Research Framework

INFR11158/11230 Usable Security and Privacy

Dr. Jingjie Li

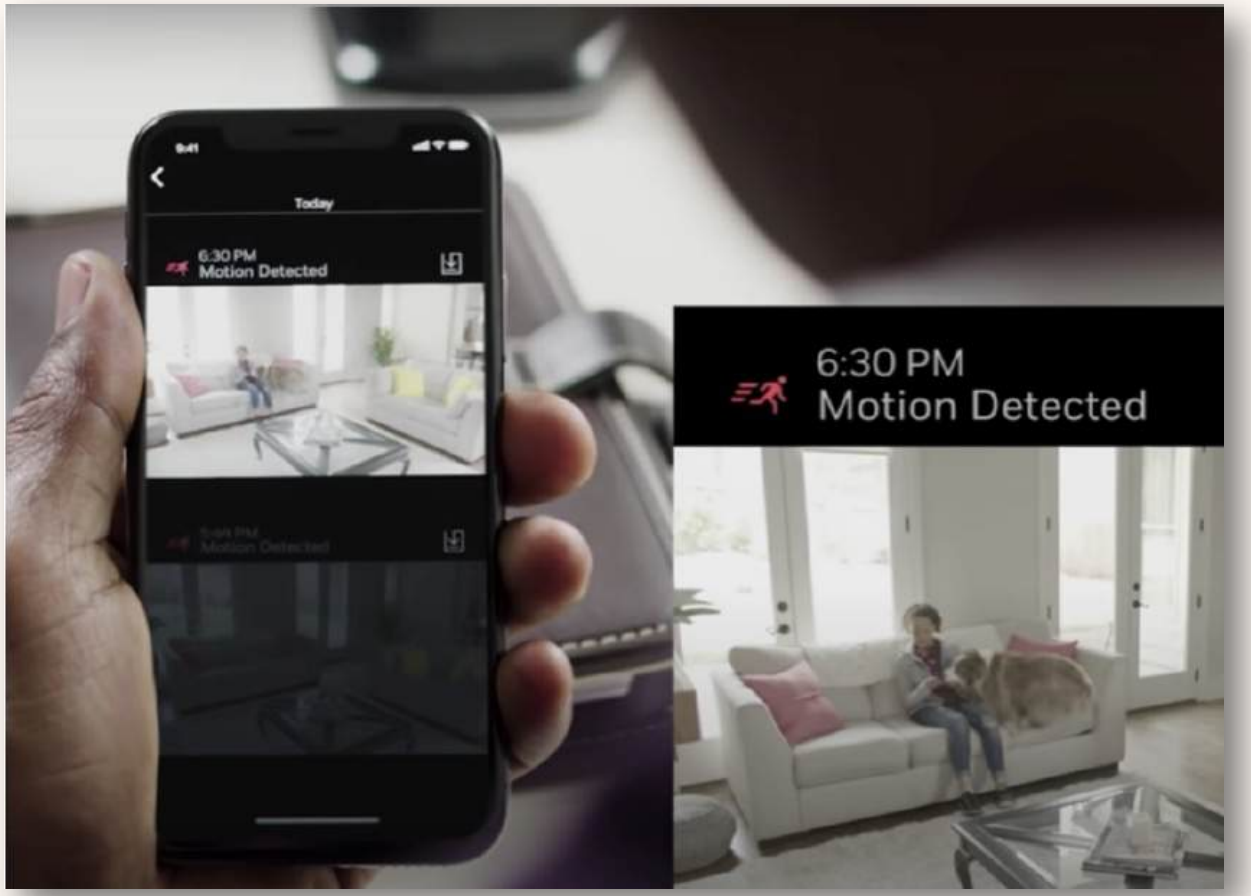
14/02/2024

THE UNIVERSITY *of* EDINBURGH

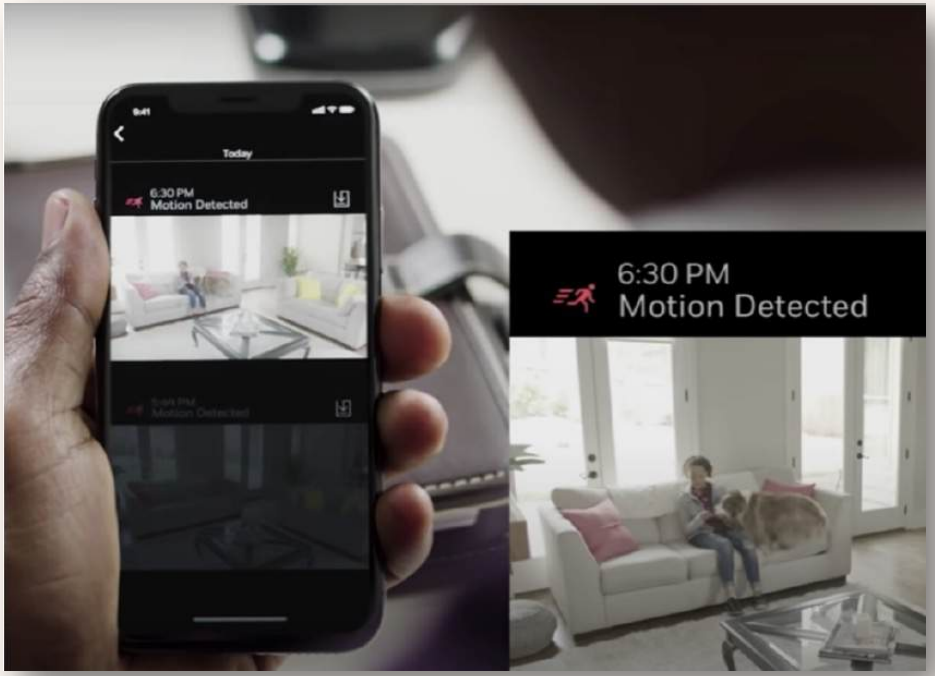
Overview

- Recap
- Human in the loop
- Planning studies – putting together
- Take-home

Give a try!



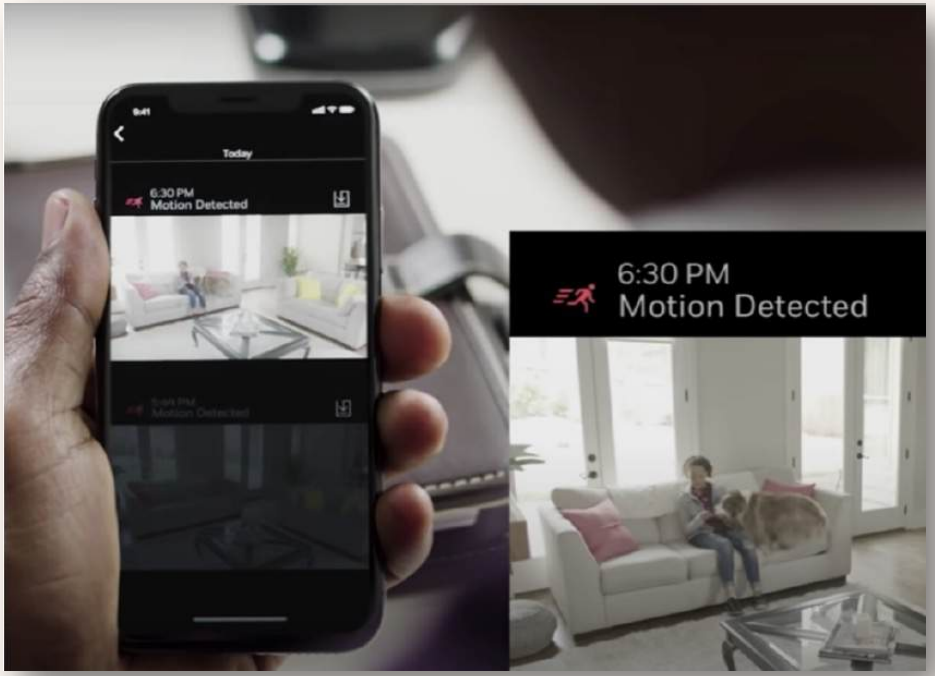
Give a try!



- Snapshot of an app on a phone
- Child playing with dog
- Edited picture
- Motion detection enabled

....

Give a try!



What are the privacy concerns?
- Children being monitored
- Camera placed in the living room

....

Codebook

Analysis of child safety and privacy-related features and information		
Code/subcode		**Explanation**
If safety is mentioned (in text)	Webpage providing additional safety resources	The webpage provides extra links and resources for safety,
	Webpage not providing additional safety resources	The webpage provides no extra links or resources for safety, and it just mentions words about safety.
If privacy is mentioned (in text)	Webpage providing additional privacy resources	The webpage provides extra links and resources for privacy, and it just mentions words about privacy.
	Webpage not providing additional privacy resources	The webpage provides no extra links or resources for privacy, and it just mentions words about privacy.
Presence of product features addressing privacy or safety		The webpage shows features, such as child safety lock or privacy shutter buttons.
Presence of parental control-related information		The webpage shows how they offer parental control for the products, e.g., child-specific access control.
Presence of information for parents		The webpage shows information on what parents could do or not do with the children and the device, e.g., parents should supervise children when using the device.
Presence of child-specific content, features, functions, and profile		The webpage shows products offering child mode, content, and child profiles.
Presence of information/disclaimer of potential risks		The webpage shows potential safety risks or misuse, or explains what parents need to be aware of regarding the potential privacy or safety risks.
Presence of information on child data collection and use		The webpage shows how the company handles child-related data. Examples include data collection and use, sharing, and selling.
Indication of child existence (when child image is not presented)		The webpage shows no children, but the image shows child-related signals, including children's bedroom or toys/stuff around the smart home device or in the room.
Presence of smart home guide, e.g., teaching children about device use		The webpage shows instructions or materials directly addressing children, teaching children about device use.

- Read the data, iteratively refine the codes (in discussion with others), organize codes in a hierarchy...

Inductive coding vs deductive coding

- **Inductive:** look for any ideas that interest you from different aspects
    - Snapshot of an app on a phone
    - Child playing with dog
    - Edited picture
    - Motion detection enabled
    ....

- **Deductive:** start with some hypothesis
    - Children being monitored by app (privacy concern)
    - Camera placed in the living room (place of the scene)

We often use inductive and deductive coding together!

Reliability

- Stability (intra-rater reliability): whether the same coder codes the data in a consistent way throughout the process
- Reproducibility (inter-coder reliability): whether different coders code the same piece of data consistently (metric: Cohen's Kappa)

Do we still need a human coder?

CollabCoder: A Lower-barrier, Rigorous Workflow for Inductive Collaborative Qualitative Analysis with Large Language Models

Jie Gao
Singapore University of Technology and Design, Singapore
Singapore-MIT Alliance for Research and Technology, Singapore
gaojie056@gmail.com

Yuchen Guo
Singapore University of Technology and Design, Singapore
yuchen_guo@mymail.sutd.edu.sg

Gionnieve Lim
Singapore University of Technology and Design, Singapore
gionnieve_lim@mymail.sutd.edu.sg

Tianqin Zhang
Singapore University of Technology and Design, Singapore
tianqin_zhang@mymail.sutd.edu.sg

Zheng Zhang
University of Notre Dame, USA
zzhang37@nd.edu

Toby Jia-Jun Li
University of Notre Dame, USA
toby.j.li@nd.edu

Simon Tangi Perrault
Singapore University of Technology and Design, Singapore
simon_perrault@sutd.edu.sg

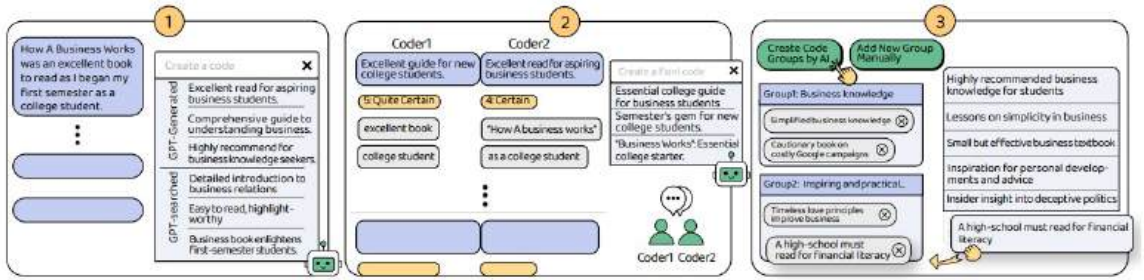
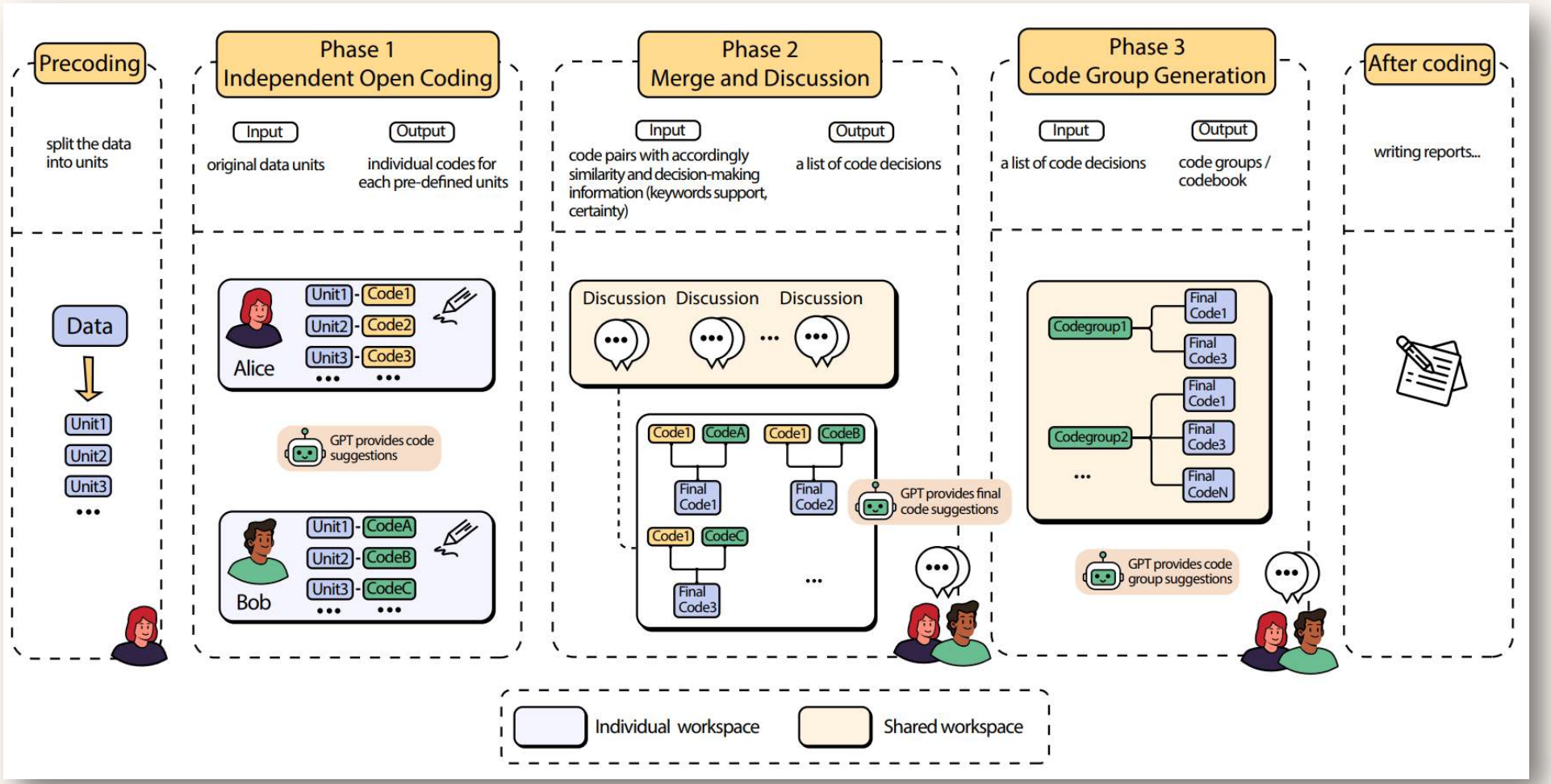
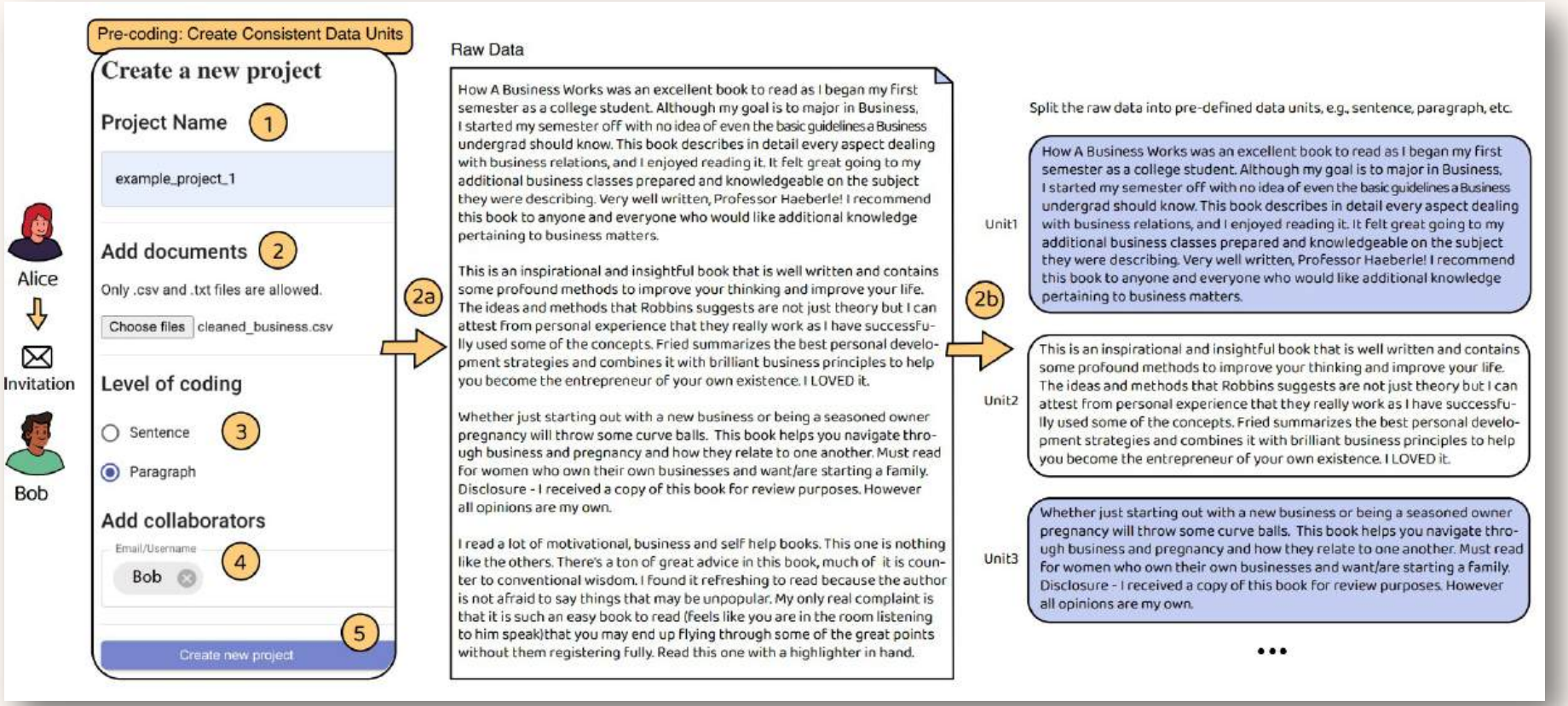
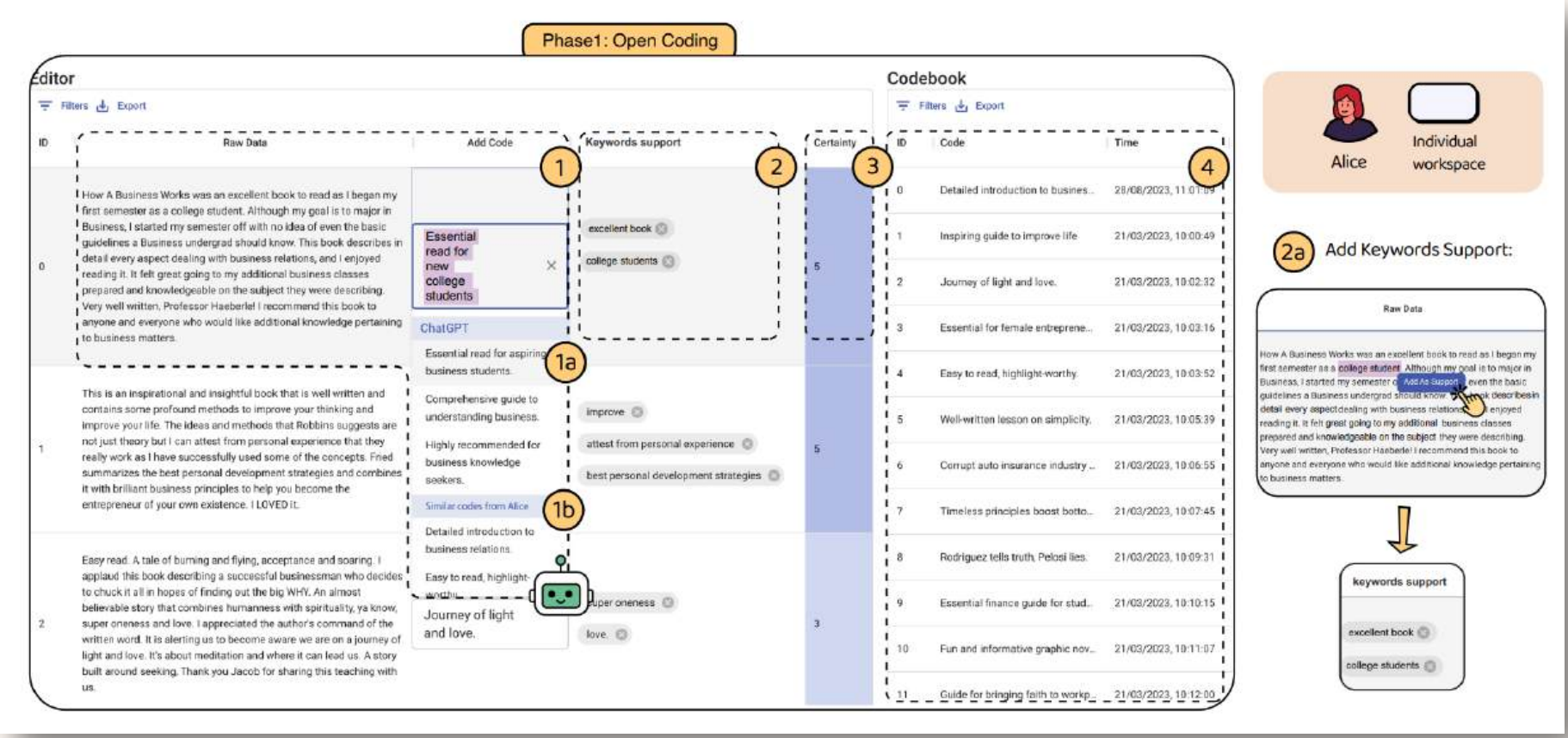


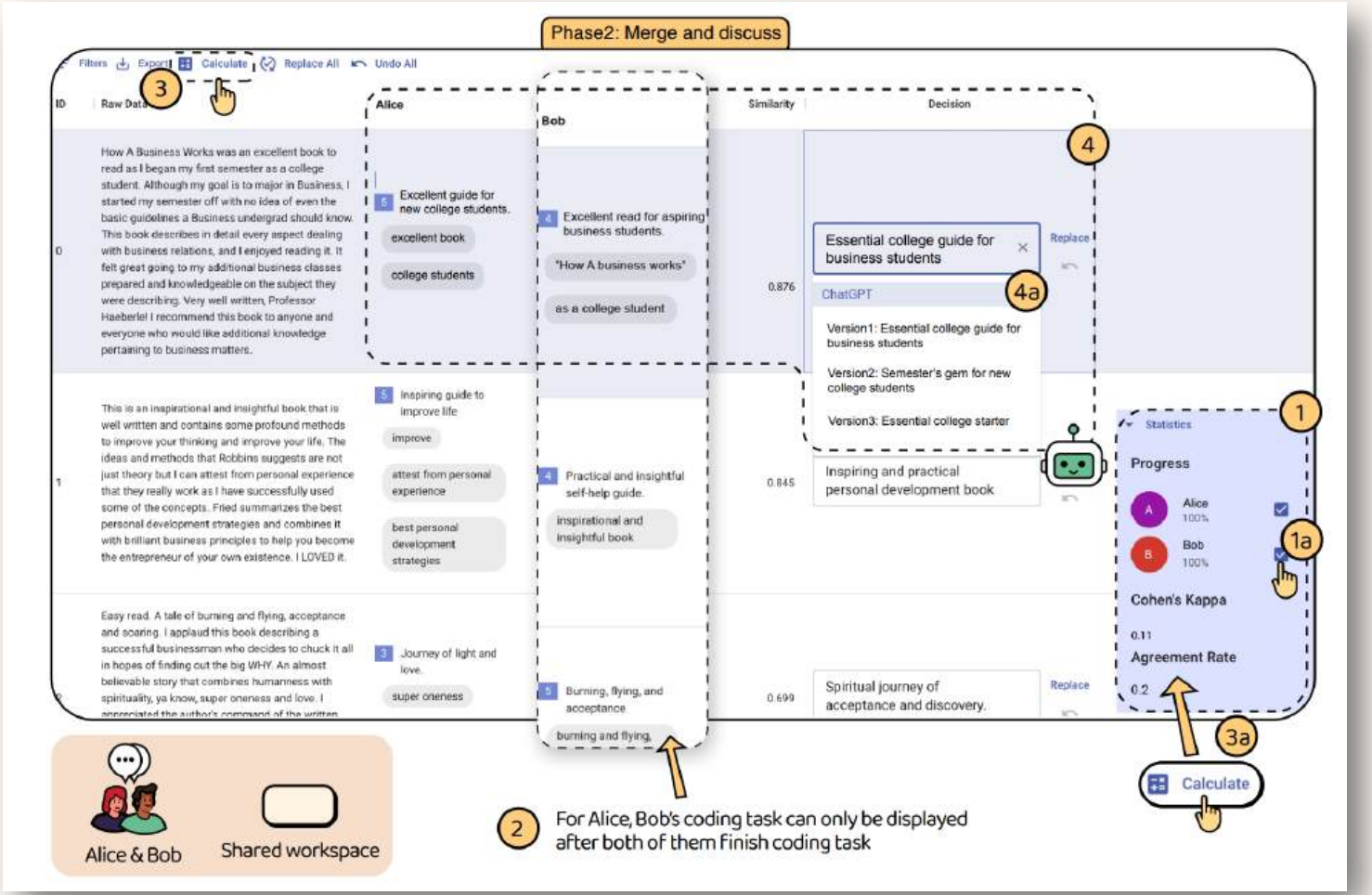
Figure 1: CollabCoder, An Integrated Workflow for Collaborative Qualitative Analysis. The workflow consists of three key stages: 1) Independent Open Coding, facilitated by on-demand code suggestions from LLMs, yielding initial codes; 2) Iterative Discussion, focusing on conflict mediation within the coding team, producing a list of agreed-upon code decisions; 3) Codebook Development, where code groups may be formed through LLM-generated suggestions, based on the list of decided codes.


Precoding
split the data into units
Data
Unit1
Unit2
Unit3
Phase 1
Independent Open Coding
Input
original data units
Output
individual codes for each pre-defined units
Alice
Unit1 - Code1
Unit2 - Code2
Unit3 - Code3
GPT provides code suggestions
Bob
Unit1 - CodeA
Unit2 - CodeB
Unit3 - CodeC
Phase 2
Merge and Discussion
Input
code pairs with accordingly similarity and decision-making information (keywords support, certainty)
Output
a list of code decisions
Discussion Discussion Discussion
Code1 CodeA Code1 CodeB
Final Code1
Final Code2
Code1 CodeC
Final Code3
GPT provides final code suggestions
Phase 3
Code Group Generation
Input
a list of code decisions
Output
code groups / codebook
Codegroup1
Final Code1
Final Code3
Codegroup2
Final Code1
Final Code3
Final CodeN
GPT provides code group suggestions
After coding
writing reports...
Individual workspace
Shared workspace

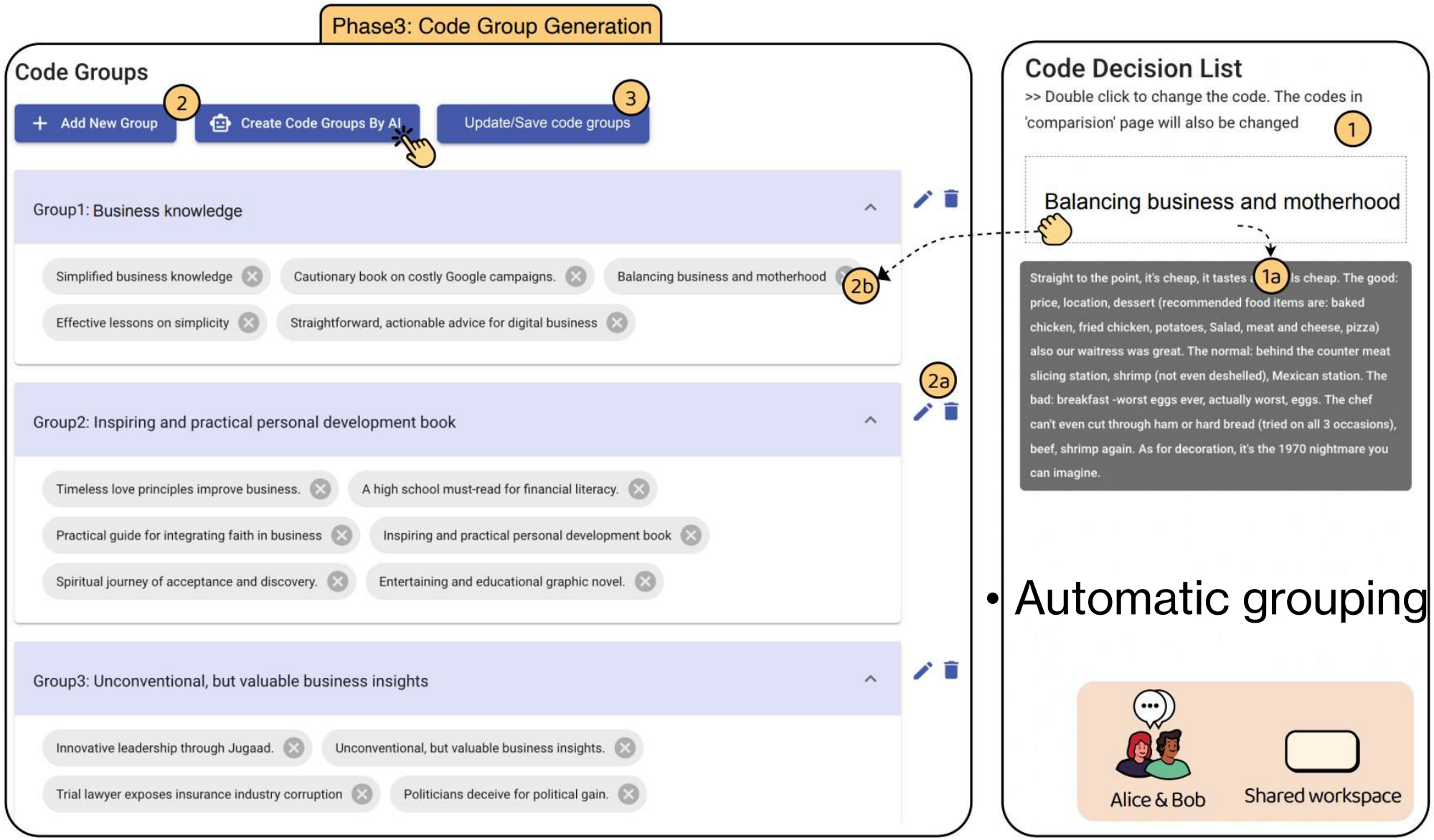

Pre-coding: Create Consistent Data Units
Create a new project
Project Name
1
example_project_1
Add documents
2
Only .csv and .txt files are allowed.
Choose files
cleaned_business.csv
Level of coding
Sentence
3
Paragraph
Add collaborators
Email/Username
Bob
4
5
Create new project
Alice
Invitation
Bob
2a
Raw Data
How A Business Works was an excellent book to read as I began my first semester as a college student. Although my goal is to major in Business, I started my semester off with no idea of even the basic guidelines a Business undergrad should know. This book describes in detail every aspect dealing with business relations, and I enjoyed reading it. It felt great going to my additional business classes prepared and knowledgeable on the subject they were describing. Very well written, Professor Haeberle! I recommend this book to anyone and everyone who would like additional knowledge pertaining to business matters.
This is an inspirational and insightful book that is well written and contains some profound methods to improve your thinking and improve your life. The ideas and methods that Robbins suggests are not just theory but I can attest from personal experience that they really work as I have successfully used some of the concepts. Fried summarizes the best personal development strategies and combines it with brilliant business principles to help you become the entrepreneur of your own existence. I LOVED it.
Whether just starting out with a new business or being a seasoned owner pregnancy will throw some curve balls. This book helps you navigate through business and pregnancy and how they relate to one another. Must read for women who own their own businesses and want/are starting a family. Disclosure - I received a copy of this book for review purposes. However all opinions are my own.
I read a lot of motivational, business and self help books. This one is nothing like the others. There's a ton of great advice in this book, much of it is counter to conventional wisdom. I found it refreshing to read because the author is not afraid to say things that may be unpopular. My only real complaint is that it is such an easy book to read (feels like you are in the room listening to him speak)that you may end up flying through some of the great points without them registering fully. Read this one with a highlighter in hand.
2b
Split the raw data into pre-defined data units, e.g., sentence, paragraph, etc.
Unit1
How A Business Works was an excellent book to read as I began my first semester as a college student. Although my goal is to major in Business, I started my semester off with no idea of even the basic guidelines a Business undergrad should know. This book describes in detail every aspect dealing with business relations, and I enjoyed reading it. It felt great going to my additional business classes prepared and knowledgeable on the subject they were describing. Very well written, Professor Haeberle! I recommend this book to anyone and everyone who would like additional knowledge pertaining to business matters.
Unit2
This is an inspirational and insightful book that is well written and contains some profound methods to improve your thinking and improve your life. The ideas and methods that Robbins suggests are not just theory but I can attest from personal experience that they really work as I have successfully used some of the concepts. Fried summarizes the best personal development strategies and combines it with brilliant business principles to help you become the entrepreneur of your own existence. I LOVED it.
Unit3
Whether just starting out with a new business or being a seasoned owner pregnancy will throw some curve balls. This book helps you navigate through business and pregnancy and how they relate to one another. Must read for women who own their own businesses and want/are starting a family. Disclosure - I received a copy of this book for review purposes. However all opinions are my own.
...



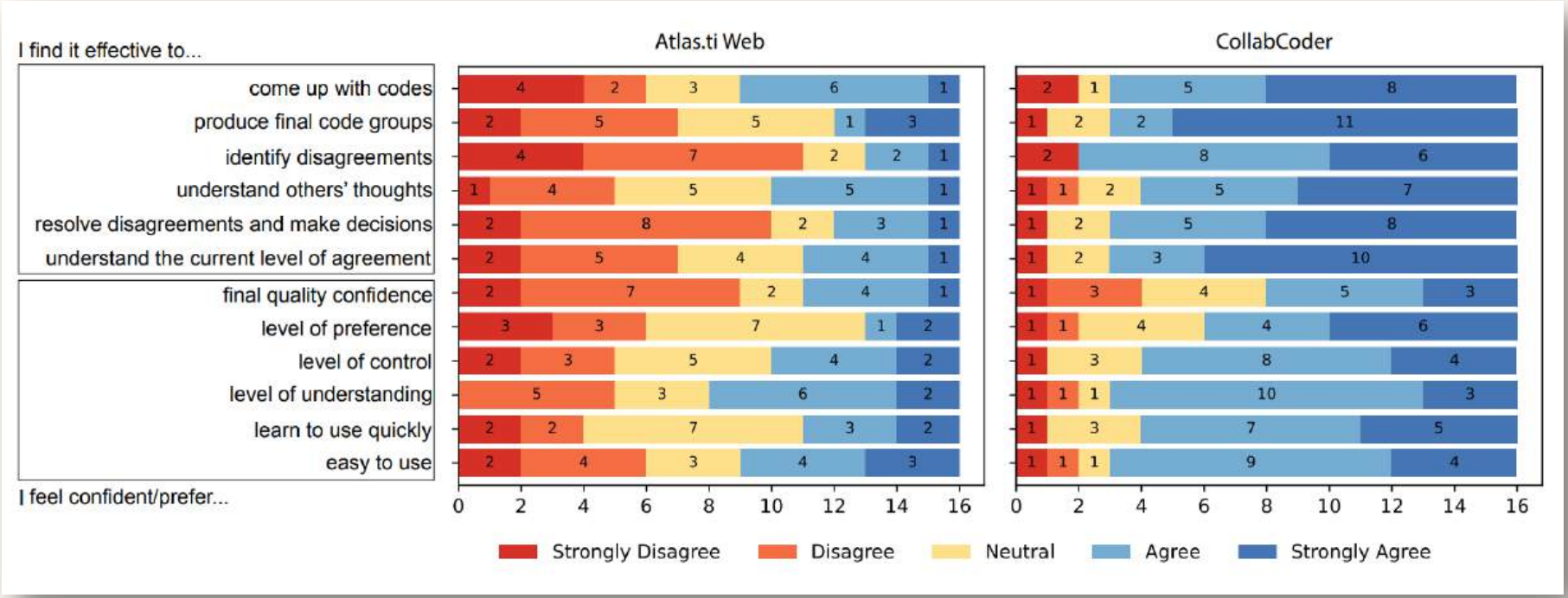
- Keywords & codes suggested by AI and crafted by human



- Compute similarity scores for merging codes

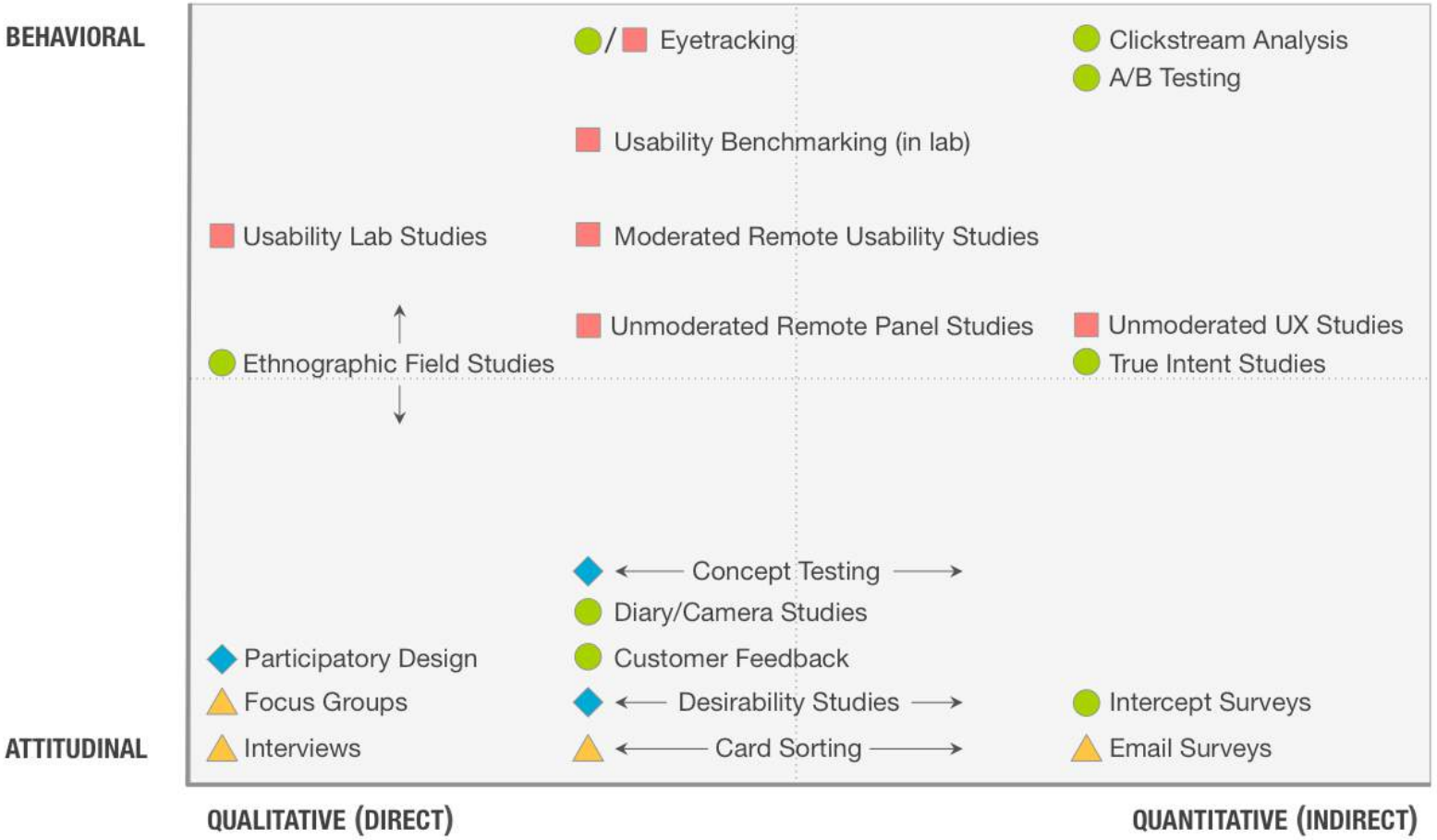


- Automatic grouping


I find it effective to...
come up with codes
produce final code groups
identify disagreements
understand others' thoughts
resolve disagreements and make decisions
understand the current level of agreement
I feel confident/prefer...
final quality confidence
level of preference
level of control
level of understanding
learn to use quickly
easy to use
Atlas.ti Web
CollabCoder
0 2 4 6 8 10 12 14 16
Strongly Disagree
Disagree
Neutral
Agree
Strongly Agree

Discuss: can AI replace human in qualitative analysis?

A Landscape of User Research Methods



KEY FOR CONTEXT OF PRODUCT USE DURING DATA COLLECTION

- Natural use of product
- Scripted (often lab-based) use of product
- De-contextualized / not using product
- Combination / hybrid

https://www.nngroup.com/articles/which-ux-research-methods/

What do you want to know or learn?

What hypothesis do you want to test?

What metrics do you want to apply?

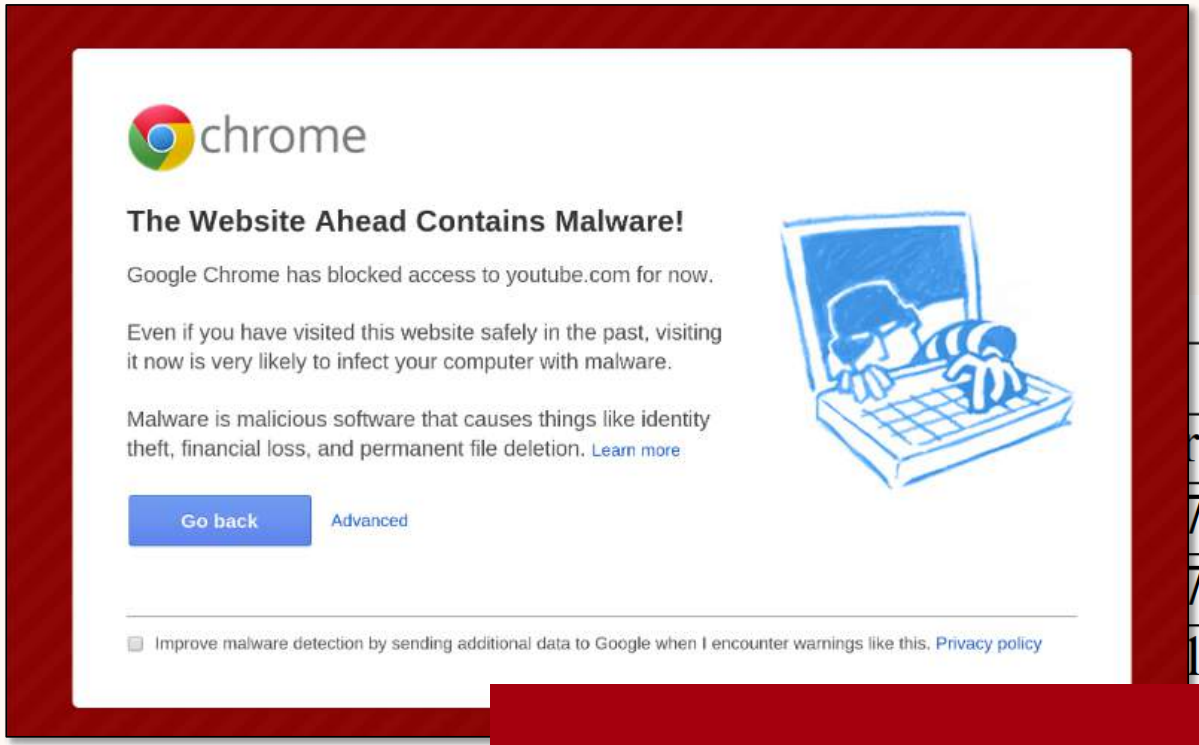
Is [my tool] usable?

Needs to be more specific to be testable.

Some of research questions:

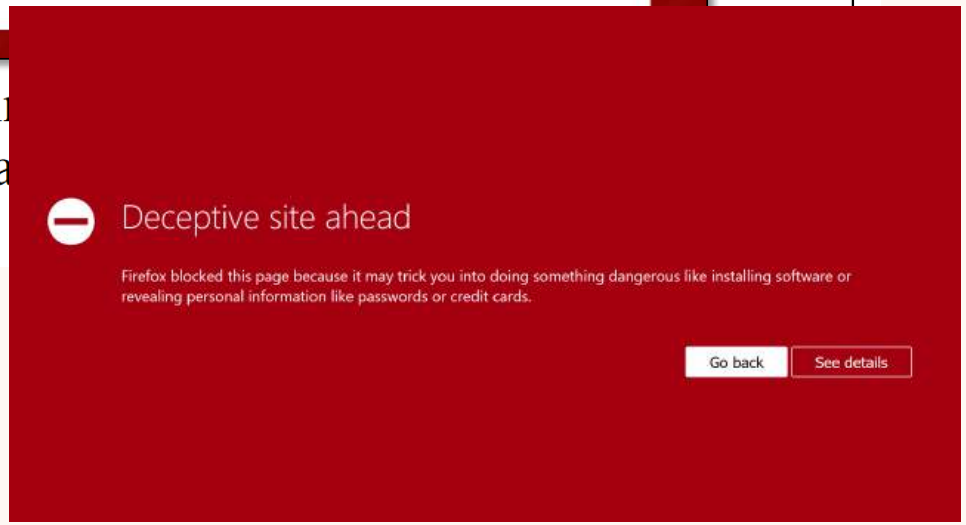
- Can people differentiate between a subdomain and a domain when reading a URL?
- Does [my new system] help people differentiate between malicious URLs and safe ones?
- Can users use [my new password manager] faster and with less errors than [the old password manager]?
- Does knowing how an app will use its permissions impact app installation decisions?
- What factors impact end-users' willingness to update software?
- Using [website], can users successfully opt-out of cookie tracking without forming inaccurate mental models?

Why do Chrome users ignore malware errors more often than Firefox users?



rome
7.9%
7.0%
1.0%

Table 1: User operati
ware and phishing wa
release) versions.



Why do people ignore malware warnings and what can we do about it?

Your Reputation Precedes You: History, Reputation, and the Chrome Malware Warning

Hazim Almuhimedi
Carnegie Mellon University
hazim@cs.cmu.edu

Adrienne Porter Felt
Robert W. Reeder
Sunny Consolvo
Google, Inc.
felt, rreeder, sconsolvo@google.com

ABSTRACT

Several web browsers, including Google Chrome and Mozilla Firefox, use malware warnings to stop people from visiting infectious websites. However, users can choose to click through (i.e., ignore) these malware warnings. In Google Chrome, users click through a fifth of malware warnings on average. We investigate factors that may contribute to why people ignore such warnings. First, we examine field data to see how browsing history affects click-through rates. We find that users consistently heed warnings about websites that they have not visited before. However, users respond unpredictably to warnings about websites that they have previously visited. On some days, users ignore more than half of warnings about websites they've visited in the past. Next, we present results of an online, survey-based experiment that we ran to gain more insight into the effects of reputation on warning adherence. Participants said that they trusted high-reputation websites more than the warnings; however, their responses suggest that a notable minority of people could be swayed by providing more information. We provide recommendations for warning designers and pose open questions about the design of malware warnings.

1. INTRODUCTION

Modern browsers such as Google Chrome and Mozilla

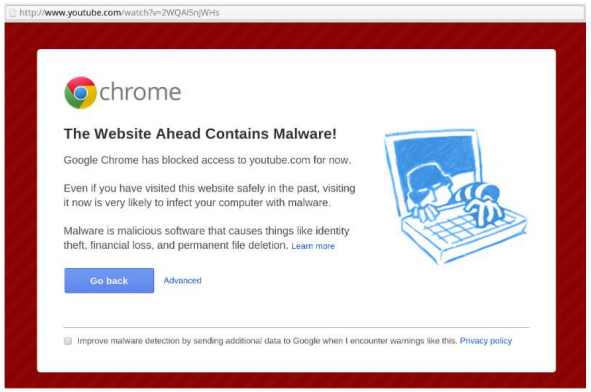


Figure 1: Malware warning in Google Chrome 32

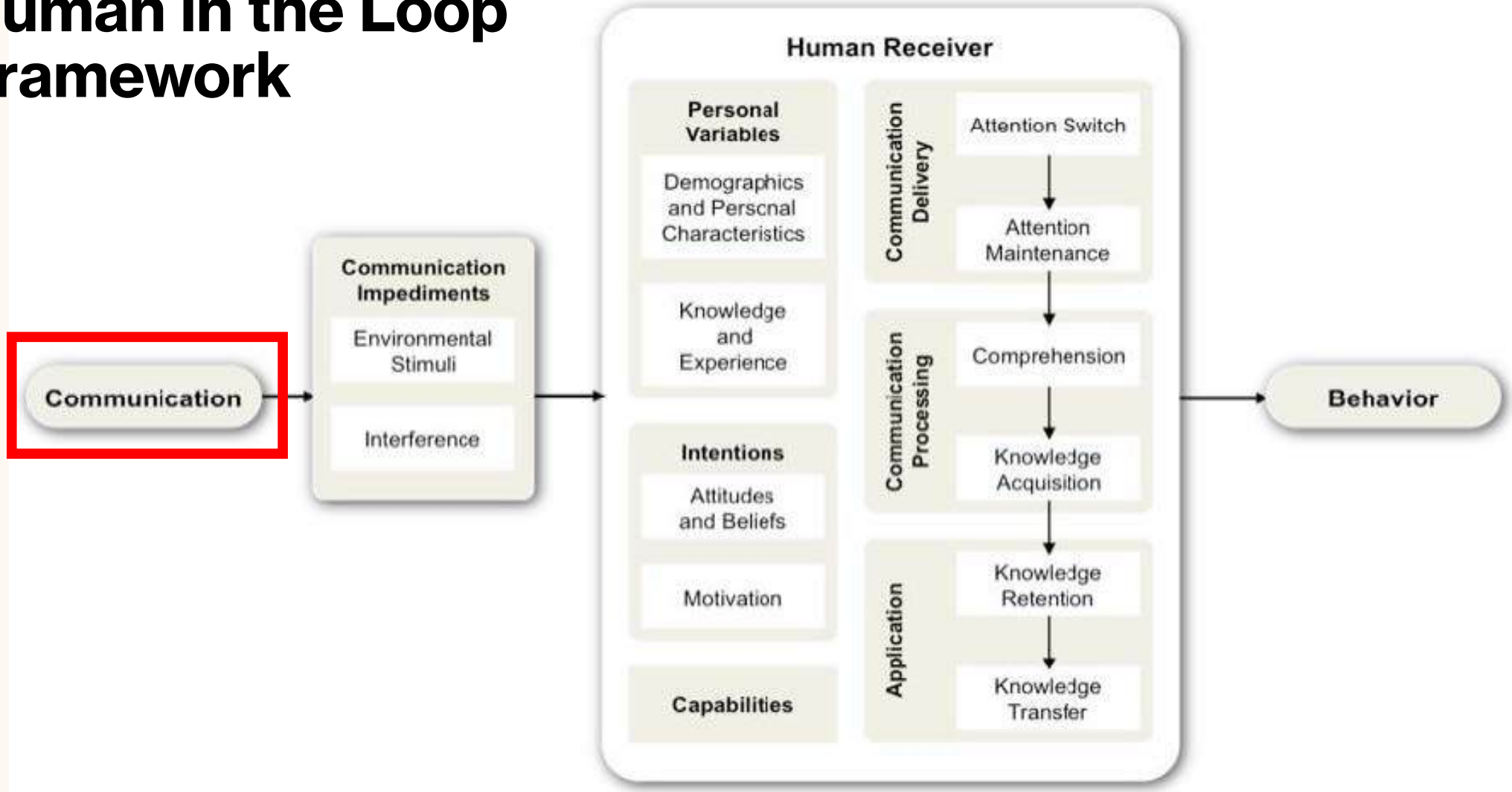
Amazon Mechanical Turk workers. We investigate the impact of people's familiarity with the website they are attempting to visit, as well as how they found out about the website. We also tested minor variations of the instrument used in our survey-based experiment to determine how small wording changes affected responses (e.g., whether or not participants were primed with the word "warning").

Our field data and Mechanical Turk experiment both support our familiarity hypothesis. In our analysis of 3,875,758

Frameworks

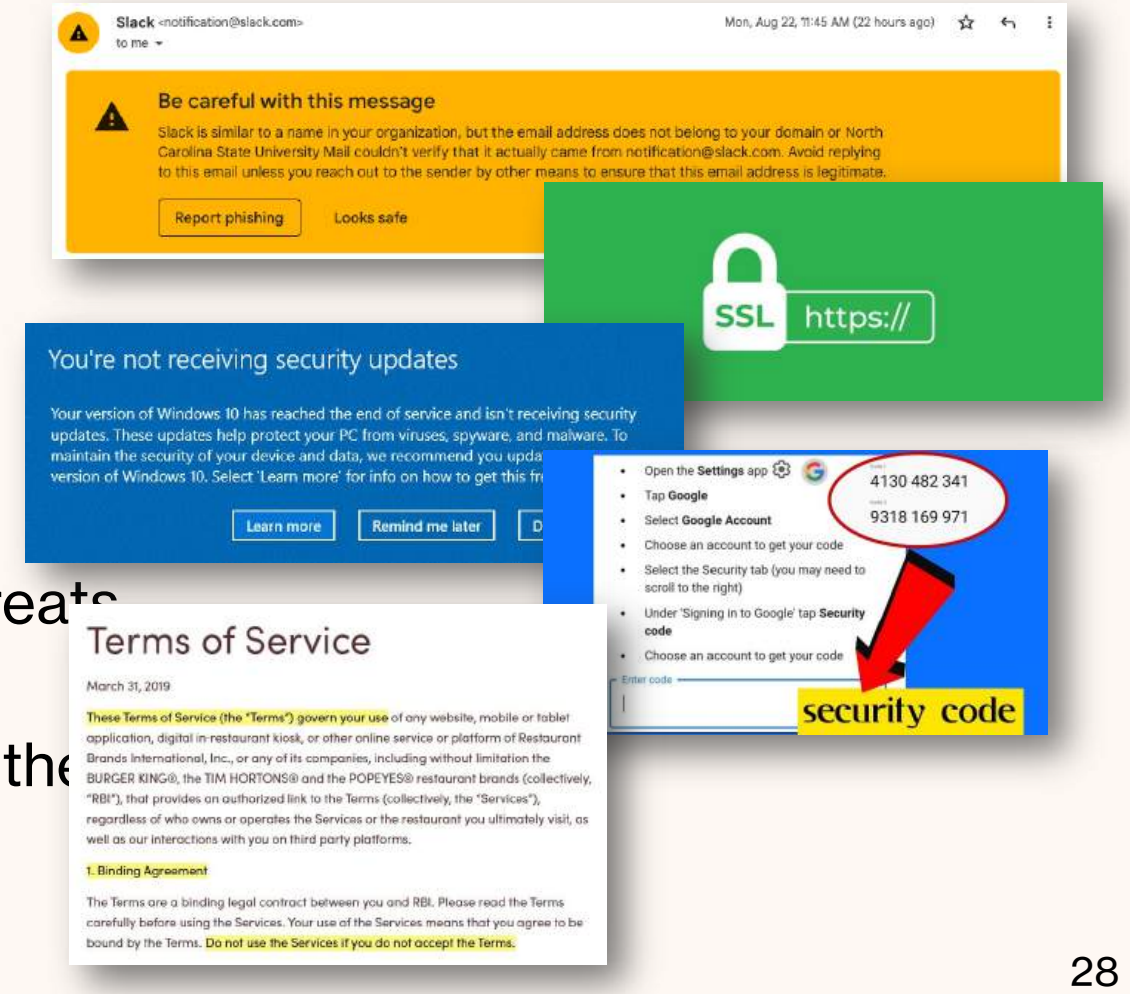
Frameworks help researchers structure their thinking around problems. Frameworks are proposed by experts in the field and represent how those people think about and break up certain types of problems.

Human in the Loop Framework

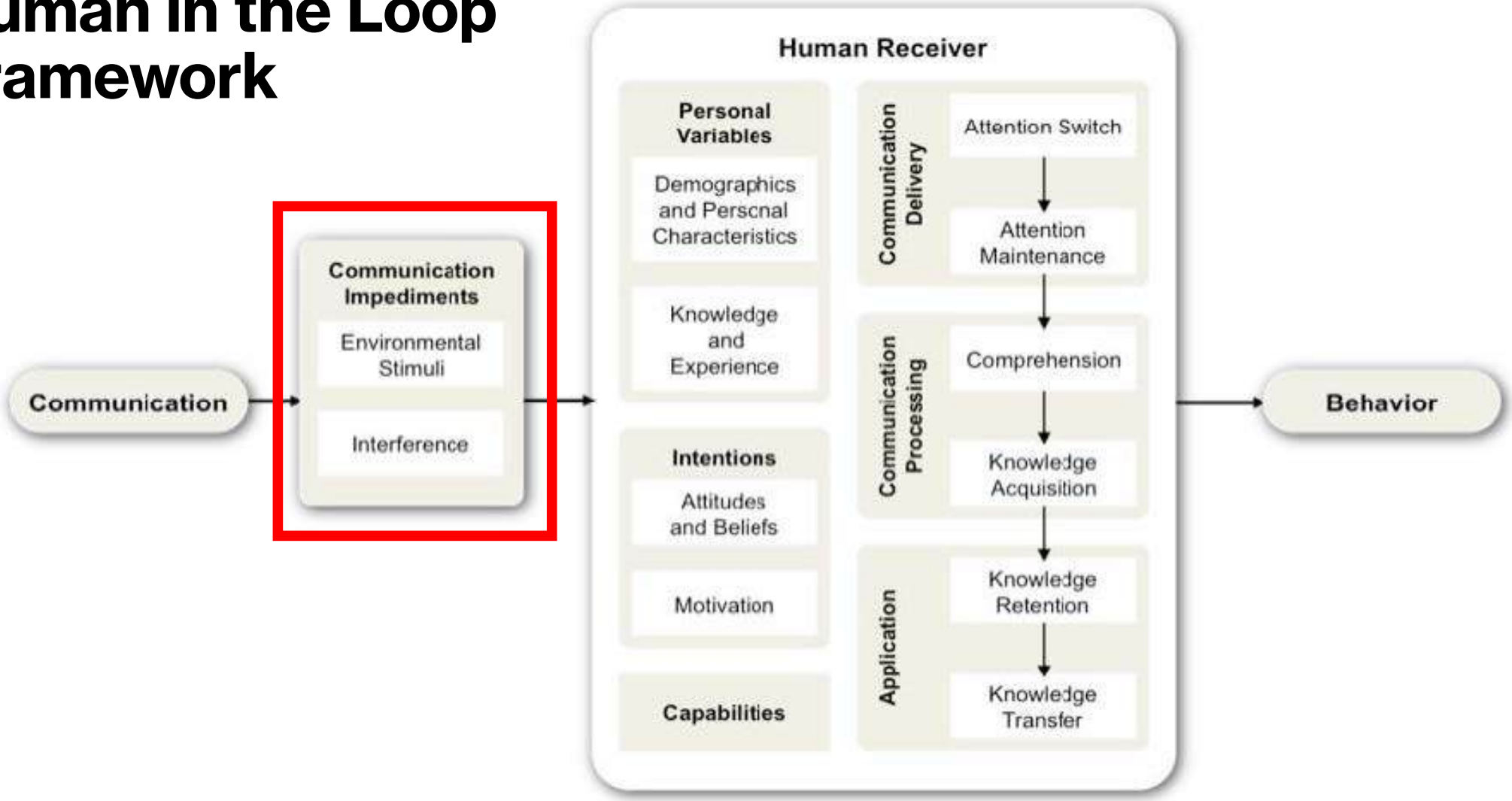


Human in the Loop: Communication

- **Warnings** alert users to avoid a hazard
- **Notices** inform users about characteristics of an object
- **Status indicators** inform users about system info.
- **Training** teaches users about threats and mitigation
- **Policy** informs users about what the are expected to comply with



Human in the Loop Framework

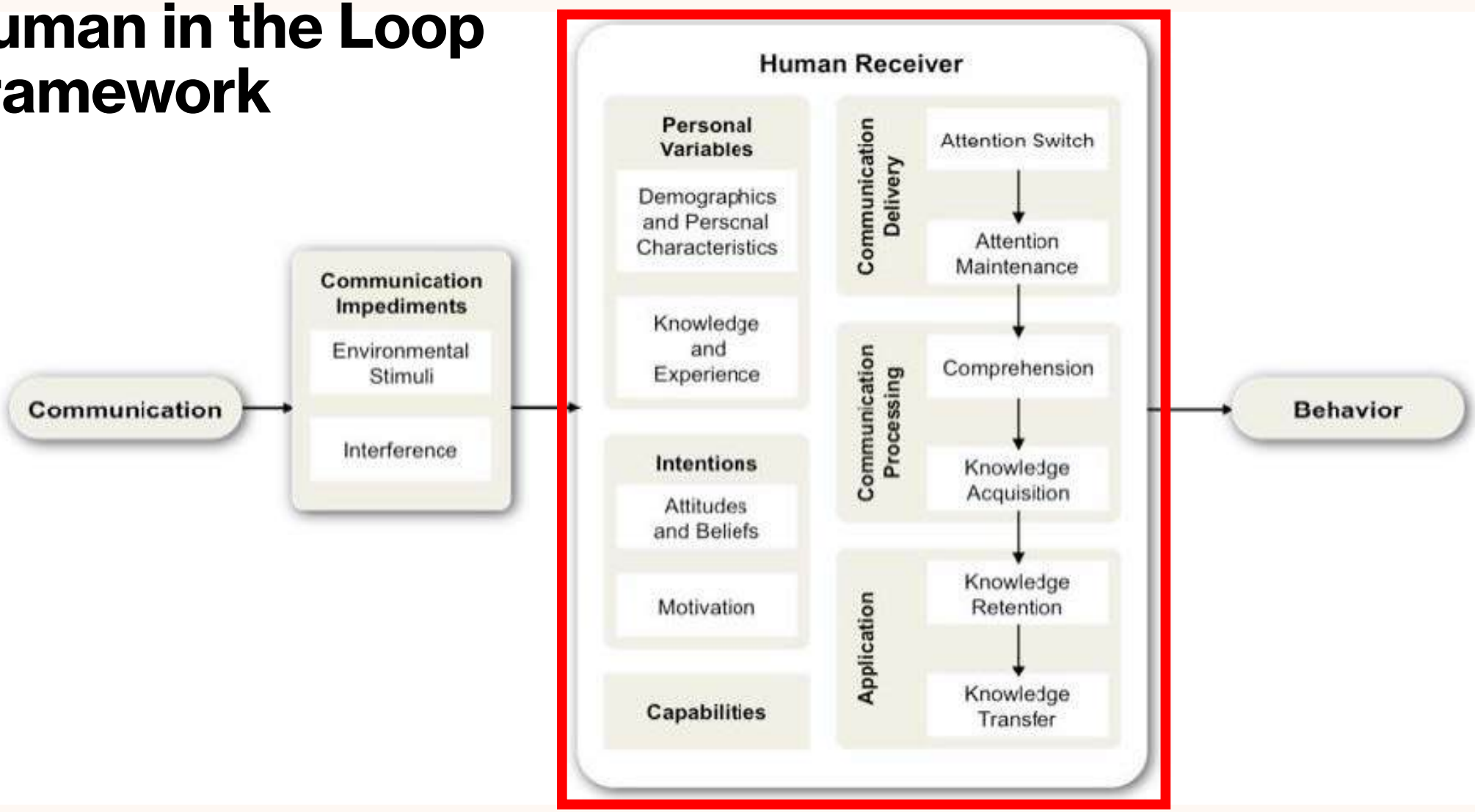



Sirani
COFFEE HOUSE
ESPRESSO DRINKS
DOUBLE ESPRESSO 2.29
AMERICANO 2.79
CAPPUCCINO 3.49
LATTE 3.79
VANILLA/MOCHA LATTE 4.29
JIRANI JUNCTION 4.29
CARAMEL MACCHIATO 4.49
ESPRESSO CON PANNA 2.99
BREWED
REG LRG
DRIP/ICED COFFEE 2.29 • 3.29
CAFÉ CON LECHE 2.59 • 3.59
POUR OVER 4.29 • —
COLD BREW 3.99 • 4.99
NITRO 3.99 • —
TEA DRINKS
HOT TEA 2.29
CHAI TEA LATTE 4.29
GREEN TEA LATTE 4.29
LONDON FOG 4.29
MORE
REG LRG
LEMONADE/ICED TEA 2.29 • 3.29
ITAL SODA 2.89 • 3.89
SMOOTHIE 4.89 • —
TRAIN FREEZE 4.89 • —
HOT CHOCOLATE 3.29 • 4.29
EXTRAS 75¢
SYRUP
ALMOND MILK
WHIPPED CREAM
thank you
Affogato
WIFI PASSWORD
YOU WILL NEED TO LOGIN TO HOMEGROWN
(NOT HOMEGROWN GUEST)
Syrups/Sauces
Vanilla
mocha
Caramel
Hazelnut
White Mocha
Peppermint
Coconut
Lavender
Sugar free Vanilla
Sugar free Mocha
Cherry

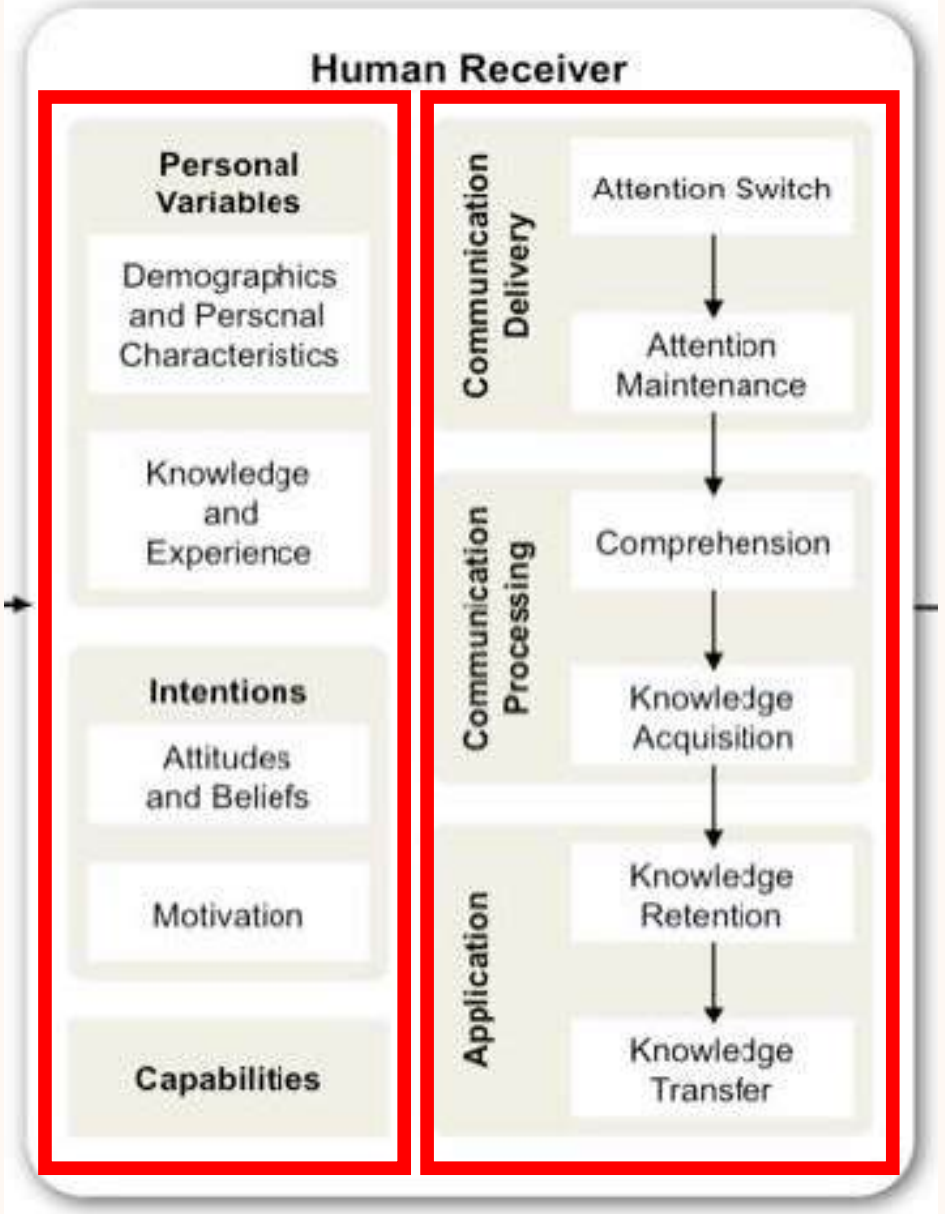
Human in the Loop: Communication Impediments

- **Environmental stimuli** (either related or unrelated) may divert users' attention away
- **Interference** prevents communication from being received as intended (can be malicious)

Human in the Loop Framework




Caution
Hot liquid


Human Receiver
Personal Variables
Demographics and Perscnal Characteristics
Knowledge and Experience
Intentions
Attitudes and Beliefs
Motivation
Capabilities
Communication Delivery
Attention Switch
Attention Maintenance
Communication Processing
Comprehension
Knowledge Acquisition
Application
Knowledge Retention
Knowledge Transfer

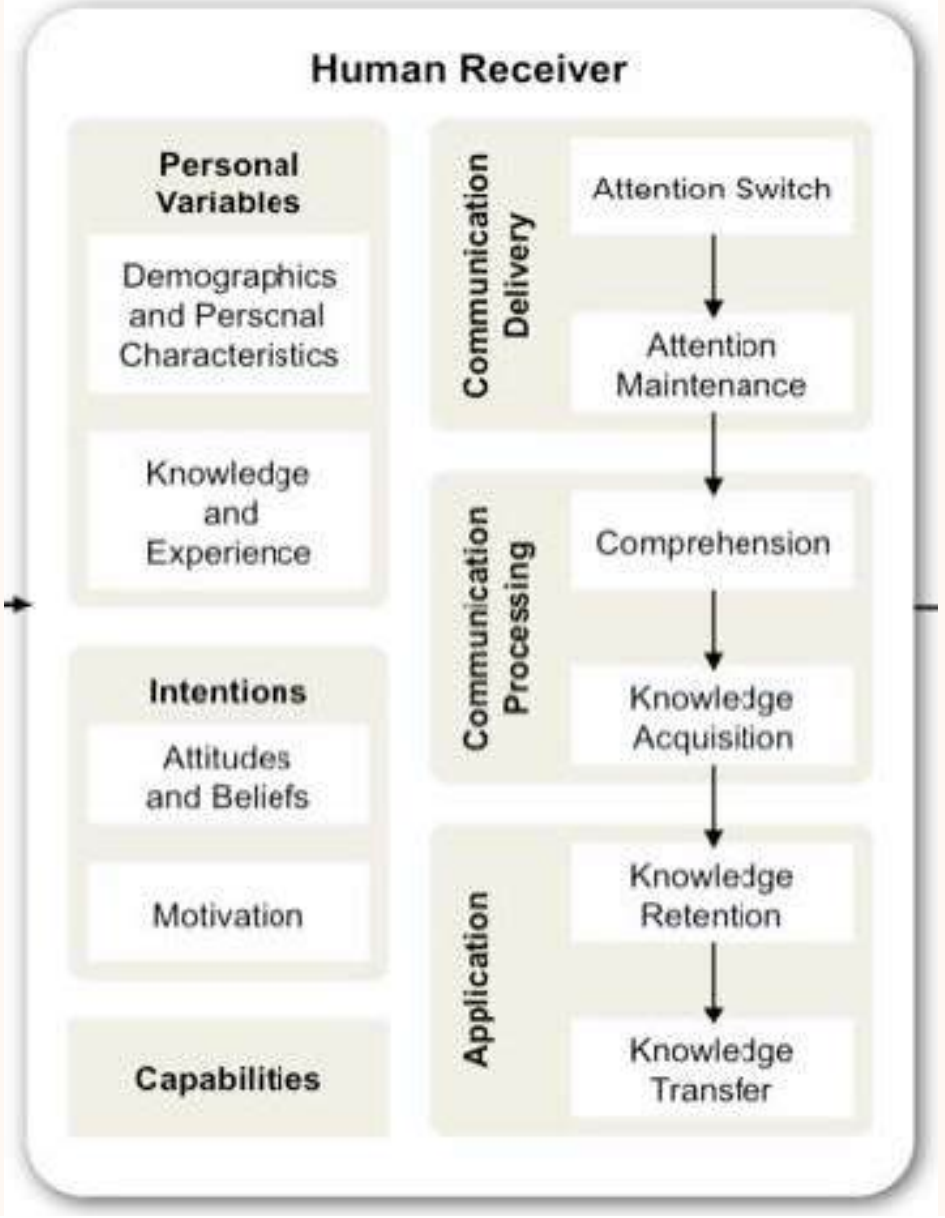
Human in the Loop: Human Receiver

- **Personal variables,** e.g., demographics, personal characteristics, knowledge , etc. – ability to comprehend and apply communications
- **Intentions** like attitudes, impacting the decision of whether to pay attention on a communication
- **Capabilities** to take proper actions

Human in the Loop: Human Receiver

- **Communication delivery:** should pay attention long enough to process it
- **Communication processing:** comprehend and acquire knowledge
- **Application:** retent the knowledge and knows when it’s applicable and to apply it


WIFI PASSWORD:
8282324340
YOU WILL NEED TO LOGIN TO
HOMEGROWN
(NOT HOMEGROWN GUEST)


Human Receiver
Personal Variables
Demographics and Perscnal Characteristics
Knowledge and Experience
Intentions
Attitudes and Beliefs
Motivation
Capabilities
Communication Delivery
Attention Switch
Attention Maintenance
Communication Processing
Comprehension
Knowledge Acquisition
Application
Knowledge Retention
Knowledge Transfer

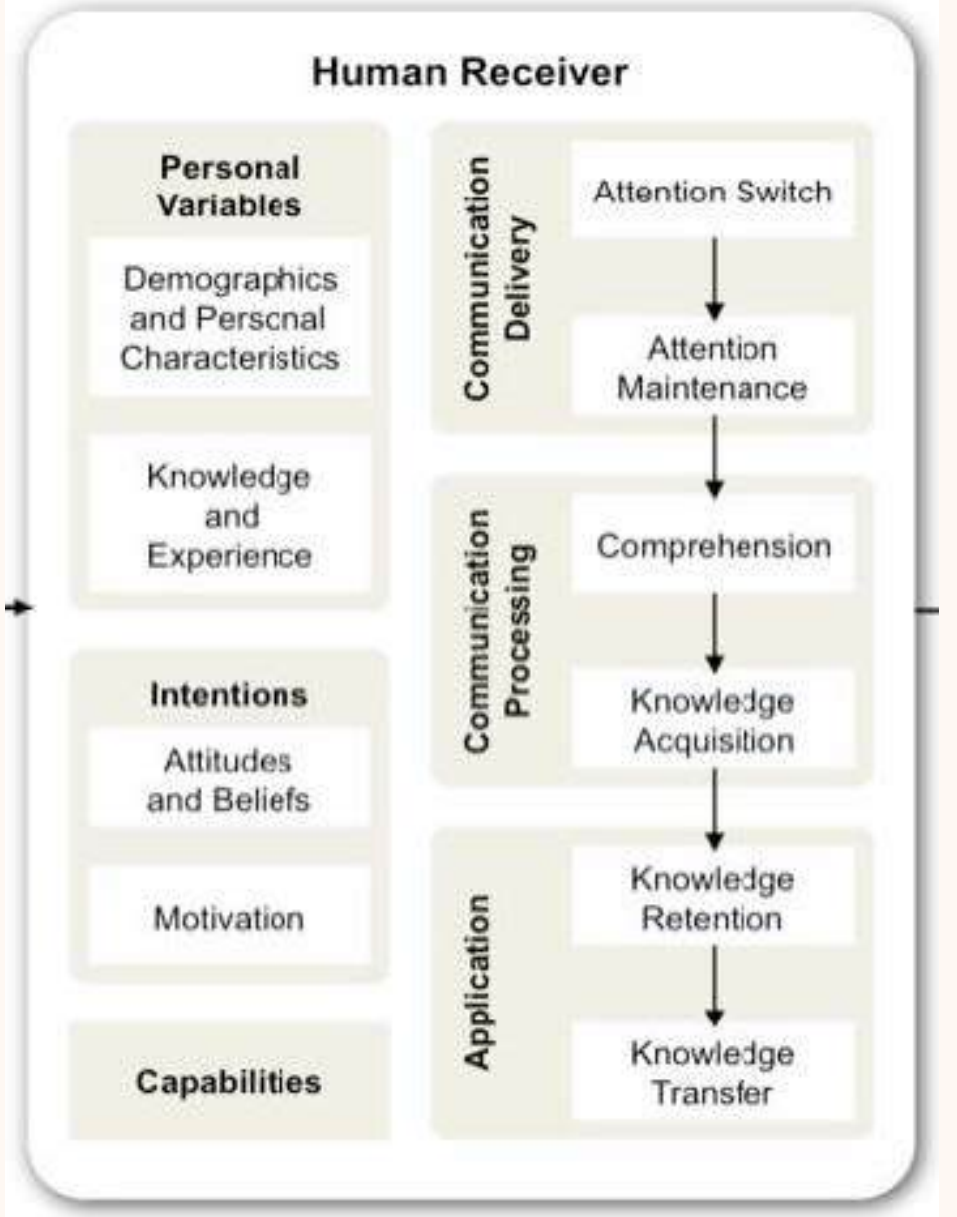
Google

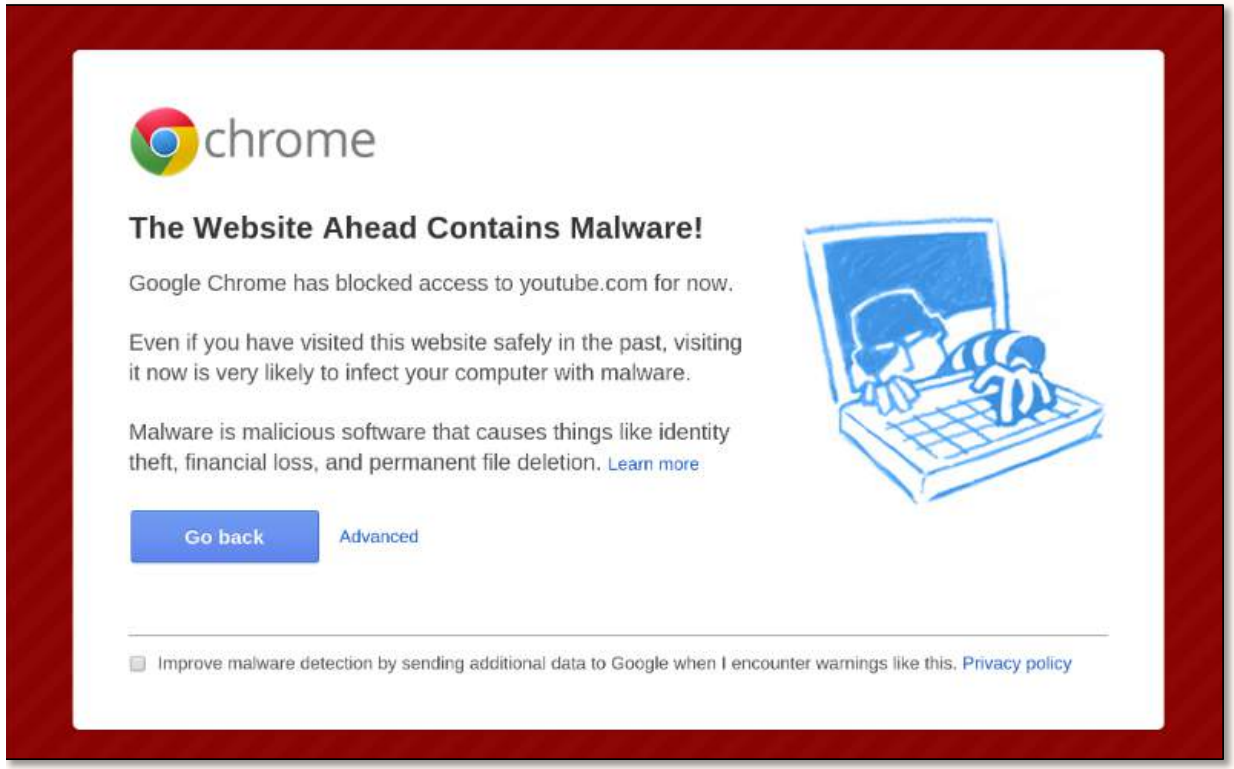
Someone knows the password to your linked Google Account

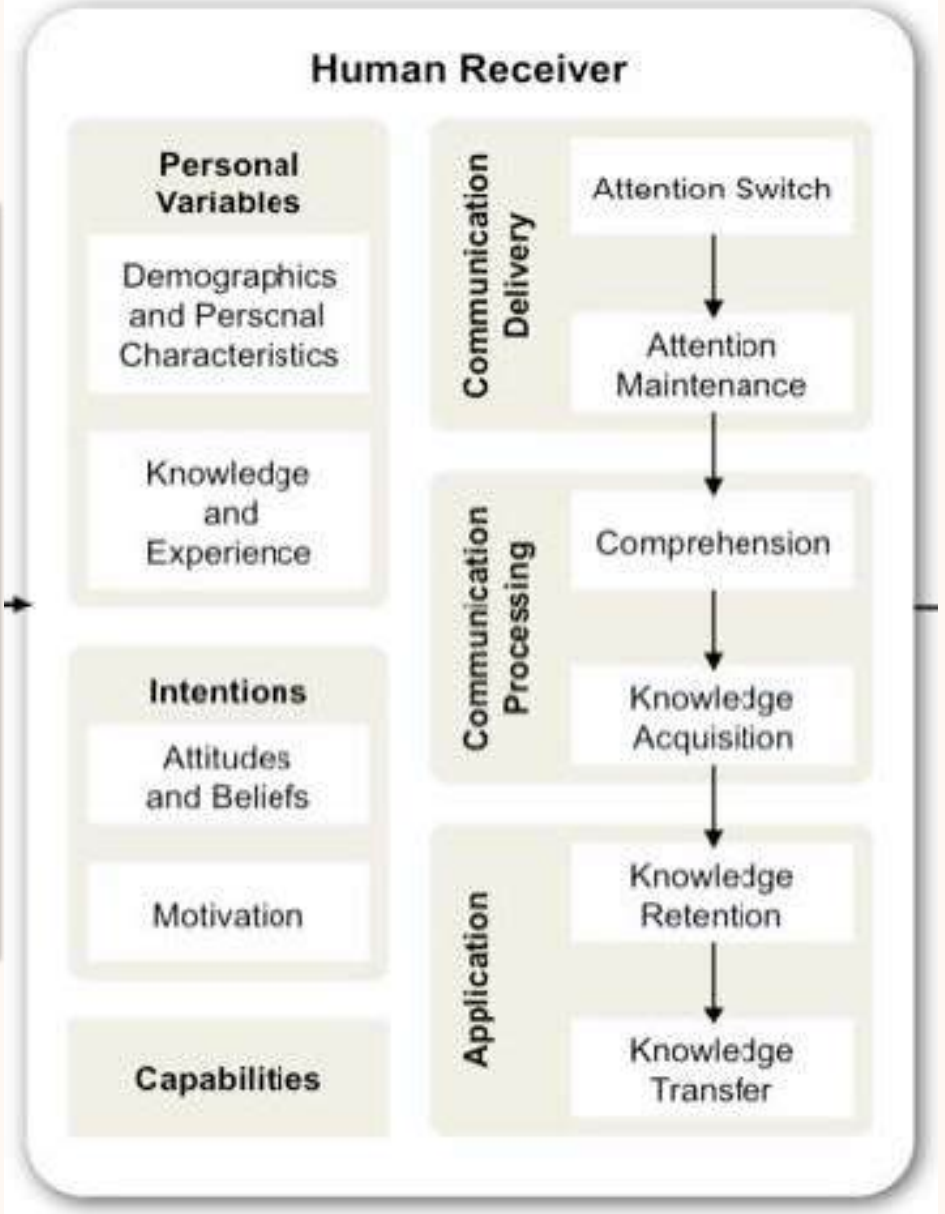
kaniea@gmail.com

Google has become aware that someone else knows your password, and we've taken steps to protect your account. Please sign back into your account now and choose a new password to secure your account.

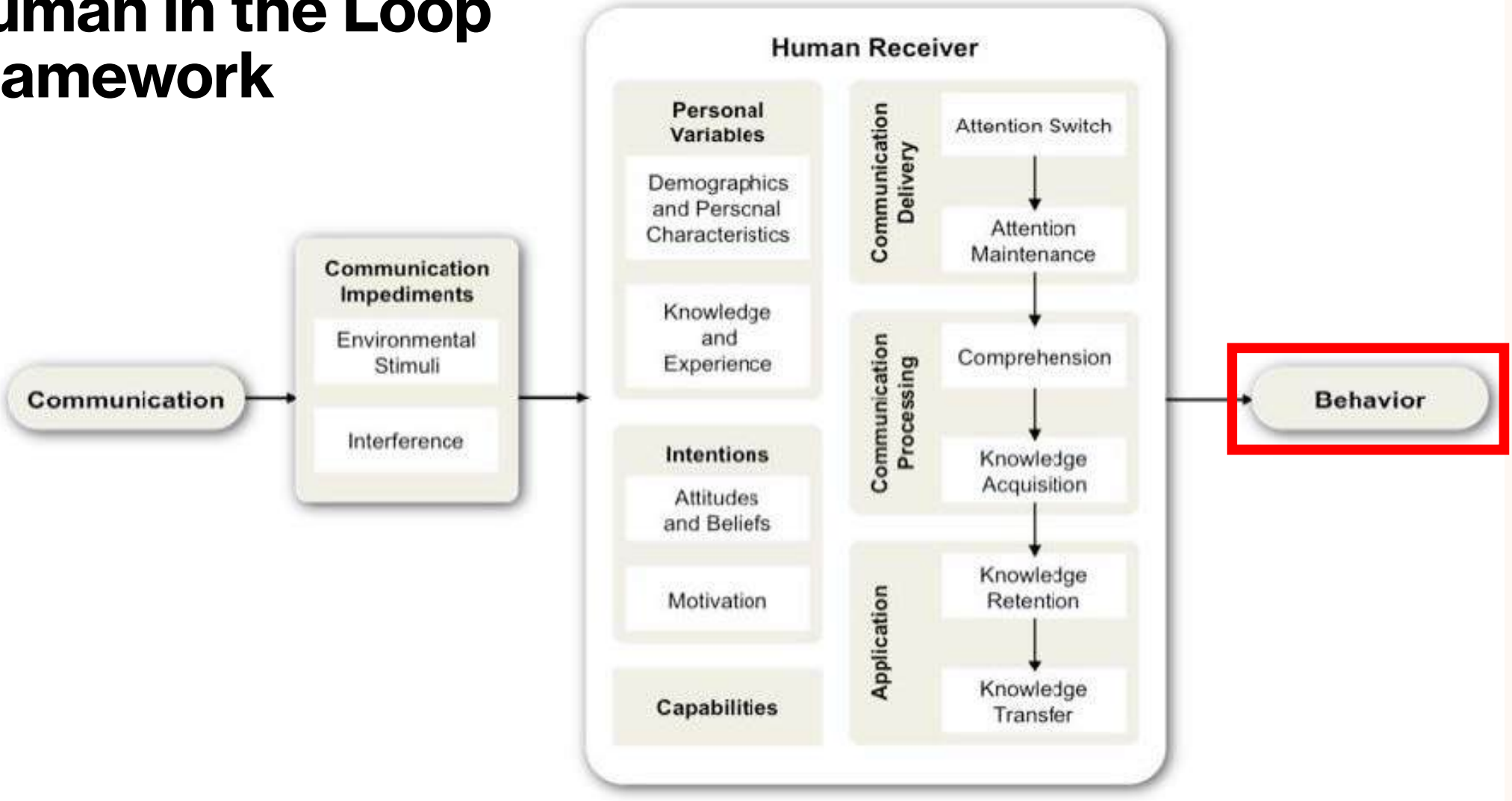
Learn more




chrome
The Website Ahead Contains Malware!
Google Chrome has blocked access to youtube.com for now.
Even if you have visited this website safely in the past, visiting it now is very likely to infect your computer with malware.
Malware is malicious software that causes things like identity theft, financial loss, and permanent file deletion. Learn more
Go back
Advanced
Improve malware detection by sending additional data to Google when I encounter warnings like this. Privacy policy


Human Receiver
Personal Variables
Demographics and Perscnal Characteristics
Knowledge and Experience
Intentions
Attitudes and Beliefs
Motivation
Capabilities
Communication Delivery
Attention Switch
Attention Maintenance
Communication Processing
Comprehension
Knowledge Acquisition
Application
Knowledge Retention
Knowledge Transfer

Human in the Loop Framework



Why do people ignore malware warnings and what can we do about it?

Your Reputation Precedes You: History, Reputation, and the Chrome Malware Warning

Hazim Almuhimedi
Carnegie Mellon University
hazim@cs.cmu.edu

Adrienne Porter Felt
Robert W. Reeder
Sunny Consolvo
Google, Inc.
felt, rreeder, sconsolvo@google.com

ABSTRACT

Several web browsers, including Google Chrome and Mozilla Firefox, use malware warnings to stop people from visiting infectious websites. However, users can choose to click through (i.e., ignore) these malware warnings. In Google Chrome, users click through a fifth of malware warnings on average. We investigate factors that may contribute to why people ignore such warnings. First, we examine field data to see how browsing history affects click-through rates. We find that users consistently heed warnings about websites that they have not visited before. However, users respond unpredictably to warnings about websites that they have previously visited. On some days, users ignore more than half of warnings about websites they've visited in the past. Next, we present results of an online, survey-based experiment that we ran to gain more insight into the effects of reputation on warning adherence. Participants said that they trusted high-reputation websites more than the warnings; however, their responses suggest that a notable minority of people could be swayed by providing more information. We provide recommendations for warning designers and pose open questions about the design of malware warnings.

1. INTRODUCTION

Modern browsers such as Google Chrome and Mozilla

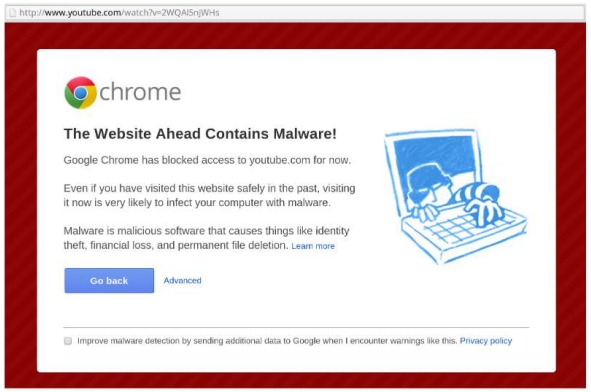


Figure 1: Malware warning in Google Chrome 32

Amazon Mechanical Turk workers. We investigate the impact of people's familiarity with the website they are attempting to visit, as well as how they found out about the website. We also tested minor variations of the instrument used in our survey-based experiment to determine how small wording changes affected responses (e.g., whether or not participants were primed with the word "warning").

Our field data and Mechanical Turk experiment both support our familiarity hypothesis. In our analysis of 3,875,758

Daily Click Through Rates (CTR) for igorning Chrome warning

Notice how they range from 10% to 27%

Something is happening on certain days

Date	CTR	N	Date	CTR	N
Tu Oct 01	15%	97,585	Tu Oct 15	16%	73,370
We Oct 02	15%	96,076	We Oct 16	18%	85,266
Th Oct 03	15%	104,075	Th Oct 17	15%	68,947
Fr Oct 04	16%	84,165	Fr Oct 18	11%	132,410
Sa Oct 05	15%	80,433	Sa Oct 19	10%	99,778
Su Oct 06	15%	77,931	Su Oct 20	12%	95,163
Mo Oct 07	16%	80,640	Mo Oct 21	14%	91,651
Tu Oct 08	17%	90,356	Tu Oct 22	21%	131,700
We Oct 09	21%	145,893	We Oct 23	18%	121,944
Th Oct 10	21%	96,159	Th Oct 24	24%	151,387
Fr Oct 11	23%	93,059	Fr Oct 25	27%	117,002
Sa Oct 12	15%	79,295	Sa Oct 26	14%	64,740
Su Oct 13	15%	79,134	Su Oct 27	14%	70,713
Mo Oct 14	18%	89,180	Mo Oct 28	15%	59,567

Table 1: Chrome malware warning click-through rates (CTRs) and sample sizes for October 2013. Darker shaded values indicate higher CTRs. Note the wide variance in daily CTRs.

Idea: Maybe Chrome is blocking popular websites on those days?

Counter example: On one day when YouTube was blocked, CTR dropped from the normal 15% to 8%.

Because it was a large site being blocked, social media and news both told users to heed the warning which they did.

The reality is more complex than popularity alone – news media, word of mouth, and other factors might influence user behavior

Idea: Maybe people ignore warnings on sites they visit often and think are safe.

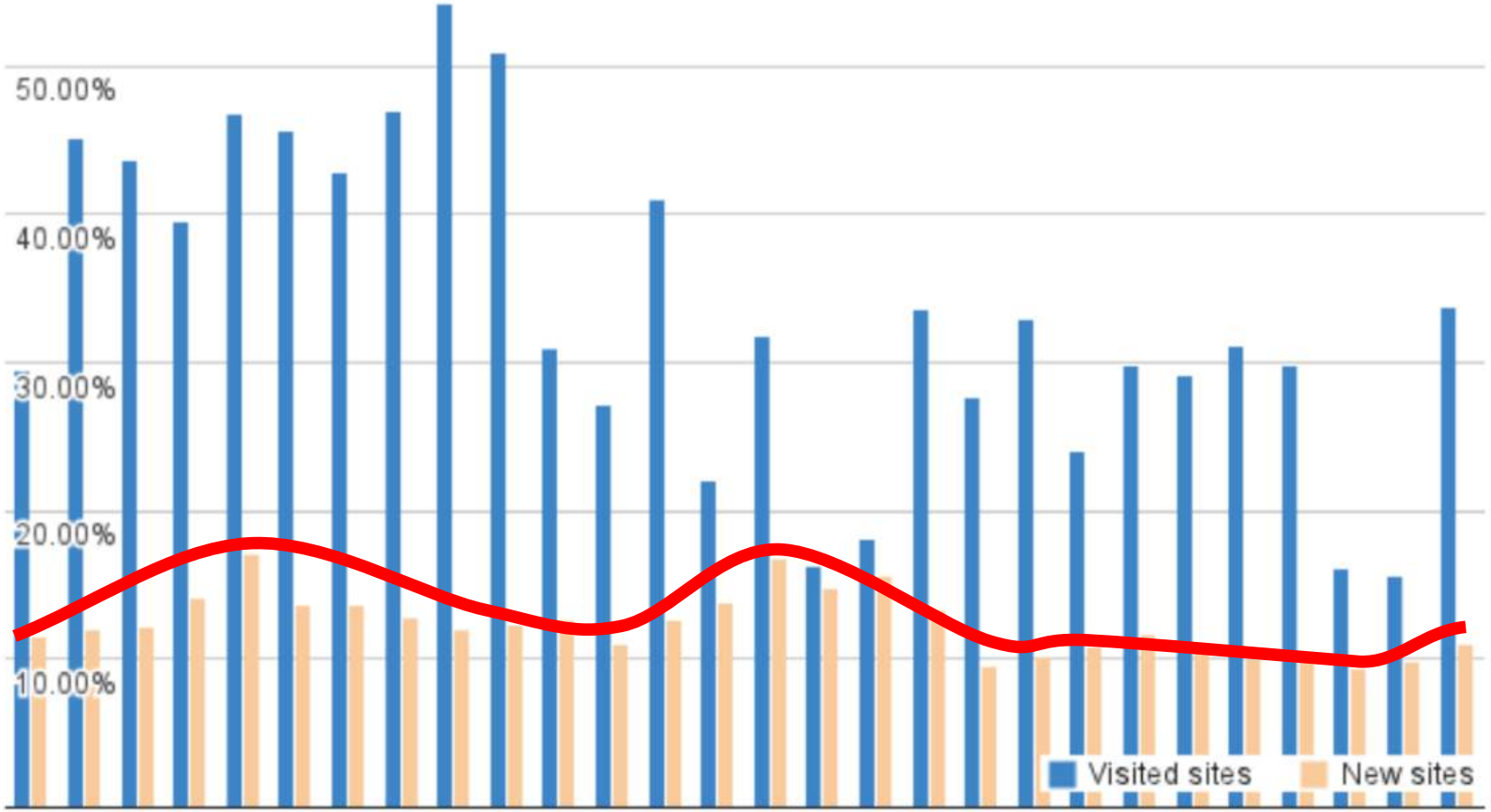
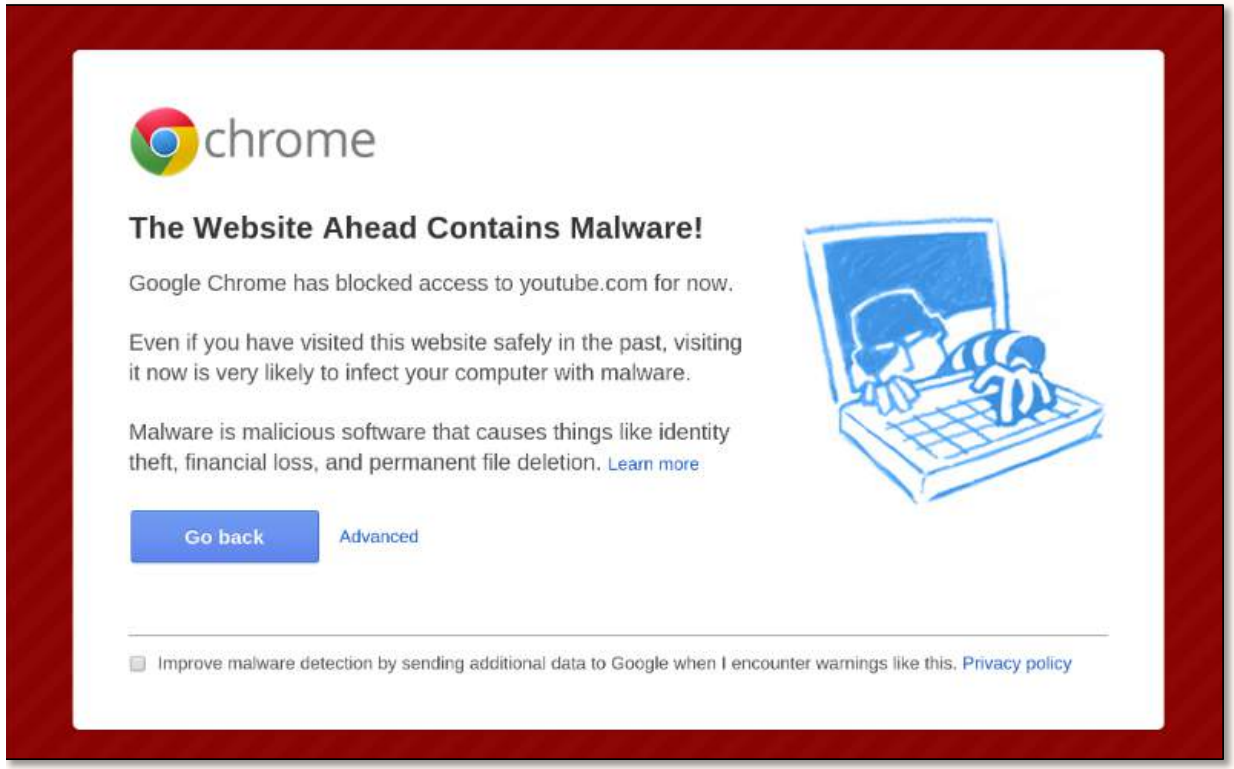
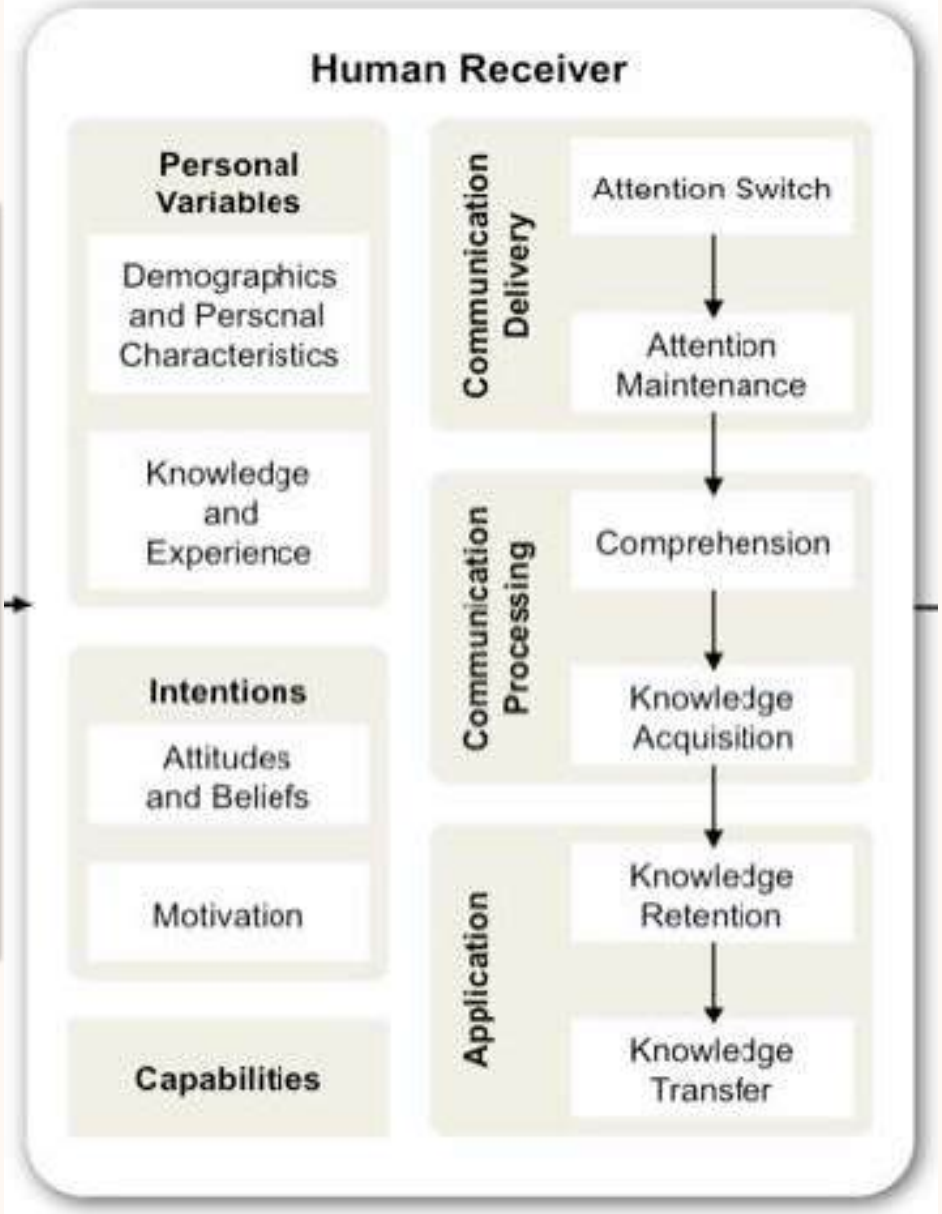


Figure 3: Daily CTR, separated by whether the website was already in the user's browsing history. For 28 days in January-February 2014.


chrome
The Website Ahead Contains Malware!
Google Chrome has blocked access to youtube.com for now.
Even if you have visited this website safely in the past, visiting it now is very likely to infect your computer with malware.
Malware is malicious software that causes things like identity theft, financial loss, and permanent file deletion. Learn more
Go back
Advanced
Improve malware detection by sending additional data to Google when I encounter warnings like this. Privacy policy


Human Receiver
Personal Variables
Demographics and Perscnal Characteristics
Knowledge and Experience
Intentions
Attitudes and Beliefs
Motivation
Capabilities
Communication Delivery
Attention Switch
Attention Maintenance
Communication Processing
Comprehension
Knowledge Acquisition
Application
Knowledge Retention
Knowledge Transfer

How might you test if this effect is really due to familiarity ?

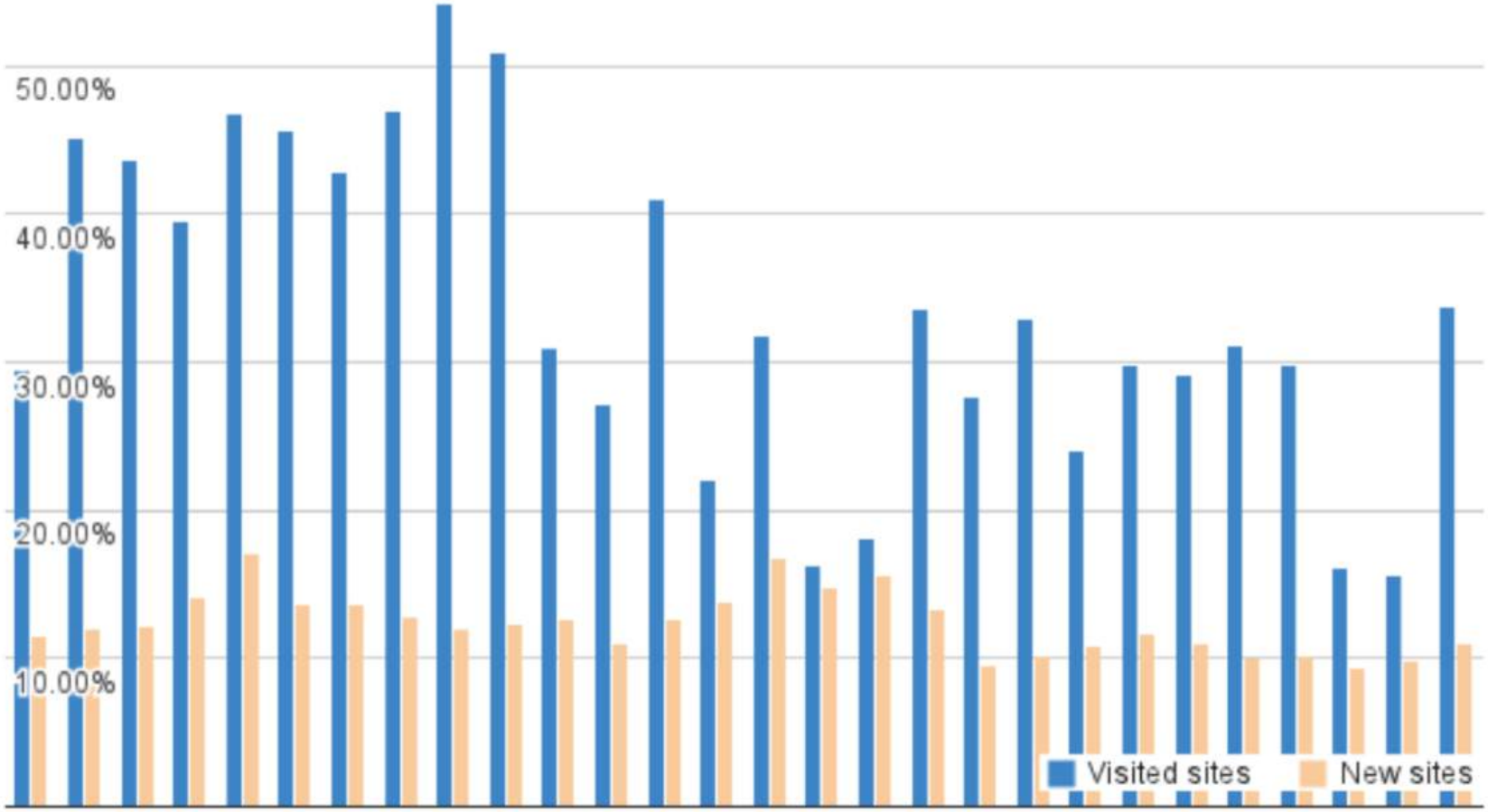


Figure 3: Daily CTR, separated by whether the website was already in the user's browsing history. For 28 days in January-February 2014.

The paper's approach:

- "We asked 1,397 Mechanical Turk workers to tell us how they would react to screenshots of Google Chrome malware warnings."
- 2X2 study design. Varying the reputation of the referring source and the reputation of the destination page.

		Referring person or site reputation	
		High	Low
Destination site reputation	High	High, High	High, Low
	Low	Low, High	Low, Low

http://www.youtube.com/watch?v=2WQAl5nJWHs

chrome

The Website Ahead Contains Malware!

Google Chrome has blocked access to youtube.com for now.

Even if you have visited this website safely in the past, visiting it now is very likely to infect your computer with malware.

Malware is malicious software that causes things like identity theft, financial loss, and permanent file deletion. Learn more

Go back Details about problems on this website Proceed at your own risk «

Improve malware detection by sending additional data to Google when I encounter warnings like this. Privacy policy

“YouTube is a well known website. I would assume the malware block is an error.”

“Because I frequent youtube.com a lot and I have never gotten any malware.”

Conclusion

- Users have complex decision making processes around security
- Testing those processes requires thought and multiple rounds of data gathering and testing.
- Studies need to be setup to handle correlation vs causation concerns.
- Don’t stop digging if you get an easy answer, validate it.

Planning research studies

Research Studies

1. Define your research question
2. Identify your variables
3. Run your study
4. Evaluate the outcome

Step 1: Define your research question

Some research questions:

- Can people differentiate between a subdomain and a domain when reading a URL?
- Does [my new system] help people differentiate between malicious URLs and safe ones?
- Can users use [my new password manager] faster and with less errors than [the old password manager]?
- Does knowing how an app will use its permissions impact app installation decisions?
- What factors impact end-users' willingness to update software?
- Using [website], can users successfully opt-out of cookie tracking without forming inaccurate mental models?

For task based lab studies

- First decide what “usable” means
- Identify what you think your users need to be able to do using your system or what kind of attitude you want them to have
- The goals need to be specific and easy to identify if they have or have not been completed
- Examples:
  - Find a stool on a shopping page and purchase it
  - Be willing to give the app 5 stars after interacting with it for the first time
- Bad examples:
  - Have fun using the site
  - Find a bus to go somewhere

“Usable” could mean:

- User can accomplish a task in Y minutes
- User can accomplish task with no unrecoverable errors
- After interacting with an interface the user has an accurate mental model of when their message is and is not encrypted
- User feels more confident in using secure messaging
- Users voluntarily select higher entropy passwords
- User creates a password that they can remember after a month of not using it

Step 2: Identify your variables

What kind of data do you want?

- Attitudinal – User attitudes and opinions

vs.

- Behavioral – What the user actually does or is capable of doing

- Qualitative – Unstructured data. Typically unstructured language data

vs.

- Quantitative – Structured data. Typically numerical data that can be summed or counted

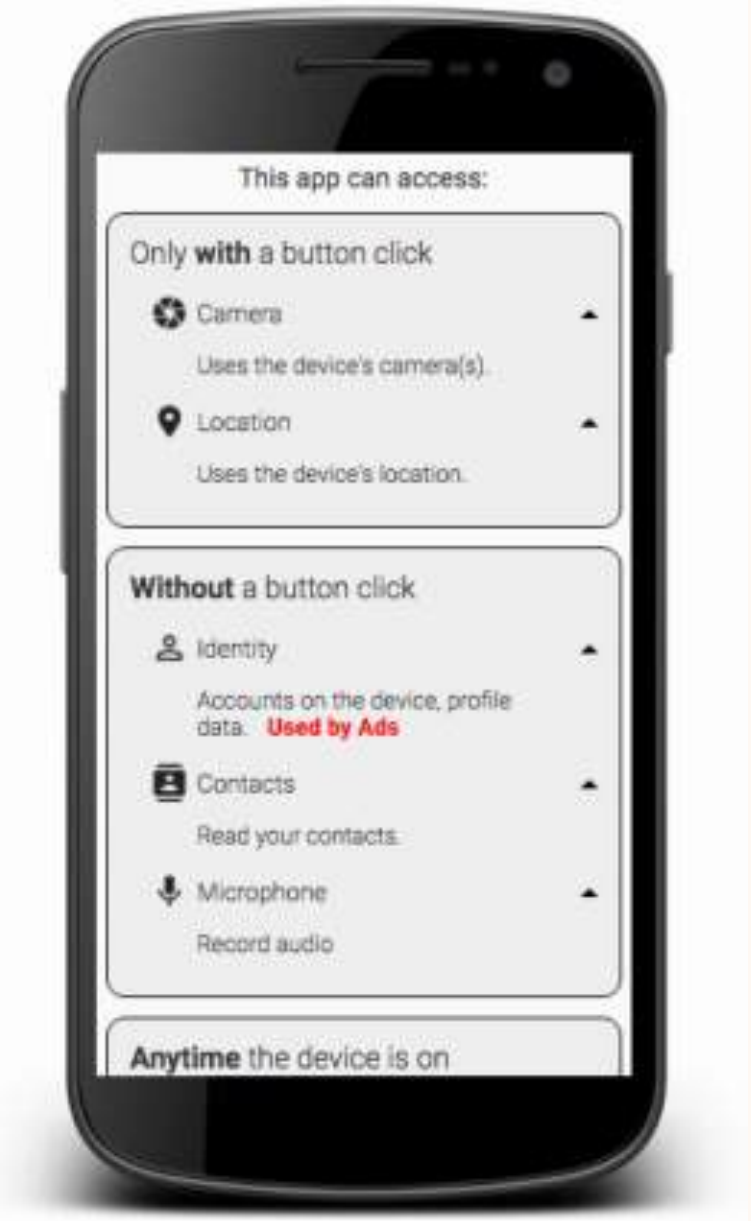
https://www.nngroup.com/articles/which-ux-research-methods/

For quantitative studies

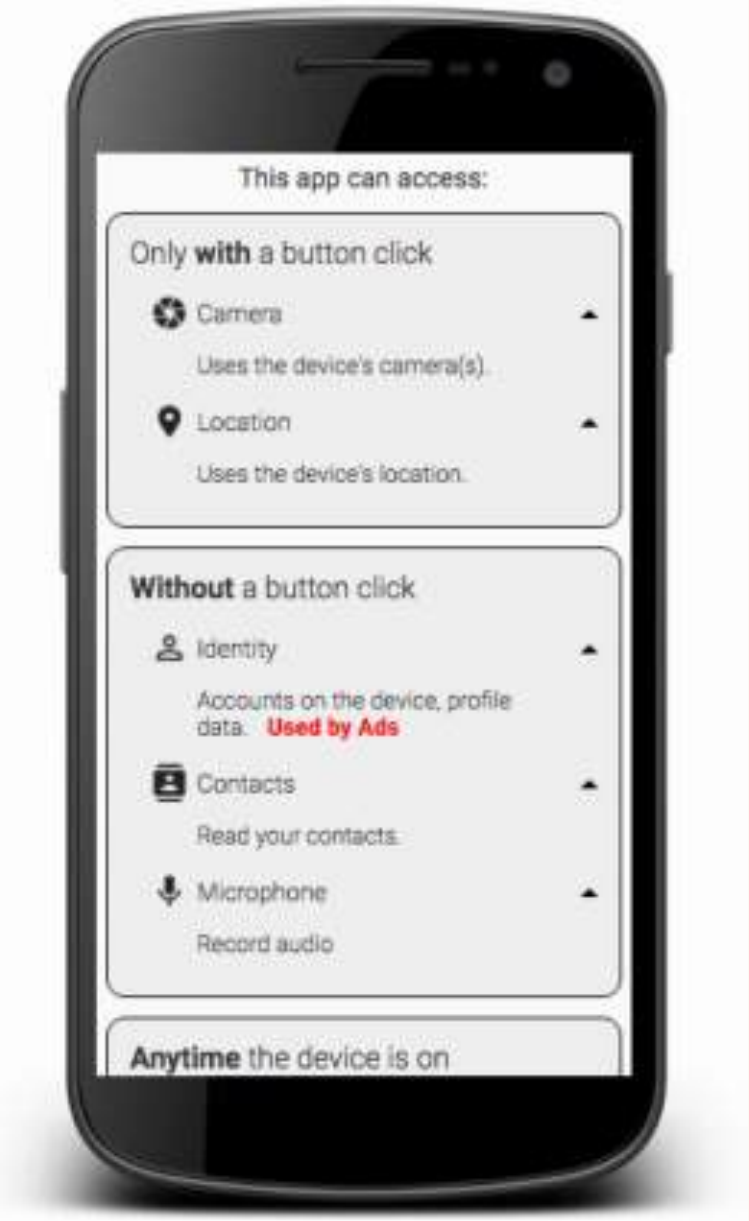
What are you going to measure?

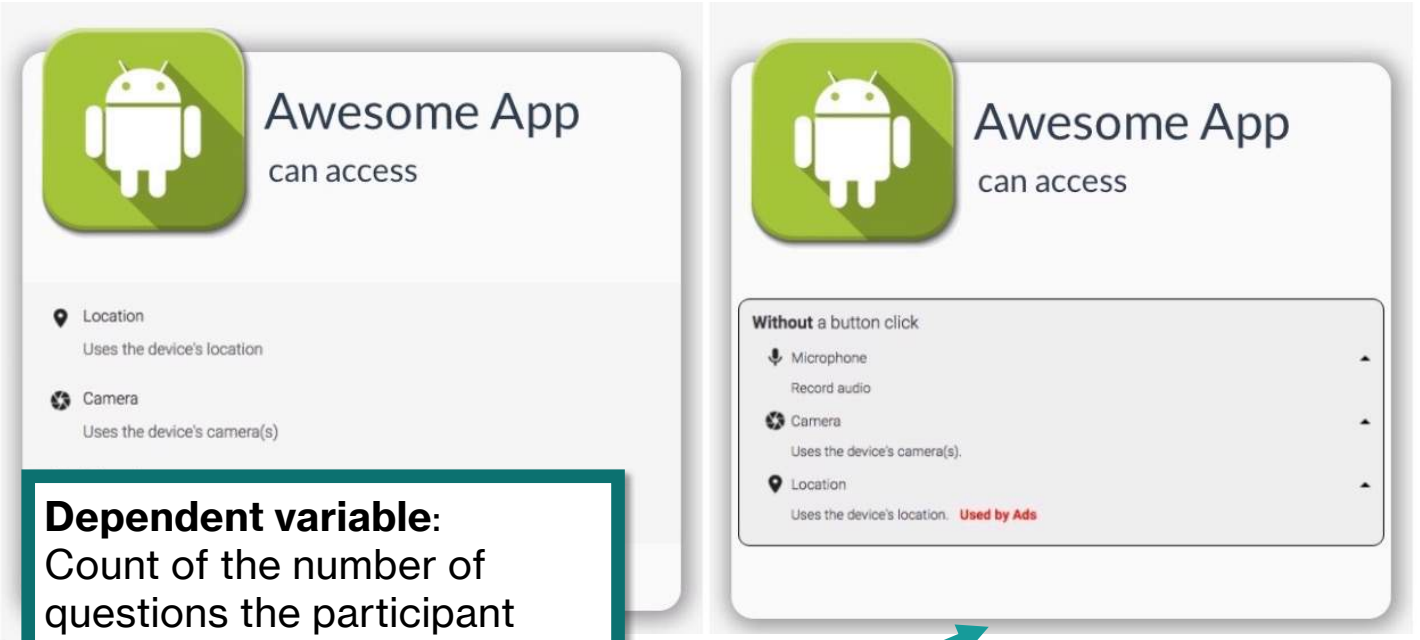
- In statistics there are classically two types of measurements (variables): dependent and independent
- Dependent
  - Also known as the outcome variable
  - “Dependent” on the study
  - Measures the usability goal
- Independent
  - Anything you are directly manipulating
  - An element of the study which is under your control
  - A pre-existing feature of your participant

Lets use this study as an example



Research Question:
Can users reliably identify if an app can or cannot perform an action directly tied to a permission.





Dependent variable:
Count of the number of questions the participant answered correctly

Independent variable:
Which of the two interfaces the participant was shown

g can this app do?

					Absolutely Possible
Charge purchases to your credit card at any time.					○
Get your location.	○	○	○	○	○
Allow ads to know your location.	○	○	○	○	○
Load ads.	○	○	○	○	○
Write on the SD card					

Variables that would make sense

- Research Question: Can users reliably identify if an app can or cannot perform an action directly tied to a permission?
- Dependent
  - Number of permissions correctly/incorrectly read
  - Time spent reading each permission screen
- Independent
  - Study group (which screen shown)
  - If the permission was privacy sensitive or not
  - Order of the tasks
  - Time of day
  - Type of most used device (laptop, mobile, PC)
  - Demographics of the participants (gender, age, native language, ...)

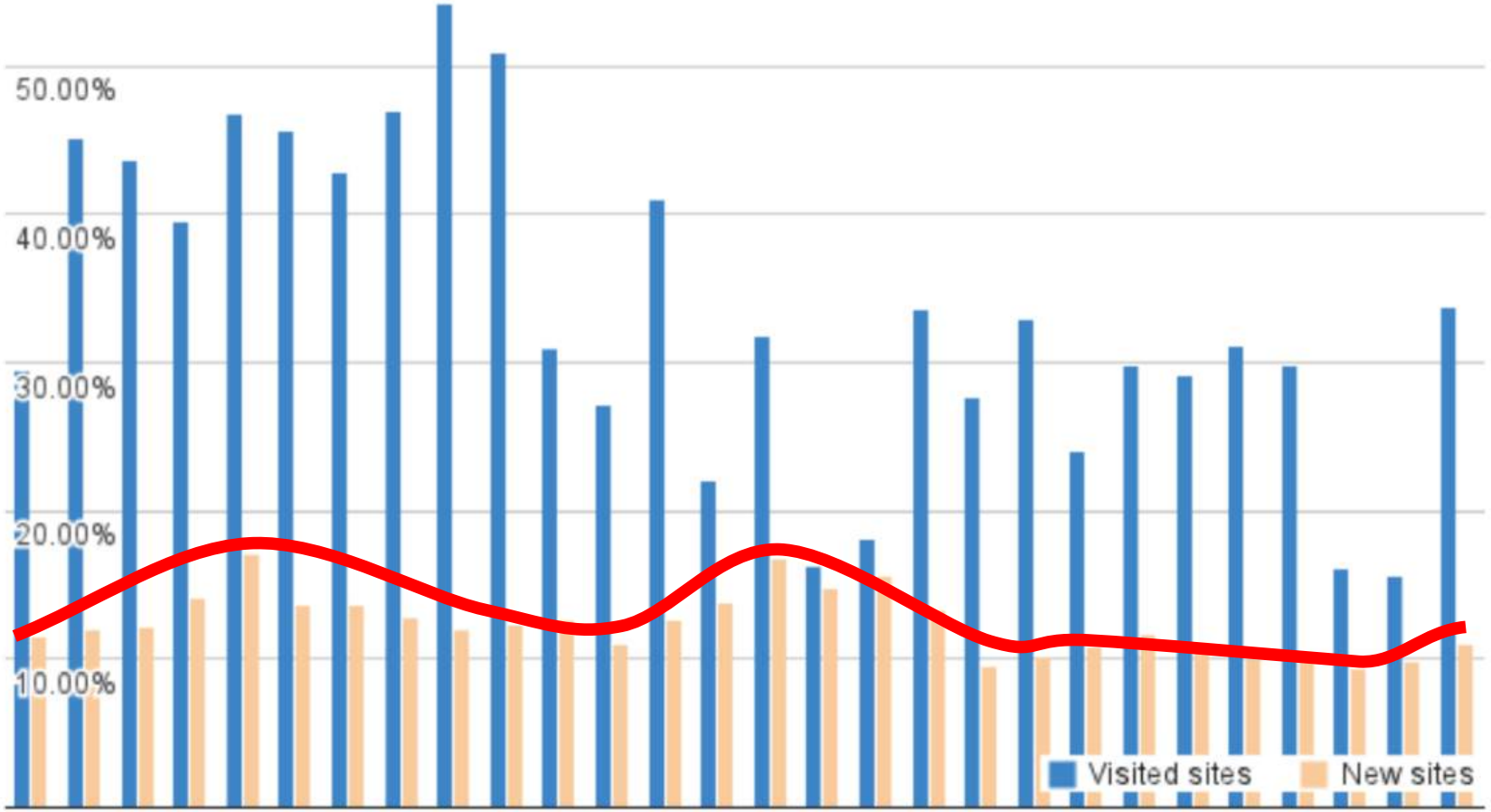


Figure 3: Daily CTR, separated by whether the website was already in the user's browsing history. For 28 days in January-February 2014.

Common dependent things to measure

- Number of dangerous errors made
- Time to complete task
- Percent of task completed
- Percent of task completed per unit of time
- Ratio of successes to failures
- Time spent in errors
- Percent or number of errors
- Percent or number of competitors better than it
- Frequency of help and documentation use

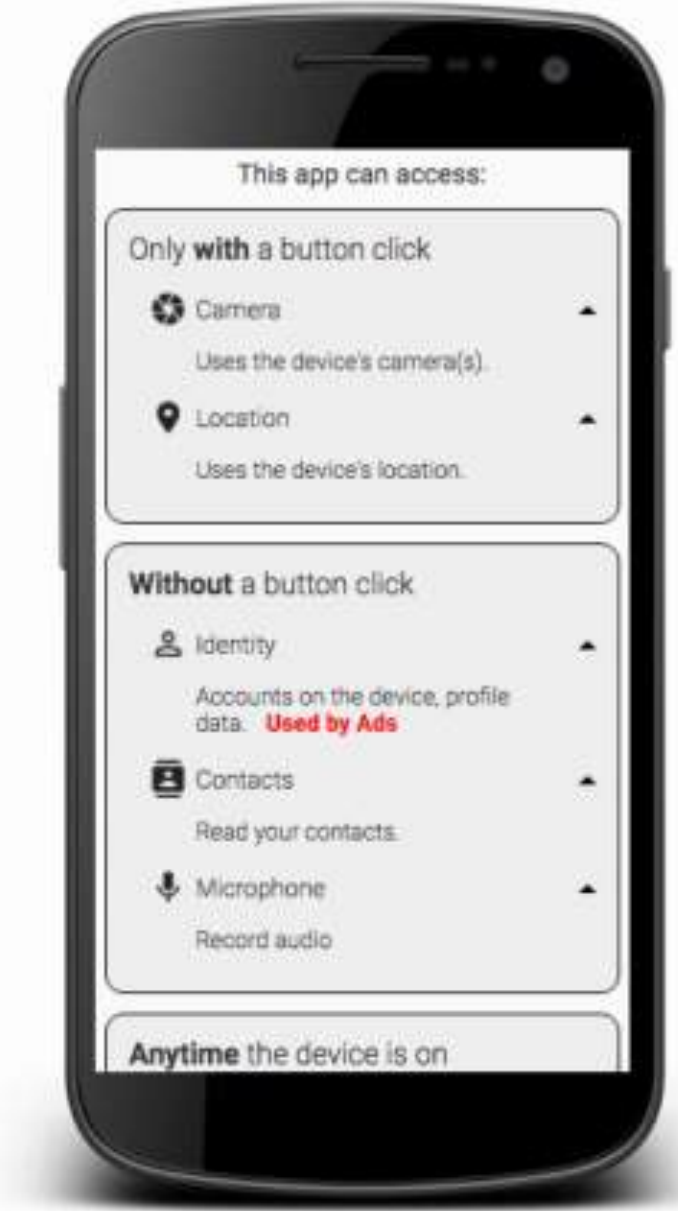
Step 3: Setup your study

Between vs. Within subjects

- Between subjects
  - Your study only shows one interface to one person
  - You are measuring how well the people randomly assigned to the A interface did compared to the people randomly assigned to the B interface
  - Lots of variability with this method
- Within subjects
  - Your study shows all interfaces to all people
  - You are measuring the difference in how they do on the two interfaces
  - Less variability (same person) but more learning effects and priming

Study design

- RQ: Does [my new interface] enable people to accurately determine what permissions an app will use?
- A/B test between the existing and new interface
- Between subjects
- 10 Tasks shown in the same order to all participants
- Dependent variables
  - Accuracy on task
- Independent variables
  - Which interface (A or B)



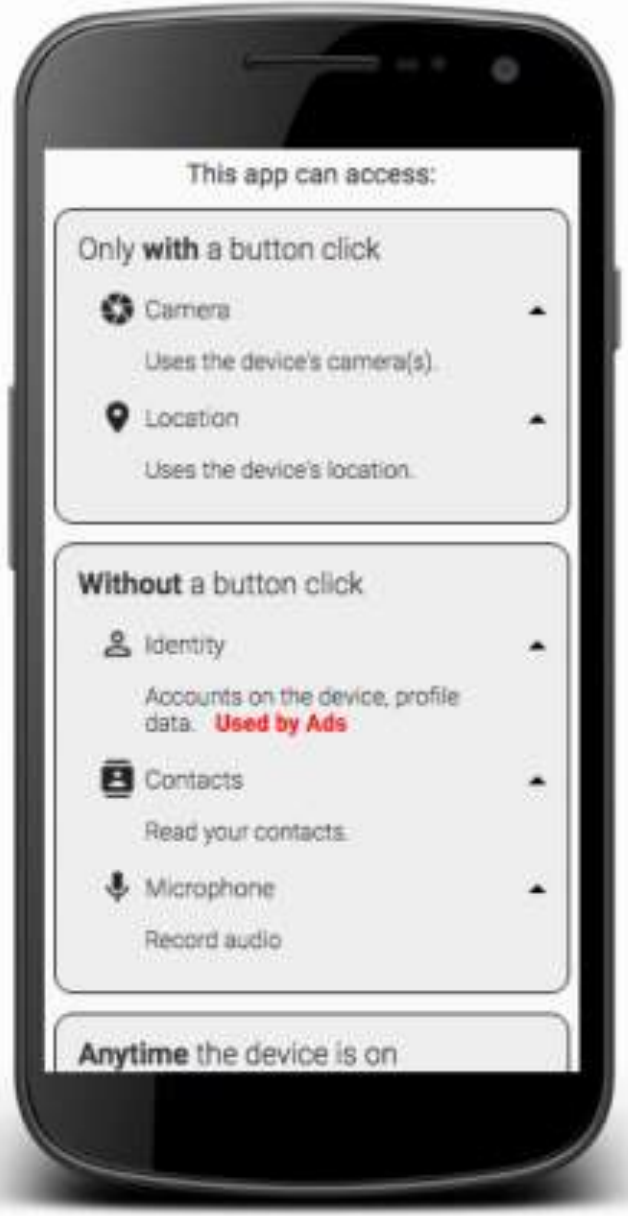
Step 4: Evaluate the outcome

Types of data

- Numeric
  - **Continuous** – Any value on the range is possible including decimal (1-5)
  - **Discrete** – Only certain values on the range are possible (1,2,3,4,5)
  - **Interval** – Only certain values on the rage are possible and each has equal distance from its neighboring values (strongly agree, agree, neutral, disagree, strongly disagree)
- Categorical
  - **Binary** – Only two possibilities (true, false)
  - **Ordinal** – The values have an ordering (slow, medium, fast)
  - **Nominal** – The values have no ordering (apple, pear, kiwi, banana)

Study design

- Accuracy on all tasks
  - Discrete
- Which interface
  - Categorical binary



Take-home

- **(Blog)** Habib, H., Zou, Y., Yao, Y., Acquisti, A., Cranor, L., Reidenberg, J., Sadeh, N. and Schaub, F., 2021, May. Toggles, dollar signs, and triangles: How to (in) effectively convey privacy choices with icons and link texts. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-25).
- **(Blog)** Guardian - UK's AI Safety Institute 'needs to set standards rather than do testing'