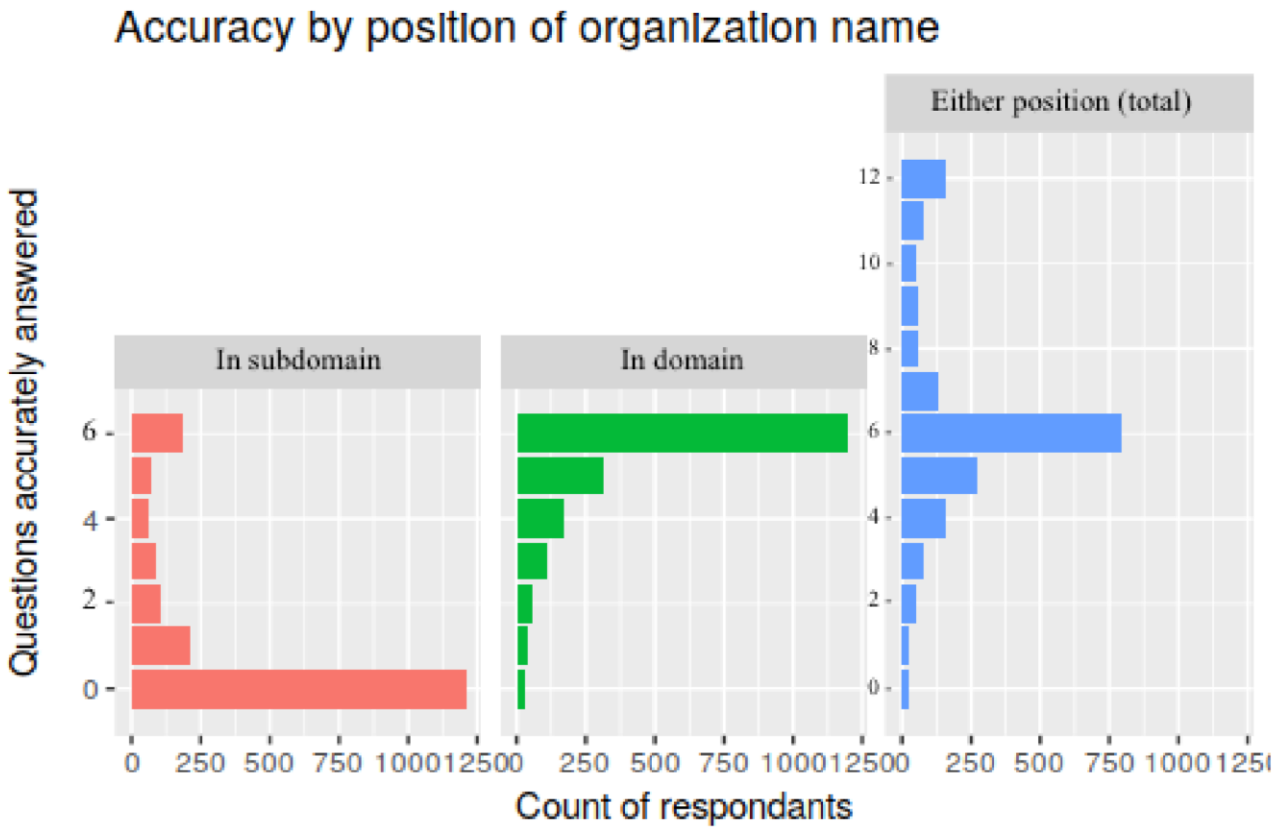
- Research question or goal
- Literature review (what have others learned or done)
- Methods planned to answer question or achieve goal
- Evaluate outcome
- **Contextualize findings**
- Writeup



How to present results to contextualize them

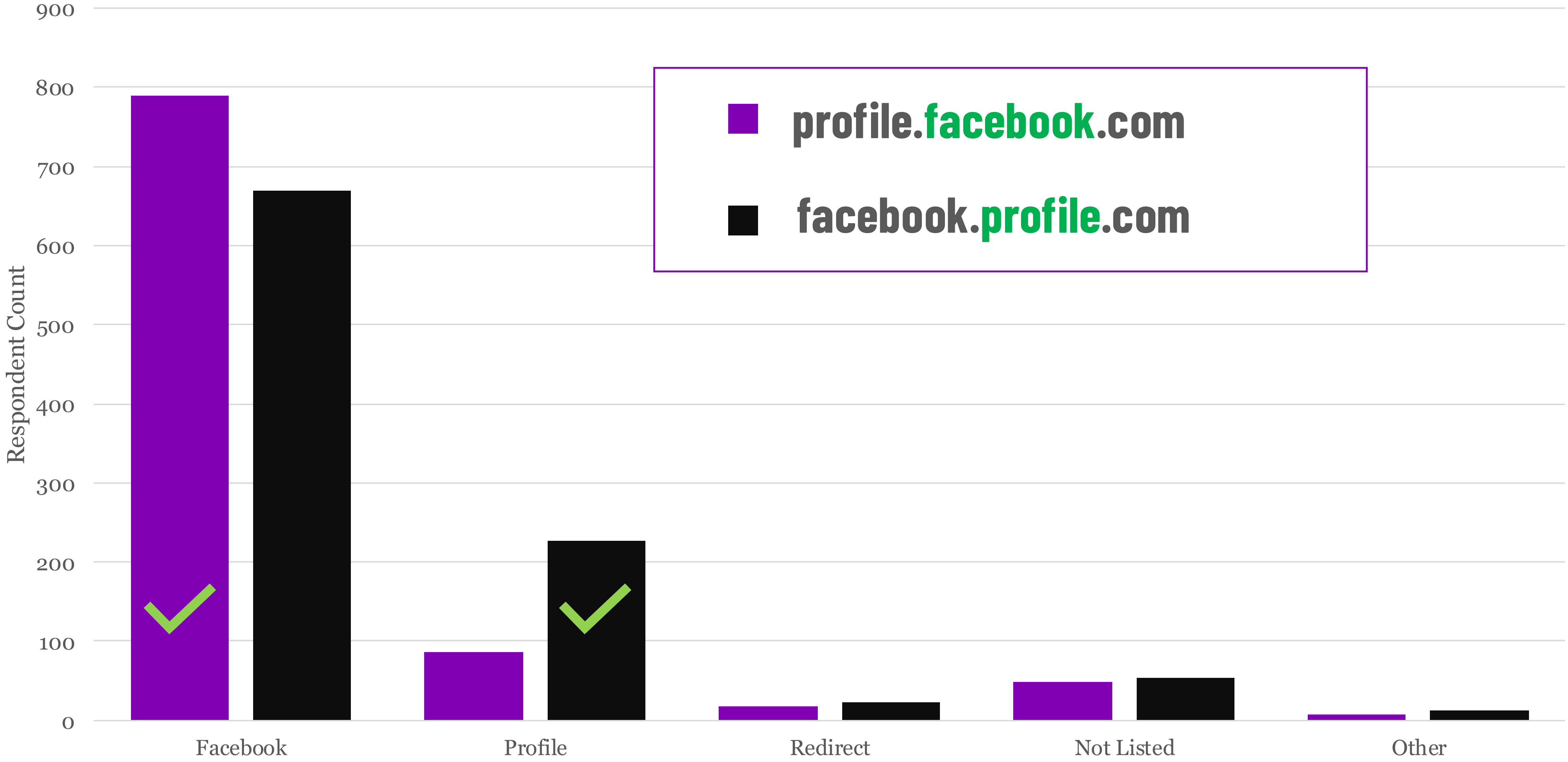


Mid-data analysis



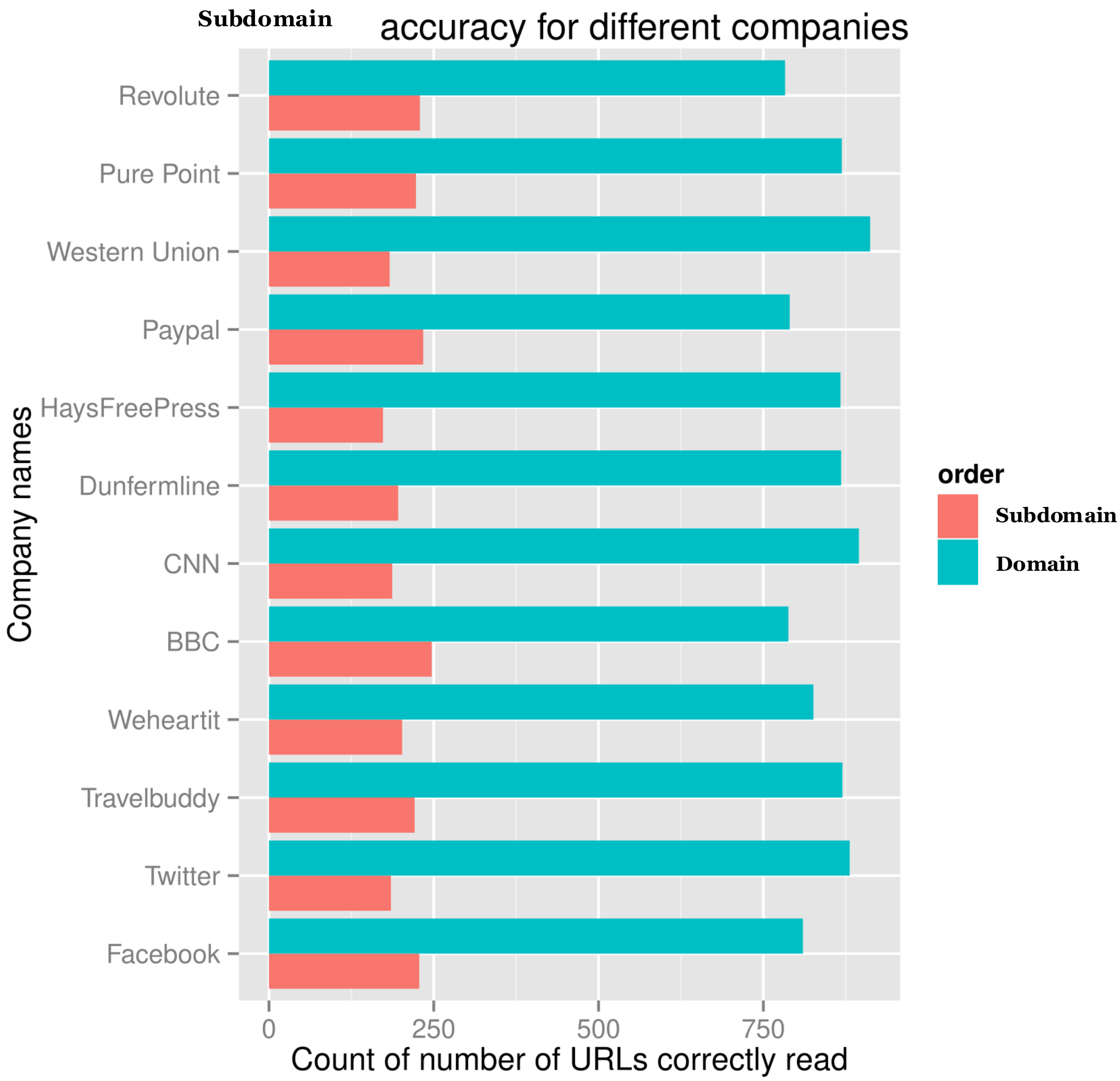
Final paper

If you were type the [below] link into a web browser, what website would open?

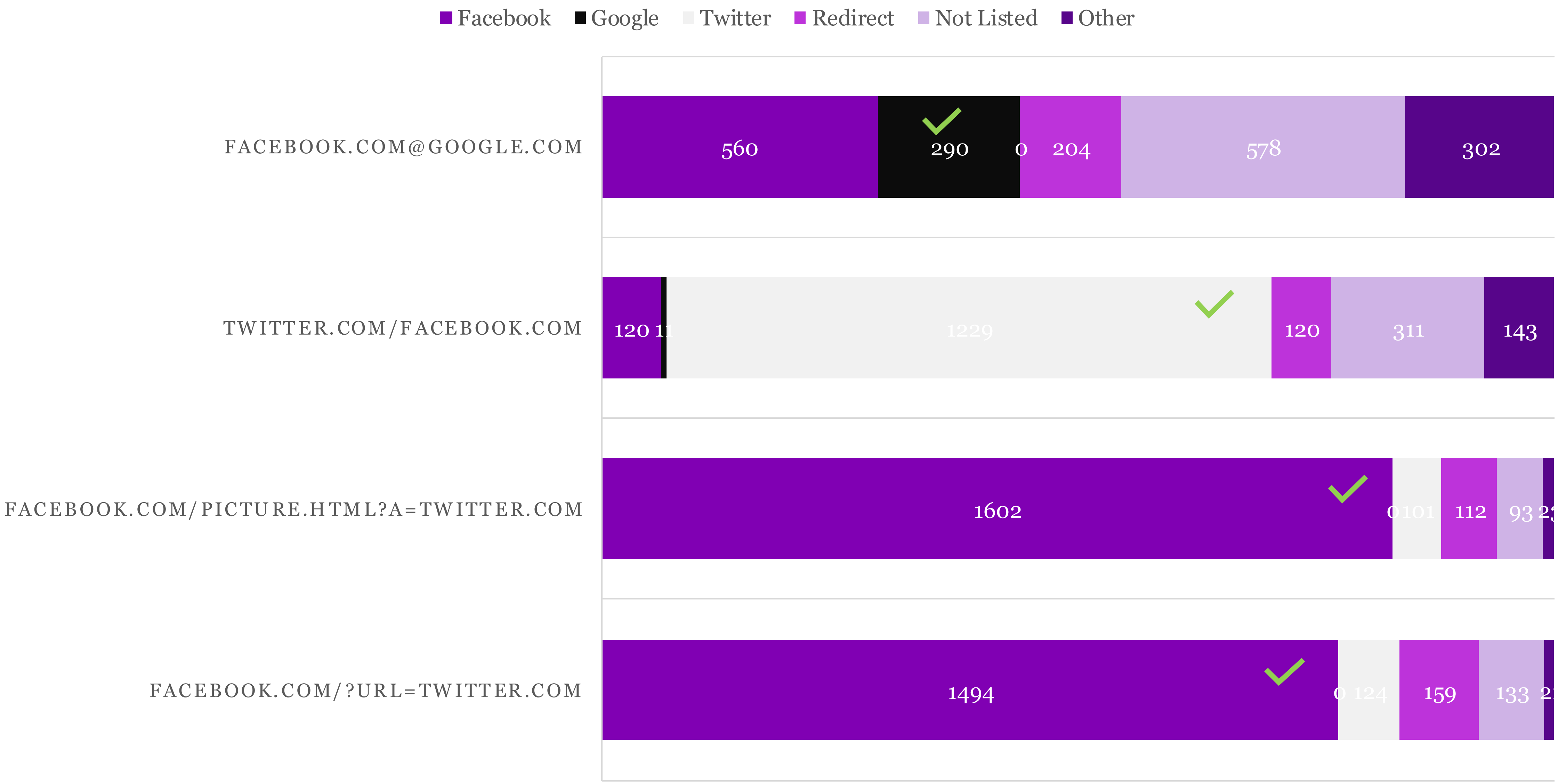


Interpreting findings

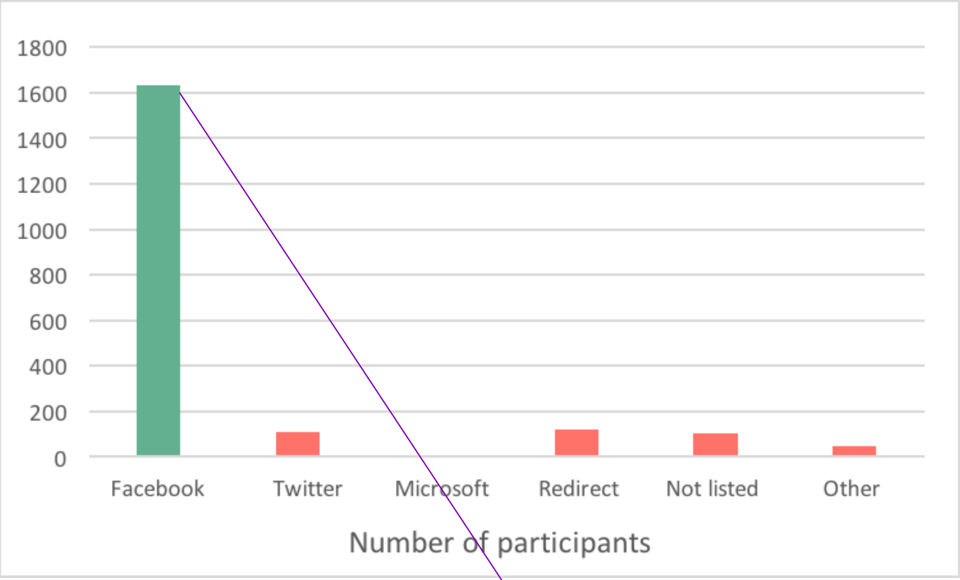
- Unsurprisingly people were worse when company name was in the subdomain position.
- But there is variation… does it mean anything?



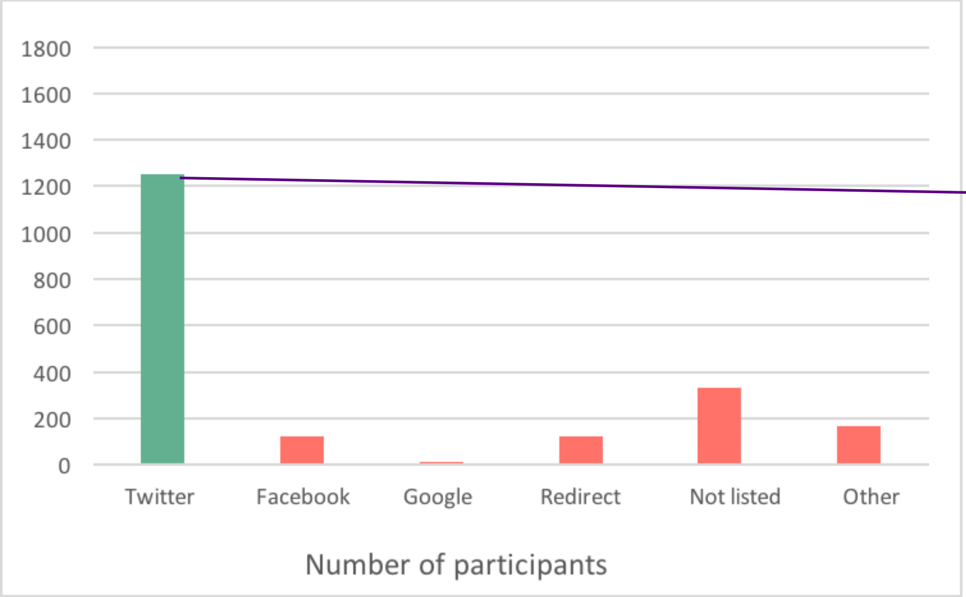
Complex URLs



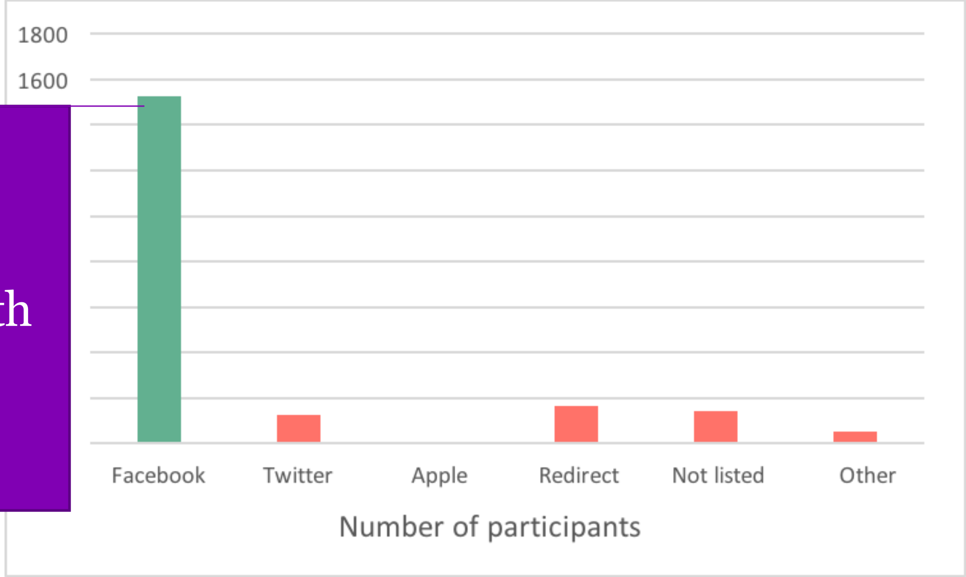
https://facebook.com/picture.html?a=twitter.com



https://twitter.com/facebook.com



https://facebook.com/?url=twitter



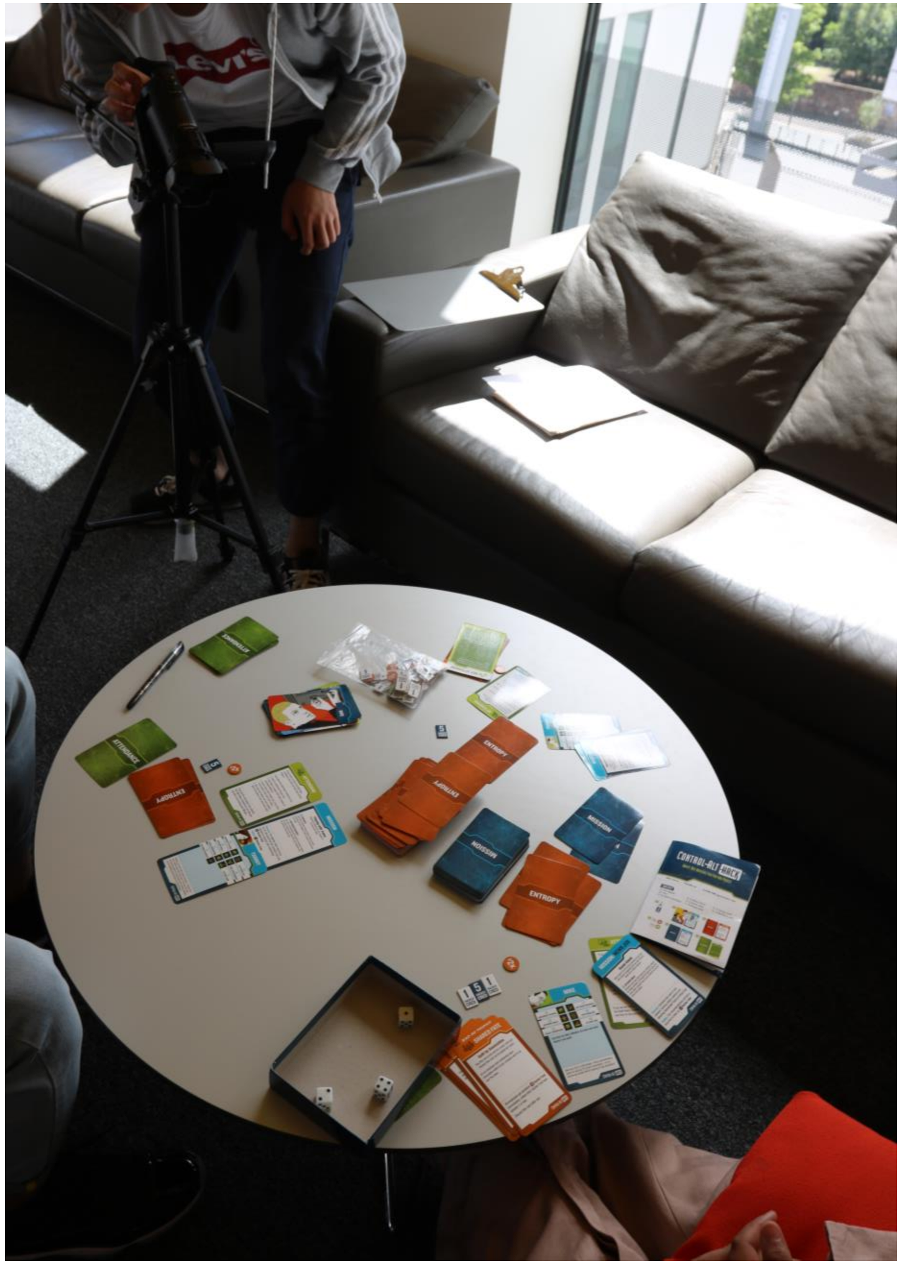
Knew how to correctly read path URLs

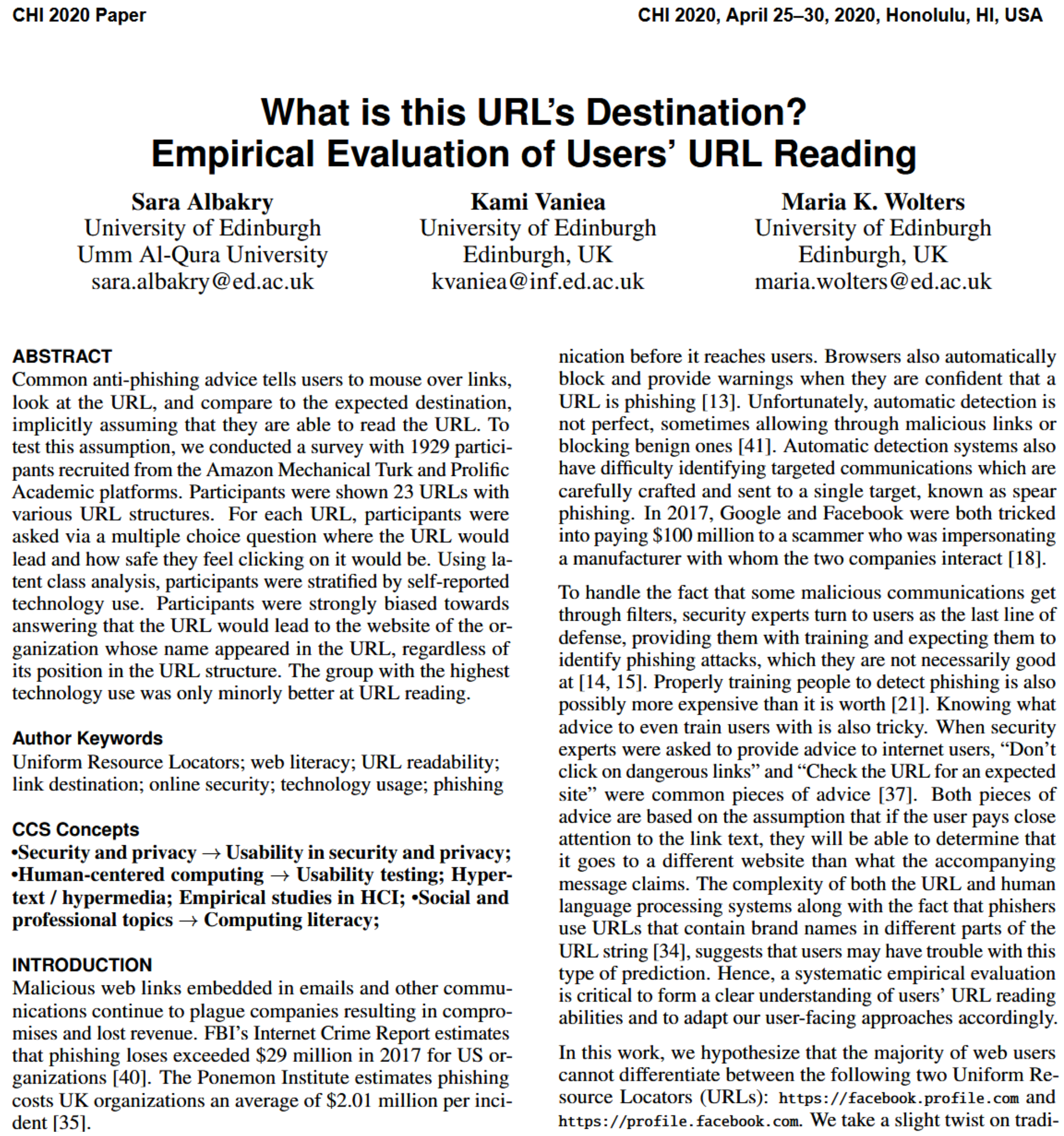
Can people read URLs under optimal conditions?

People can read basic and path URLs but struggle with subdomain URLs.

Structuring Research

- Research question or goal
- Literature review (what have others learned or done)
- Methods planned to answer question or achieve goal
- Evaluate outcome
- Contextualize findings
- **Writeup**





What is this URL's Destination? Empirical Evaluation of Users' URL Reading

Sara Albakry
University of Edinburgh
Umm Al-Qura University
sara.albakry@ed.ac.uk

Kami Vaniea
University of Edinburgh
Edinburgh, UK
kvaniea@inf.ed.ac.uk

Maria K. Wolters
University of Edinburgh
Edinburgh, UK
maria.wolters@ed.ac.uk

ABSTRACT

Common anti-phishing advice tells users to mouse over links, look at the URL, and compare to the expected destination, implicitly assuming that they are able to read the URL. To test this assumption, we conducted a survey with 1929 participants recruited from the Amazon Mechanical Turk and Prolific Academic platforms. Participants were shown 23 URLs with various URL structures. For each URL, participants were asked via a multiple choice question where the URL would lead and how safe they feel clicking on it would be. Using latent class analysis, participants were stratified by self-reported technology use. Participants were strongly biased towards answering that the URL would lead to the website of the organization whose name appeared in the URL, regardless of its position in the URL structure. The group with the highest technology use was only minorly better at URL reading.

Author Keywords

Uniform Resource Locators; web literacy; URL readability; link destination; online security; technology usage; phishing

CCS Concepts

•Security and privacy → Usability in security and privacy; •Human-centered computing → Usability testing; Hypertext / hypermedia; Empirical studies in HCI; •Social and professional topics → Computing literacy;

INTRODUCTION

Malicious web links embedded in emails and other communications continue to plague companies resulting in compromises and lost revenue. FBI's Internet Crime Report estimates that phishing loses exceeded $29 million in 2017 for US organizations [40]. The Ponemon Institute estimates phishing costs UK organizations an average of $2.01 million per incident [35].

nication before it reaches users. Browsers also automatically block and provide warnings when they are confident that a URL is phishing [13]. Unfortunately, automatic detection is not perfect, sometimes allowing through malicious links or blocking benign ones [41]. Automatic detection systems also have difficulty identifying targeted communications which are carefully crafted and sent to a single target, known as spear phishing. In 2017, Google and Facebook were both tricked into paying $100 million to a scammer who was impersonating a manufacturer with whom the two companies interact [18].

To handle the fact that some malicious communications get through filters, security experts turn to users as the last line of defense, providing them with training and expecting them to identify phishing attacks, which they are not necessarily good at [14, 15]. Properly training people to detect phishing is also possibly more expensive than it is worth [21]. Knowing what advice to even train users with is also tricky. When security experts were asked to provide advice to internet users, "Don't click on dangerous links" and "Check the URL for an expected site" were common pieces of advice [37]. Both pieces of advice are based on the assumption that if the user pays close attention to the link text, they will be able to determine that it goes to a different website than what the accompanying message claims. The complexity of both the URL and human language processing systems along with the fact that phishers use URLs that contain brand names in different parts of the URL string [34], suggests that users may have trouble with this type of prediction. Hence, a systematic empirical evaluation is critical to form a clear understanding of users' URL reading abilities and to adapt our user-facing approaches accordingly.

In this work, we hypothesize that the majority of web users cannot differentiate between the following two Uniform Resource Locators (URLs): `https://facebook.profile.com` and `https://profile.facebook.com`. We take a slight twist on tradi-

QUESTIONS