Haojian Jin

What we have covered.

- Location privacy
- Permissions for Privacy
- Policies for Privacy
- Privacy Norms/Contextual Integrity

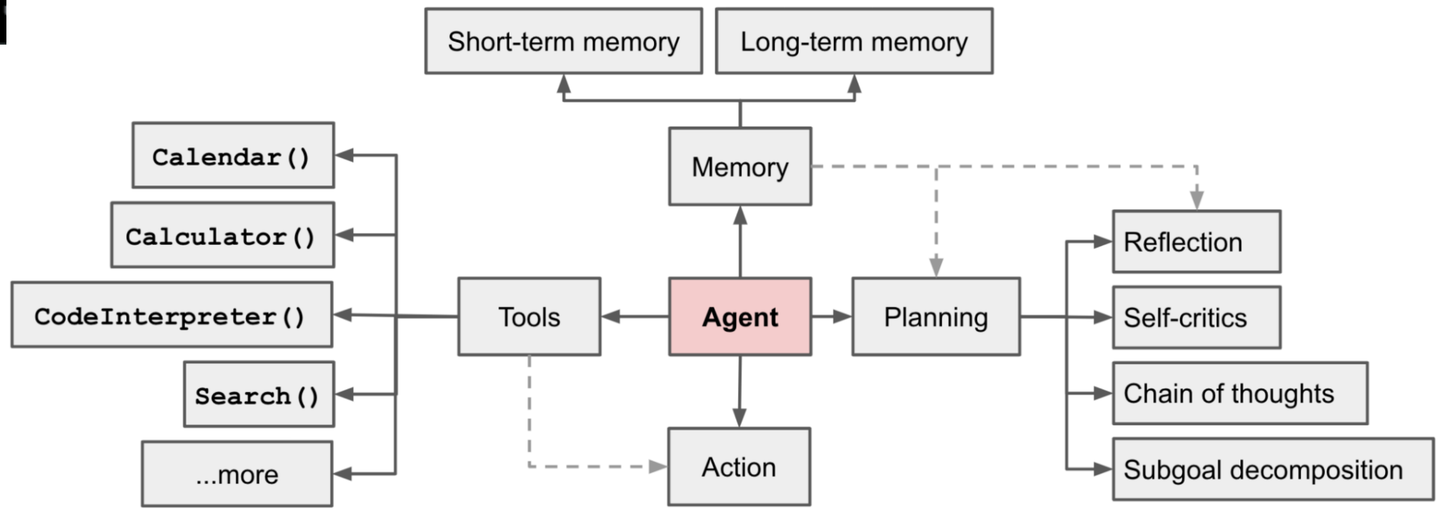
Limitations of Contextual Integrity

- Contexts framing are too rigid. May not be rich enough to capture the nuances. Too many factors in the transmission principle.
- Generalization. Unclear how the norms change when the contexts change?
- What if users do not want to protect themselves?

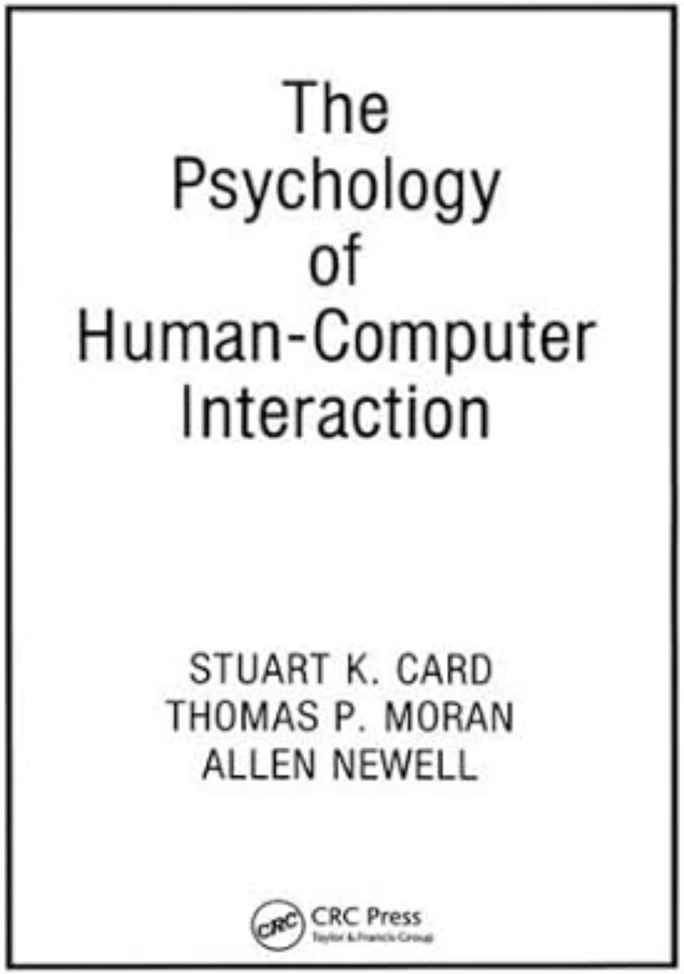
Privacy & Cognition

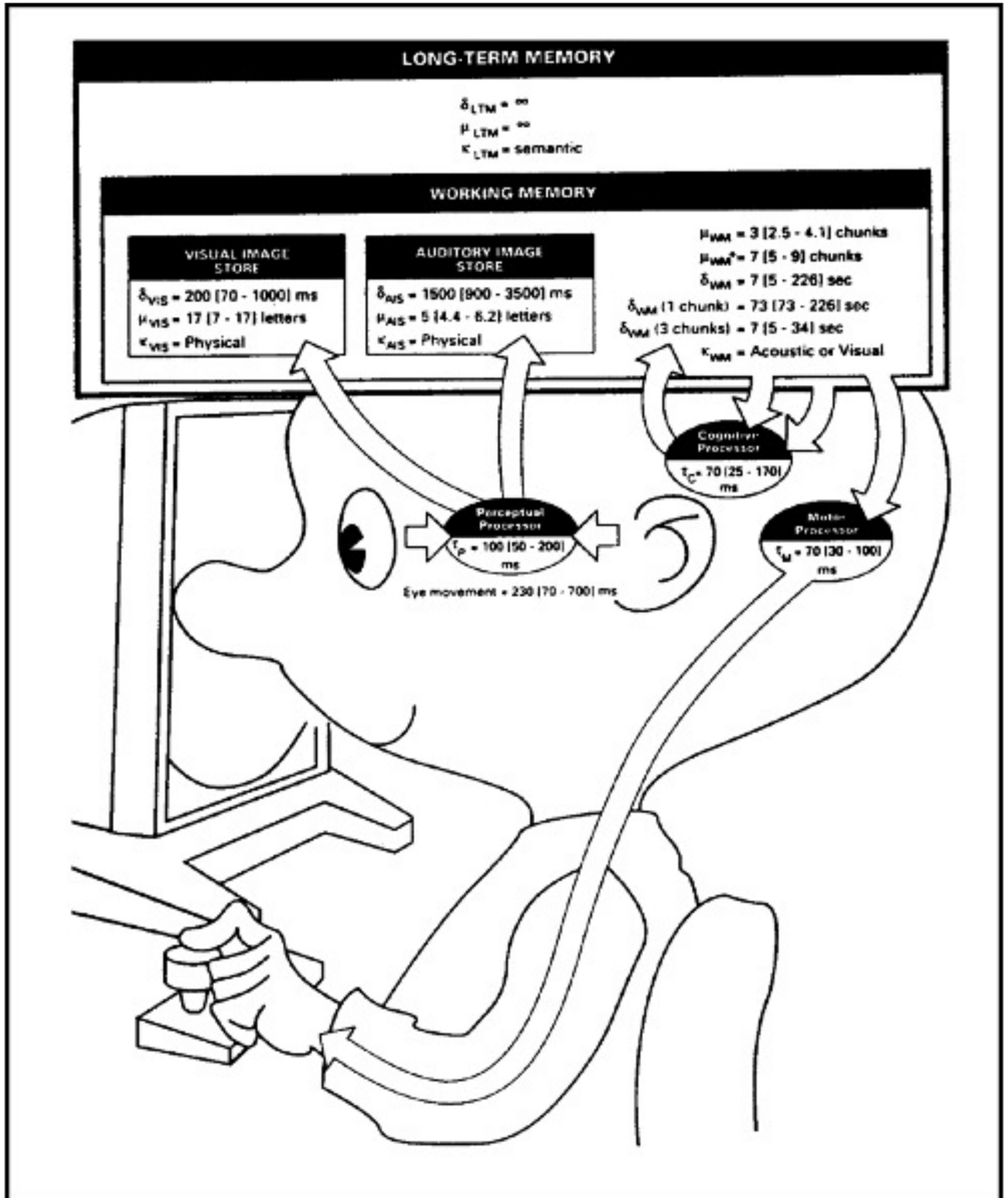
- The model human processor by Alan Newell
- Privacy Paradox
- Cognitive Biases
- A multi-layer model


Lilian Weng
@lilianweng
Agent = LLM + memory + planning skills + tool use
This is probably just a start of a new era :)
lilianweng.github.io
LLM Powered Autonomous Agents
Building agents with LLM (large language model) as its core controller is a cool concept. Several proof-of-concepts ...
8:35 PM · Jun 26, 2023 · 604.7K Views
103
862
3.6K
2.4K

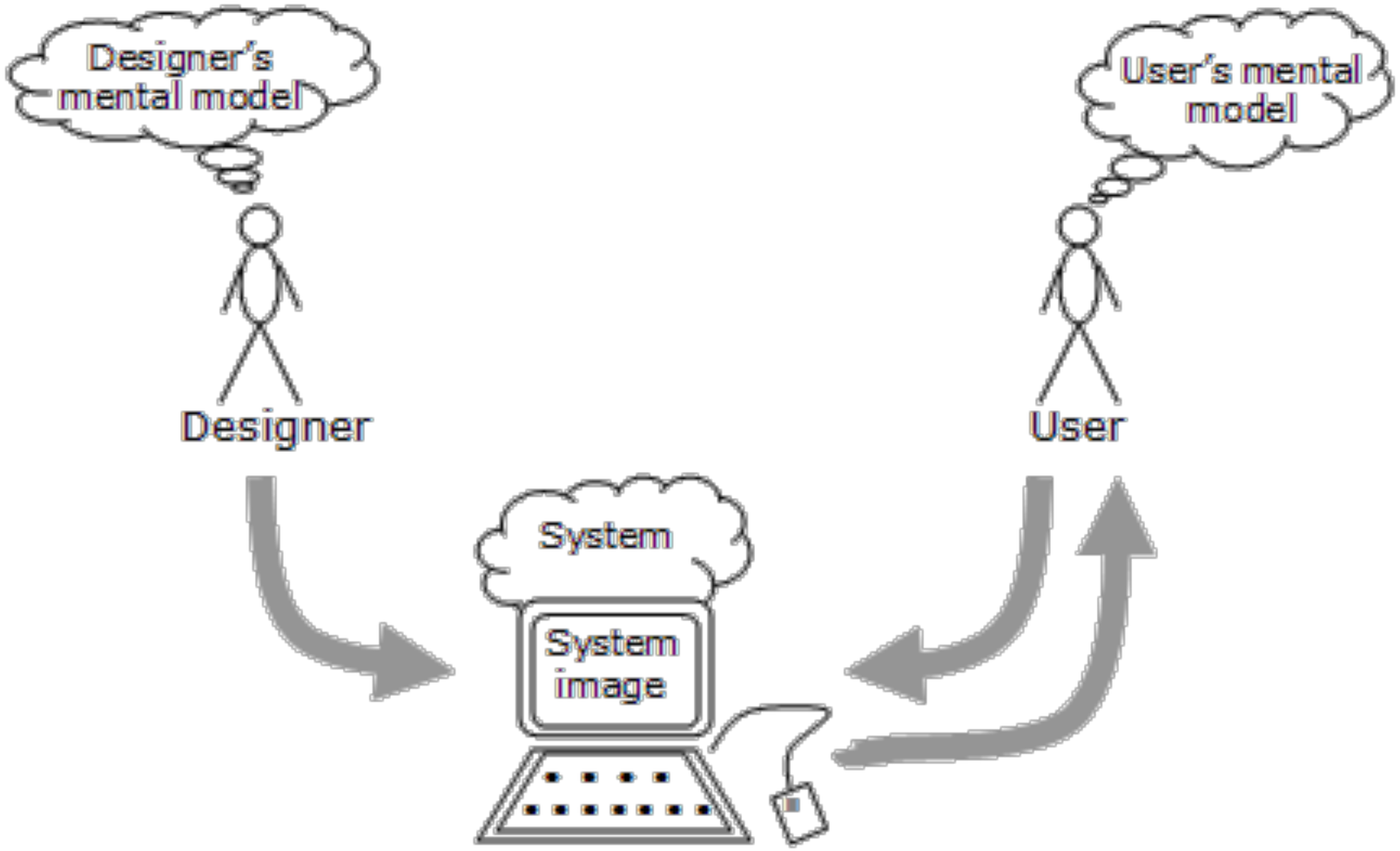

Short-term memory
Long-term memory
Memory
Calendar()
Calculator()
CodeInterpreter()
Search()
...more
Tools
Agent
Planning
Action
Reflection
Self-critics
Chain of thoughts
Subgoal decomposition

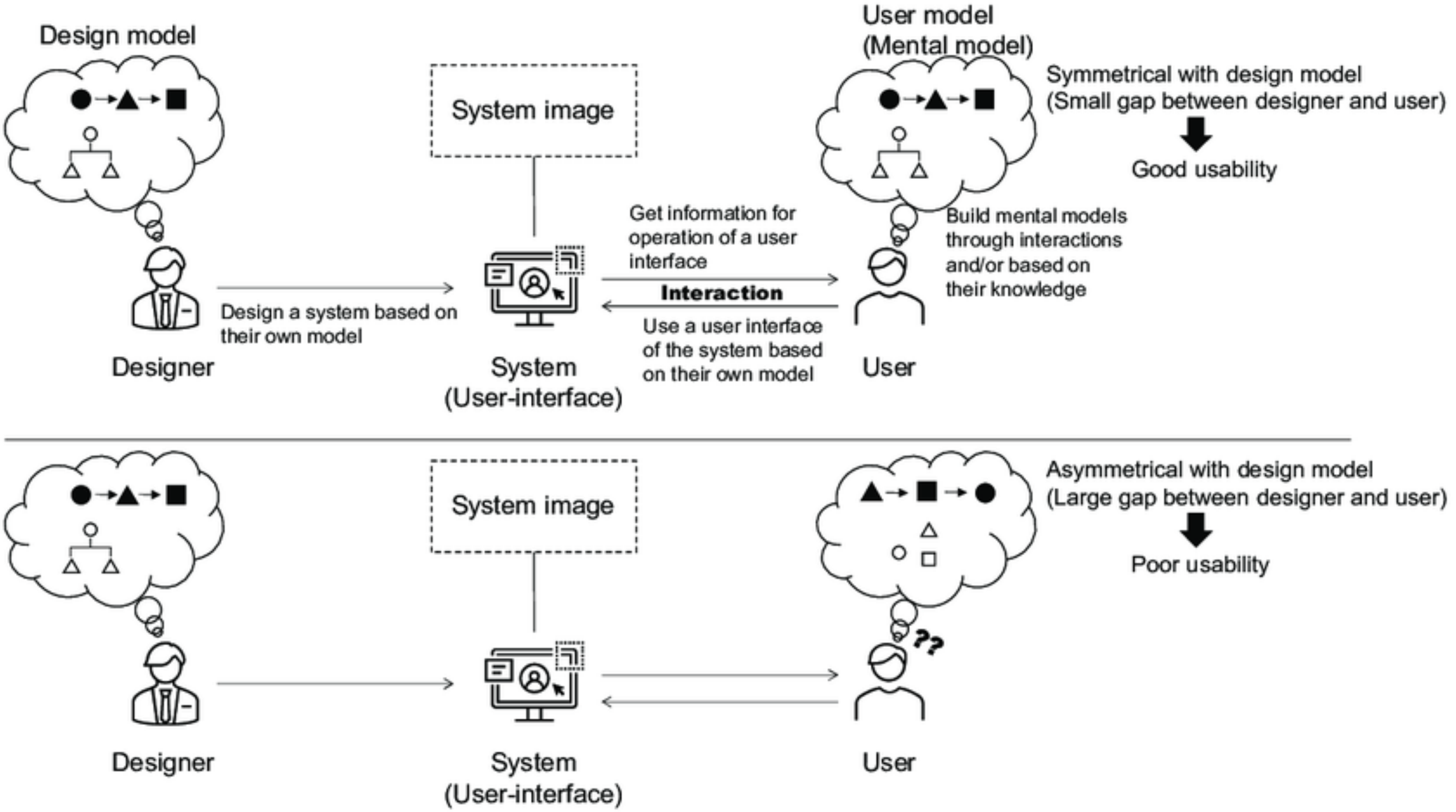
What is the nature of the mind?

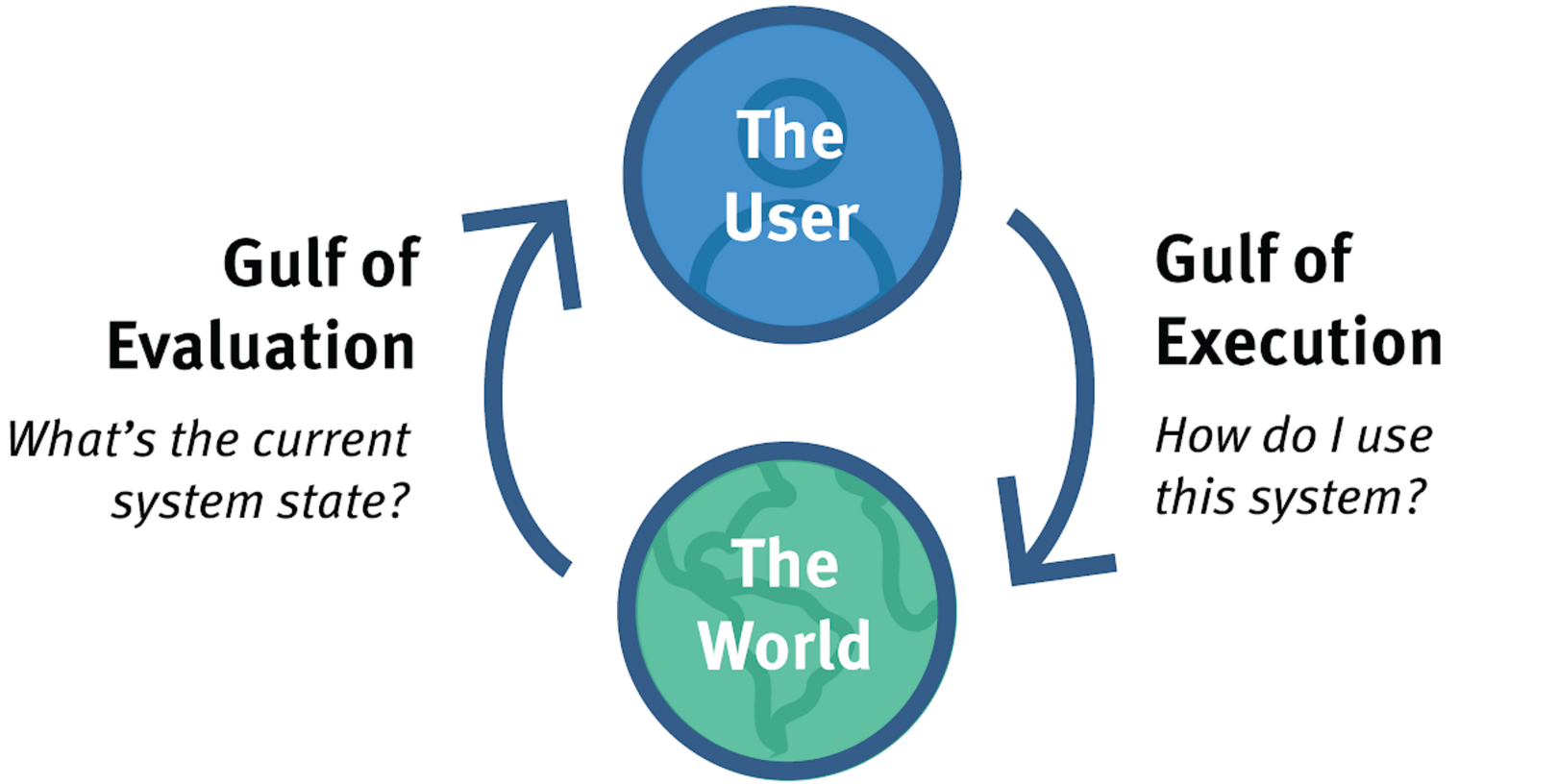




A task specific view - Mental Model & System Image

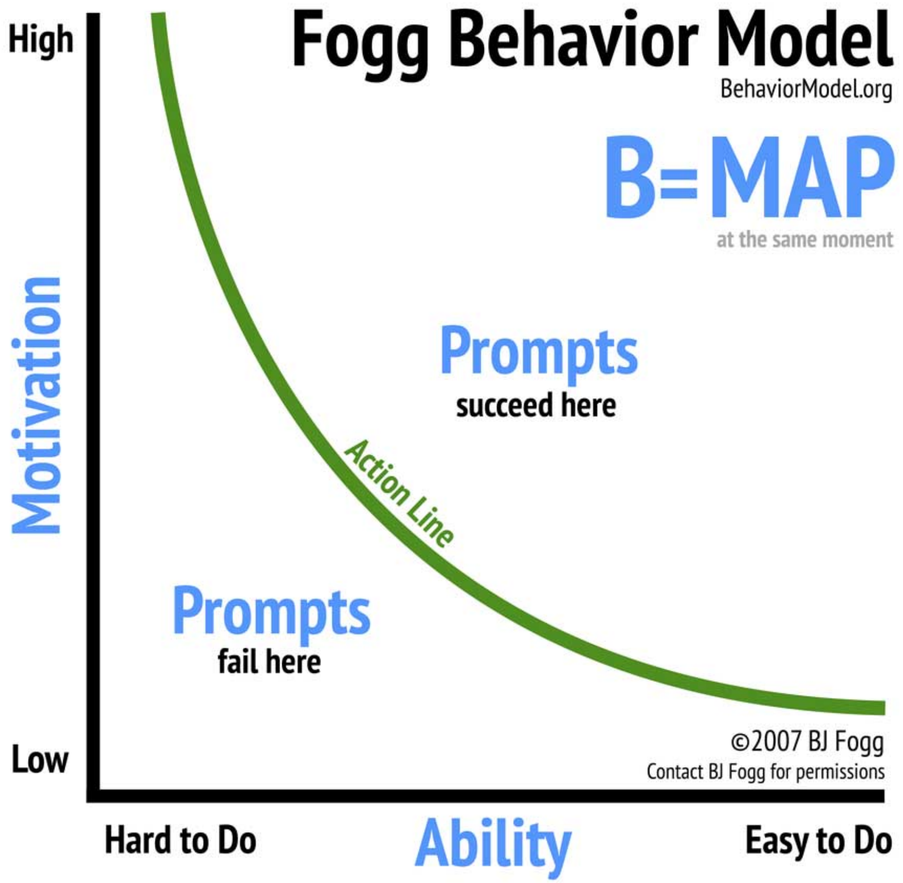



Design model
System image
User model
(Mental model)
Symmetrical with design model
(Small gap between designer and user)
Good usability
Get information for
operation of a user
interface
Interaction
Use a user interface
of the system based
on their own model
Build mental models
through interactions
and/or based on
their knowledge
Design a system based on
their own model
Designer
System
(User-interface)
User
Asymmetrical with design model
(Large gap between designer and user)
Poor usability
??
Designer
System
(User-interface)
User



nngroup.com NN/g

Fogg behavior model



8 Cognitive Biases of Human Beings

1. Anchoring bias.

People are **over-reliant** on the first piece of information they hear. In a salary negotiation, whoever makes the first offer establishes a range of reasonable possibilities in each person's mind.

2. Availability heuristic.

People **overestimate the importance** of information that is available to them. A person might argue that smoking is not unhealthy because they know someone who lived to 100 and smoked three packs a day.

3. Bandwagon effect.

The probability of one person adopting a belief increases based on the number of people who hold that belief. This is a powerful form of **groupthink** and is reason why meetings are often unproductive.

4. Blind-spot bias.

Failing to recognize your own cognitive biases is a bias in itself. People notice cognitive and motivational biases much more in others than in themselves.

5. Choice-supportive bias.

When you choose something, you tend to feel positive about it, even if that **choice has flaws**. Like how you think your dog is awesome — even if it bites people every once in a while.

6. Clustering illusion.

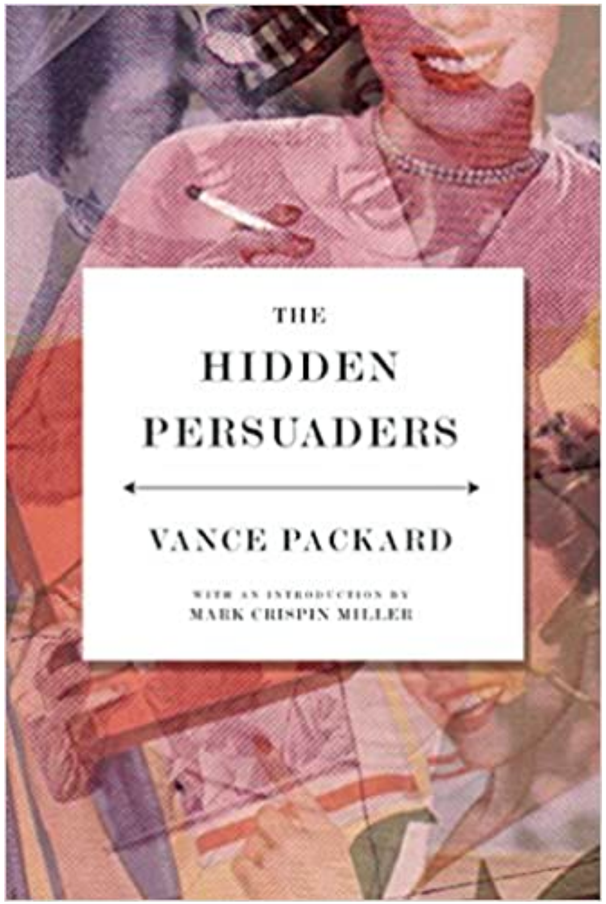
This is the tendency to **see patterns in random events**. It is key to various gambling fallacies, like the idea that red is more or less likely to turn up on a roulette table after a string of reds.

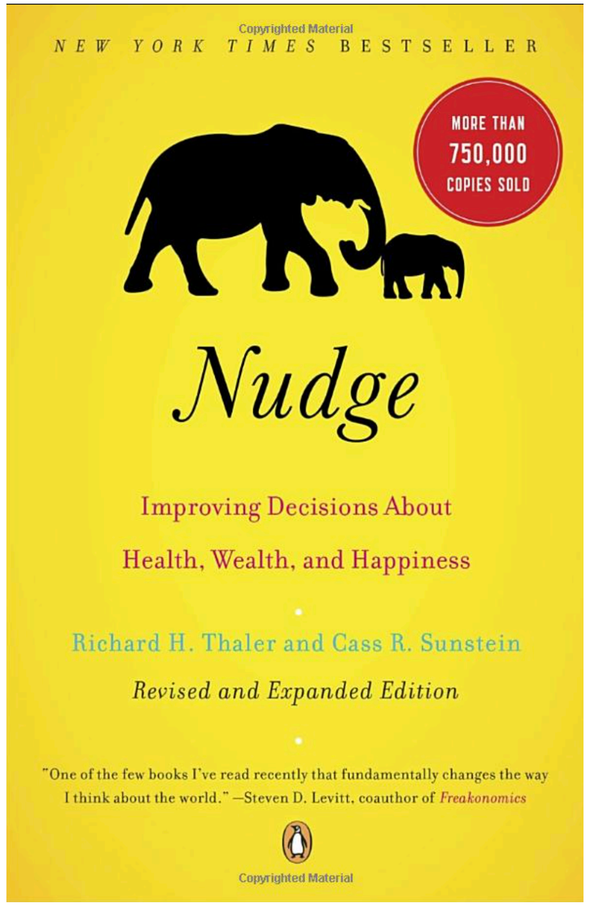
7. Confirmation bias.

We tend to listen only to information that confirms our **preconceptions** — one of the many reasons it's so hard to have an intelligent conversation about climate change.

8. Conservatism bias.

Where people favor prior evidence over new evidence or information that has emerged. People were **slow to accept** that the Earth was round because they maintained their earlier understanding that the planet was flat.

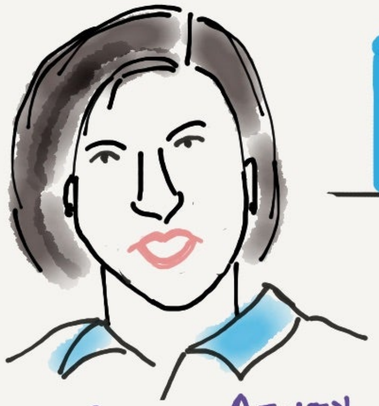

THE
HIDDEN
PERSUADERS
VANCE PACKARD
WITH AN INTRODUCTION BY
MARK CRISPIN MILLER


Copyrighted Material
NEW YORK TIMES BESTSELLER
MORE THAN
750,000
COPIES SOLD
Nudge
Improving Decisions About
Health, Wealth, and Happiness
Richard H. Thaler and Cass R. Sunstein
Revised and Expanded Edition
"One of the few books I've read recently that fundamentally changes the way I think about the world." —Steven D. Levitt, coauthor of Freakonomics
Copyrighted Material

Unpacking the Privacy Paradox

The privacy paradox is a dichotomy between **a person's intentions to protect their online privacy versus how they actually behave online** and, as a result, compromise their privacy.

PRIVACY PARADOX



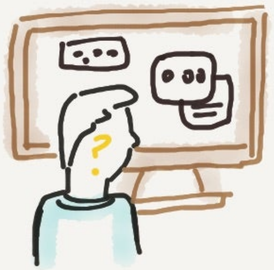
Dr. Susan Athey

STATED

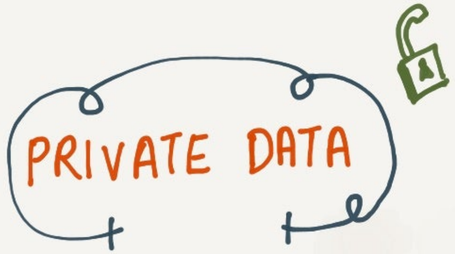
ACTUAL

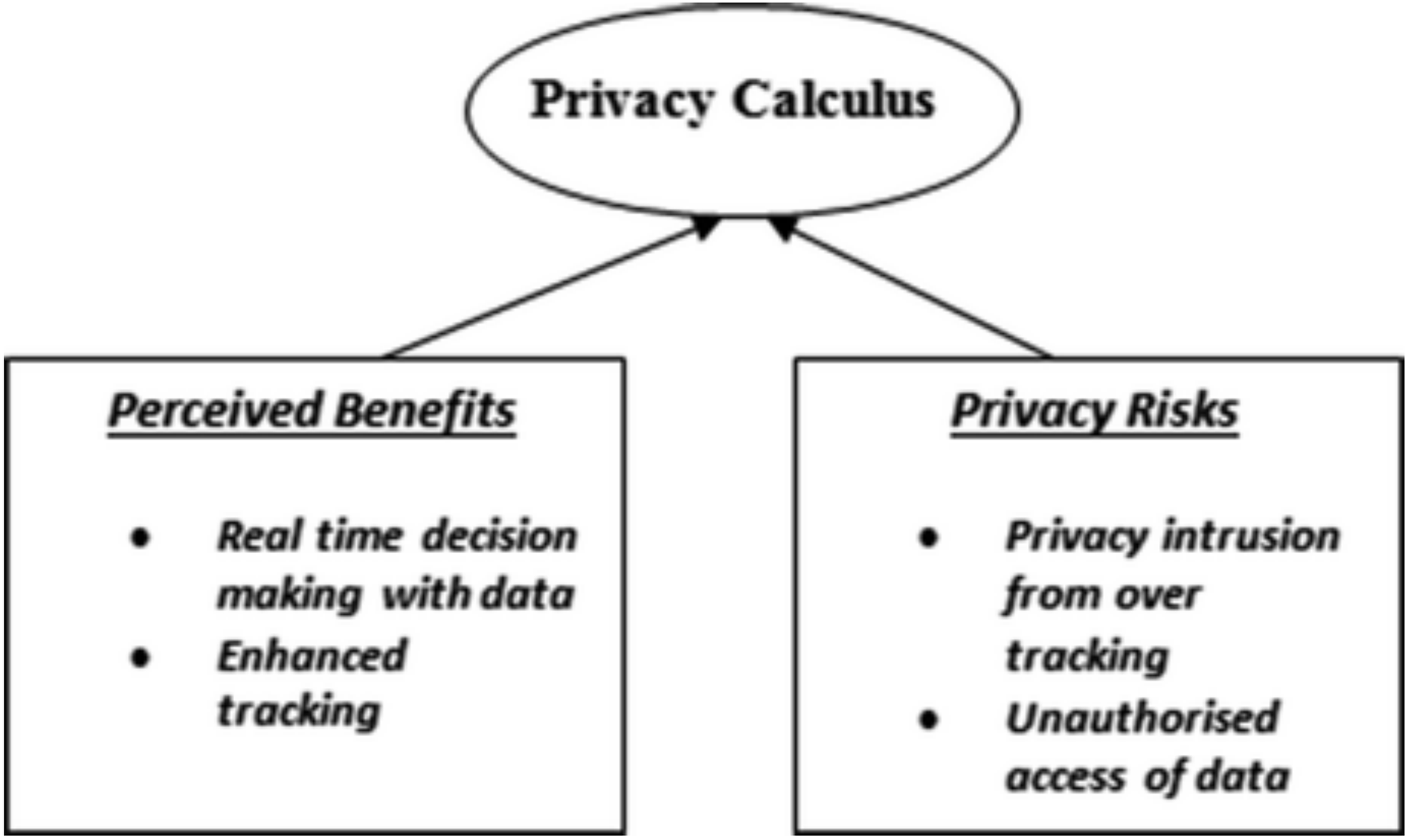
Incentive

Navigation

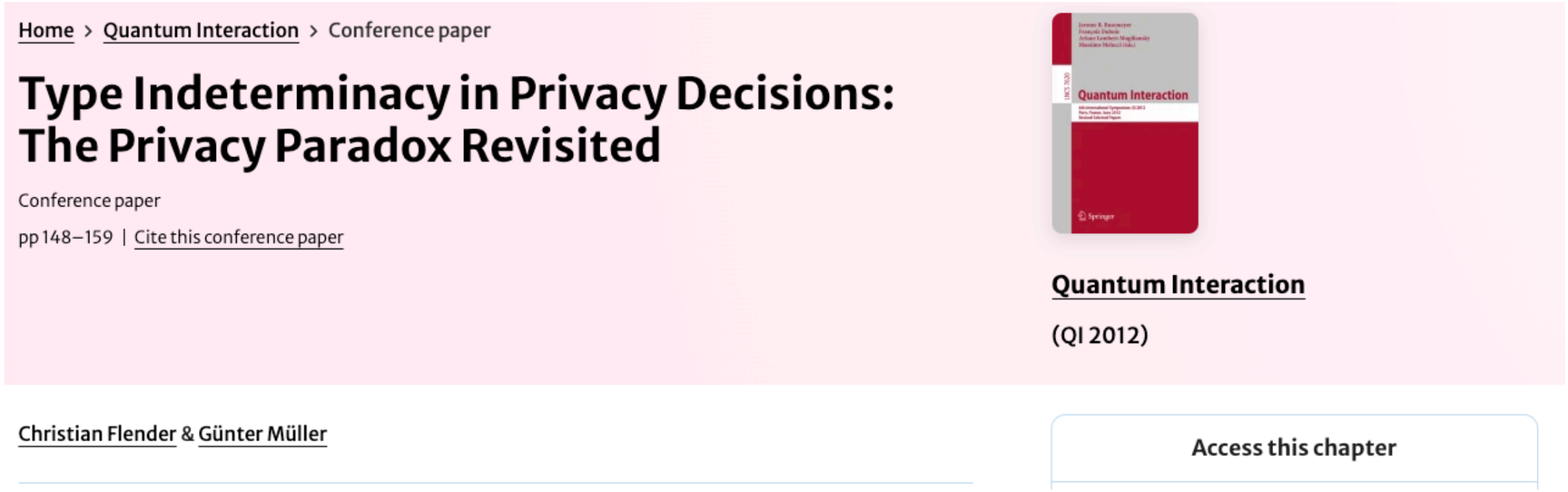


Encryption




Privacy Calculus
Perceived Benefits
Real time decision making with data
Enhanced tracking
Privacy Risks
Privacy intrusion from over tracking
Unauthorised access of data

Quantum Privacy???



a privacy decision's outcome is not settled until the moment the decision is actually made, and two distinct decisions cannot be deemed interchangeable within the context of decision-making

The Myth of the Privacy Paradox

*Daniel J. Solove**

Abstract

In this Article, Professor Daniel Solove deconstructs and critiques the privacy paradox and the arguments made about it. The "privacy paradox" is the phenomenon where people say that they value privacy highly, yet in their behavior relinquish their personal data for very little in exchange or fail to use measures to protect their privacy.

Commentators typically make one of two types of arguments about the privacy paradox. On one side, the "behavior valuation argument" contends behavior is the best metric to evaluate how people actually value privacy. Behavior reveals that people ascribe a low value to privacy or readily trade it away for goods or services. The argument often goes on to contend that privacy regulation should be reduced.

On the other side, the "behavior distortion argument" suggests that people's behavior is not an accurate metric of preferences because behavior is distorted by biases and heuristics, manipulation and skewing, and other factors.

Professor Solove argues instead that the privacy paradox is a myth created by faulty logic. The behavior involved in privacy paradox studies involves people making decisions about risk in very specific contexts. In

Why?

- Biases and Heuristics
- Framing Effects
- Behavioral Manipulation and Skewing
- Misunderstandings and Lack of Knowledge
- Inertia and Friction


copyright (c) 1999 Daniel J. Simons. All rights reserved.


elRellano.com

Framing Effects

Facebook reveals news feed experiment to control emotions

Protests over secret study involving 689,000 users in which friends' postings were moved to influence moods

Poll: Facebook's secret mood experiment: have you lost trust in the social network?

Activists and politicians called Facebook's experiment 'scandalous', 'spooky' and 'disturbing'. Photograph: Dado Ruvic/Reuters

It already knows whether you are single or dating, the first school you went to and whether you like or loathe Justin Bieber. But now Facebook, the world's biggest social networking site, is facing a storm of protest after it revealed it had discovered how to make users feel happier or sadder with a few computer key strokes.

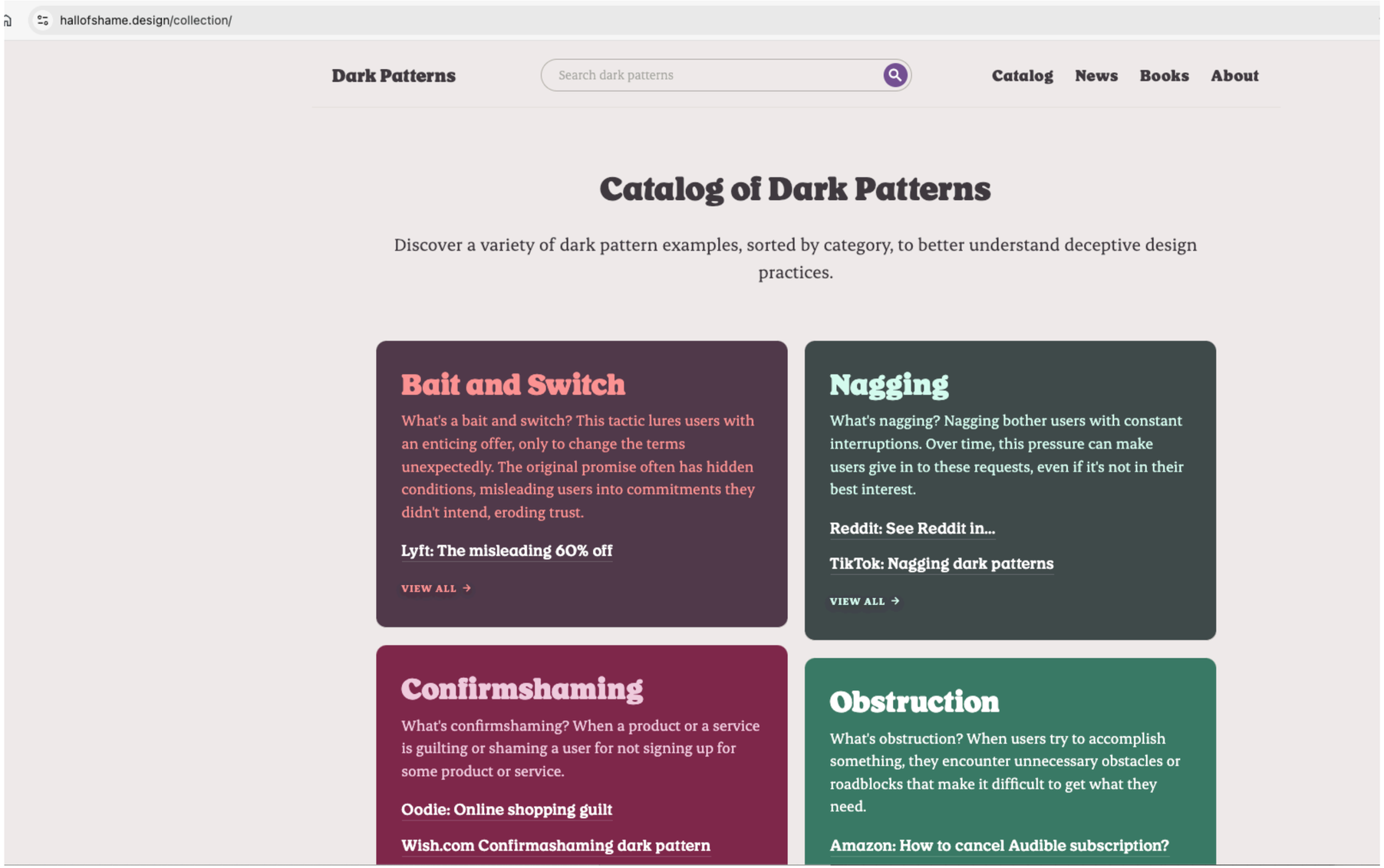
Researchers at Facebook want to study whether users are more likely to share positive (happy) thoughts if their friends have been posting positive thoughts, and whether they are more likely to share negative (unhappy) thoughts if their friends have been sharing negative thoughts.

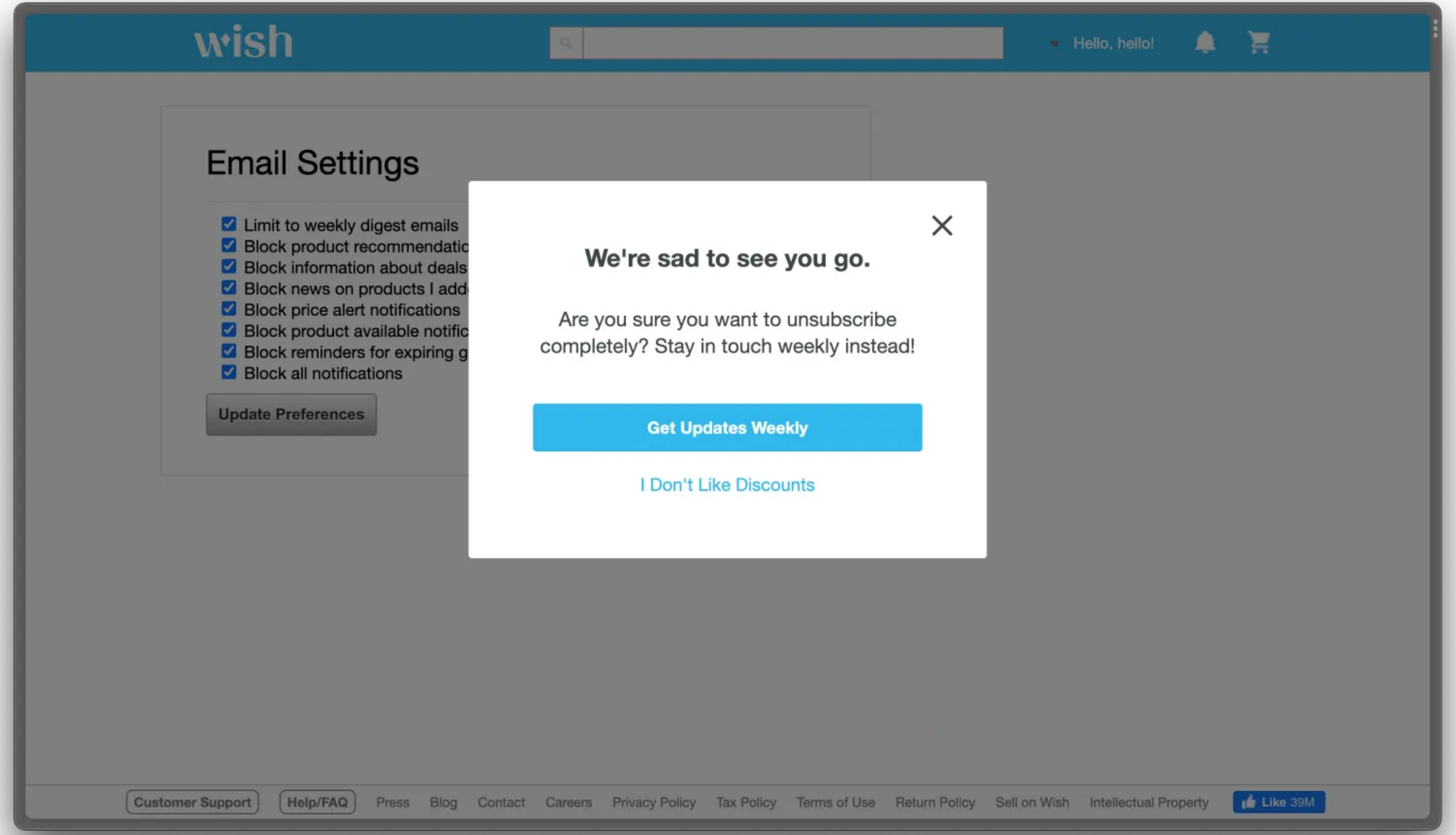
- To increase the proportion of positive posts in some users' news feeds, the researchers will randomly exclude some fraction of friends' negative posts each time the news feed is loaded.
- To increase the proportion of negative posts in some users' news feeds, the researchers will randomly exclude some fraction of friends' positive posts each time the news feed is loaded.
- The researchers will use an automated algorithm to measure whether users' posts are of a positive or negative mood.
- The researchers will publish the anonymized aggregate results of the experiment in a scientific paper.
- Participants will not be identified and will remain anonymous.

If the researchers are not allowed to perform this experiment, they will not be able to make a valid scientific determination of whether users' moods are affected by the moods of their friends' posts. Therefore, the researchers will not be able to produce features that might protect the moods of psychologically-vulnerable users.

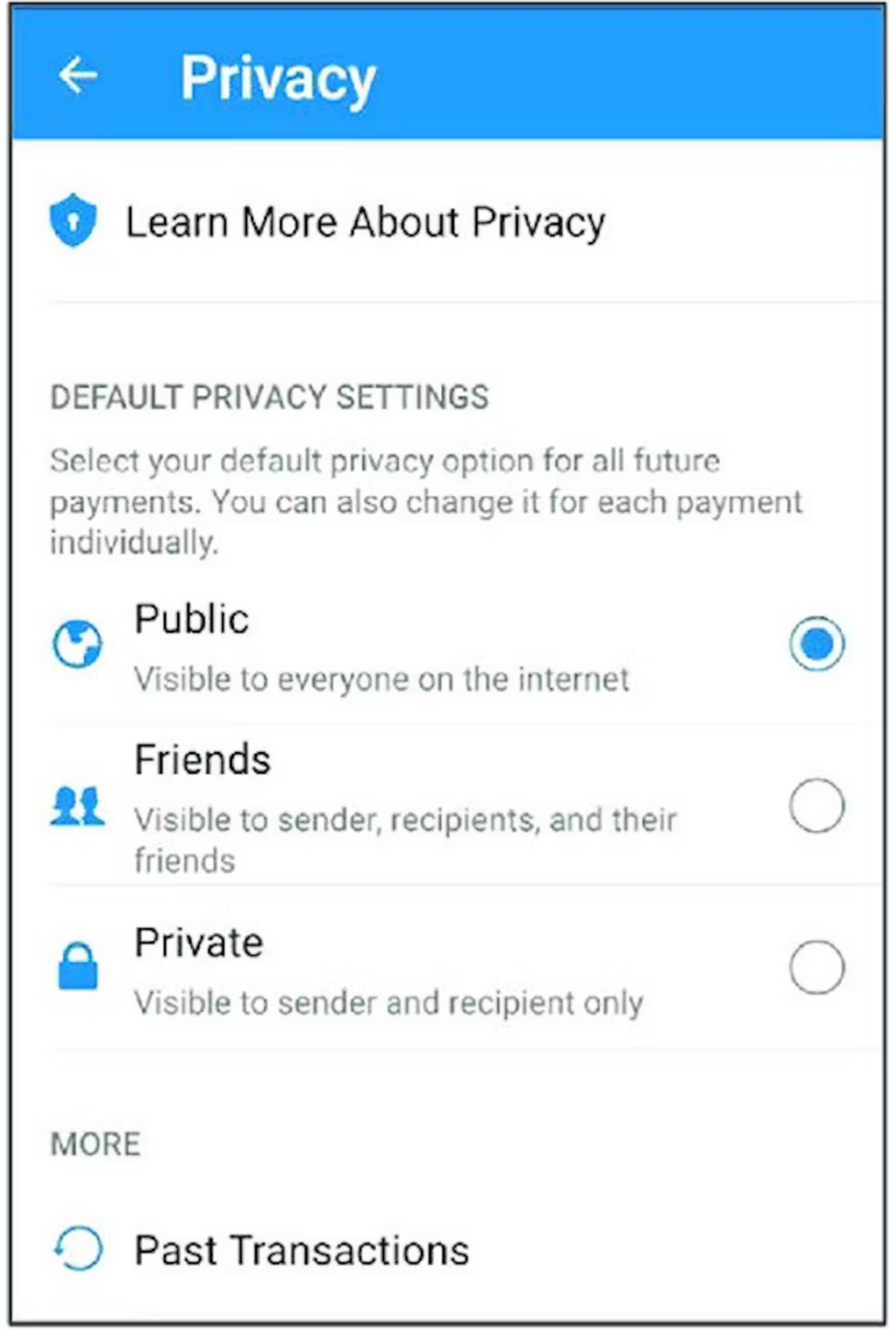
Chrome plugin lets users experience Facebook's 'emotion contagion' experiment

Dark patterns




wish
Hello, hello!
Email Settings
Limit to weekly digest emails
Block product recommendatic
Block information about deals
Block news on products I add
Block price alert notifications
Block product available notific
Block reminders for expiring g
Block all notifications
Update Preferences
We're sad to see you go.
Are you sure you want to unsubscribe completely? Stay in touch weekly instead!
Get Updates Weekly
I Don't Like Discounts
Customer Support
Help/FAQ
Press
Blog
Contact
Careers
Privacy Policy
Tax Policy
Terms of Use
Return Policy
Sell on Wish
Intellectual Property
Like 39M

Default Privacy Settings



State of the Art Research

- Alan Westin. Privacy Index.
- Clustering Privacy Preferences
- Contextual Integrity for Privacy Preferences Predictions
- Contextual Labels
- Precision Privacy

(1) *Consumers have lost all control over how personal information is collected and used by companies.*
(2) *Most businesses handle the personal information they collect about consumers in a proper and confidential way.*
(3) *Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.*

Westin privacy category	Percentage of participants
Privacy unconcerned	11.6%
Privacy pragmatists	55.9%
Privacy fundamentalists	32.5%

Would a privacy fundamentalist sell their DNA for $1000... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences

Allison Woodruff
Google
1600 Amphitheatre Pkwy
Mountain View, CA 94043
woodruff@acm.org

Vasyl Pihur
Google
1600 Amphitheatre Pkwy
Mountain View, CA 94043
vpihur@google.com

Sunny Consolvo
Google
1600 Amphitheatre Pkwy
Mountain View, CA 94043
sconsolvo@google.com

Lauren Schmidt
Google
1600 Amphitheatre Pkwy
Mountain View, CA 94043
schmidtl@google.com

Laura Brandimarte
Carnegie Mellon University
5000 Forbes Av. HBH 2105C
Pittsburgh, PA 15213
lbrandim@andrew.cmu.edu

Alessandro Acquisti
Carnegie Mellon University
5000 Forbes Av. HBH 2105C
Pittsburgh, PA 15213
acquisti@andrew.cmu.edu

ABSTRACT

Westin's Privacy Segmentation Index has been widely used to measure privacy attitudes and categorize individuals into three privacy groups: fundamentalists, pragmatists, and unconcerned. Previous research has failed to establish a robust correlation between the Westin categories and actual or intended behaviors. Unexplored however is the connection between the Westin categories and individuals' responses to the *consequences* of privacy behaviors. We use a survey of 884 Amazon Mechanical Turk participants to investigate the relationship between the Westin Privacy Segmentation Index and attitudes and behavioral intentions for both privacy-sensitive scenarios and privacy-sensitive consequences. Our results indicate a lack of correlation between the Westin categories and behavioral intent, as well as a lack of correlation between the Westin categories and consequences. We discuss potential implications of this attitude-consequence gap.

Nonetheless, concerns have long existed regarding the predictive power of Westin's categories and the assumptions underlying his Privacy Segmentation Index. First, previous research has failed to establish a significant correlation between the Westin categories (which capture broad, generic privacy attitudes) and context-specific, privacy-related behaviors, either actual or intended [13, 23, 29]. Second, researchers have raised concerns regarding unstated assumptions underlying the index, which presumes individuals make privacy decisions that are highly rational, reflective, and informed [42]. Instead, scholars have posited that incomplete information or decision-making biases, among other factors, may cause a gap between the general attitudes captured by the Westin categories and actual, specific privacy behavior [4]. Third, the instrument has not been updated since approximately 1995, and it is not obvious that it remains current in our Internet-centric world.

It is perhaps unsurprising that generic attitudes (such as

Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings

Jialiu Lin Bin Liu Norman Sadeh Jason I. Hong
School of Computer Science, Carnegie Mellon University
{jialiul, bliu1, sadeh, jasonh}@cs.cmu.edu

ABSTRACT

In this paper, we investigate the feasibility of identifying a small set of privacy profiles as a way of helping users manage their mobile app privacy preferences. Our analysis does not limit itself to looking at permissions people feel comfortable granting to an app. Instead it relies on static code analysis to determine the purpose for which an app requests each of its permissions, distinguishing for instance between apps relying on particular permissions to deliver their core functionality and apps requesting these permissions to share information with advertising networks or social networks. Using privacy preferences that reflect people's comfort with the purpose for which different apps request their permissions, we use clustering techniques to identify privacy profiles. A major contribution of this work is to show that, while people's mobile app privacy preferences are diverse, it is possible to identify a small number of privacy profiles that collectively do a good job at capturing these diverse preferences.

This paper investigates the feasibility of organizing end-users into a small set of clusters and of identifying default privacy profiles for each such cluster as a way of both simplifying and enhancing mobile app privacy. We use data obtained through static code analysis and crowdsourcing, and analyze it using machine learning techniques to highlight the limitations of today's interfaces as well as opportunities for significantly improving them. Specifically, our results were obtained by collecting 21,657 preference ratings from 725 users on 837 free Android apps. These preference ratings were collected on over 1200 app-permission-purpose triples. Each such preference rating captures a user's willingness to grant a given permission to a given app for a particular purpose. Identification of the purpose(s) associated with a given app's permission was inferred using static code analysis, while distinguishing between different types of 3rd-party libraries responsible for requesting access to a given permission. For example, if location data is used by an app only because of an ad library bundled with the app, we can infer that location is used

Privacy Expectations and Preferences in an IoT World

Pardis Emami-Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer,
Lorrie Faith Cranor, Norman Sadeh
Carnegie Mellon University
Pittsburgh, PA, USA
{pardis, srutib, lbauer, lorrie}@cmu.edu
{htq, degeling, sadeh}@cs.cmu.edu

ABSTRACT

With the rapid deployment of Internet of Things (IoT) technologies and the variety of ways in which IoT-connected sensors collect and use personal data, there is a need for transparency, control, and new tools to ensure that individual privacy requirements are met. To develop these tools, it is important to better understand how people feel about the privacy implications of IoT and the situations in which they prefer to be notified about data collection. We report on a 1,007-participant vignette study focusing on privacy expectations and preferences as they pertain to a set of 380 IoT data collection and use scenarios. Participants were presented with 14 scenarios that varied across eight categorical factors, including the type of data collected (e.g. location, biometrics, temperature), how the data is used (e.g., whether it is shared, and for what purpose), and other attributes such as the data retention period. Our findings show that privacy preferences are diverse and context dependent; participants were more comfortable with data being collected in public settings rather than in private places, and are more likely to consent to data being collected for uses they find beneficial. They are less

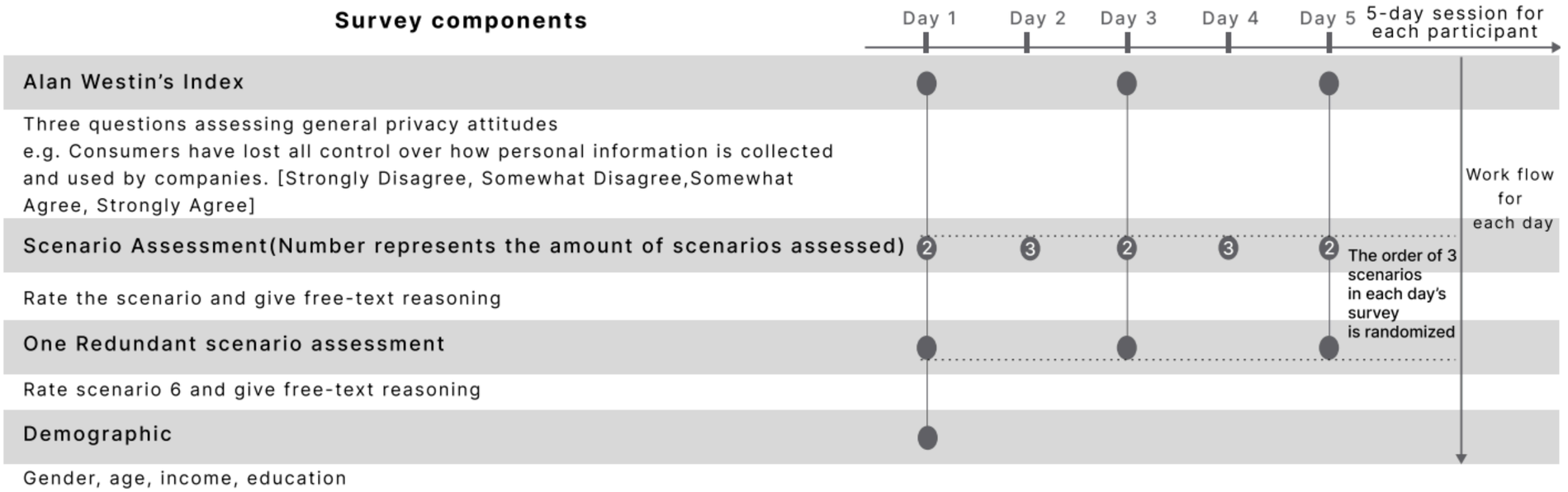
practices and offer privacy choices that respect individual privacy preferences. Gaining traction on this problem requires nuanced understanding of societal norms and context, as well as individual needs [31, 35]. For example, most people tacitly accept being recorded on cameras and CCTV outdoors in public spaces, but express disdain for installing video surveillance systems inside the walls of their homes. As more complex IoT scenarios become possible, many other factors may play a role in determining individuals' privacy preferences. While some may feel comfortable with their location being tracked for the purpose of traffic prediction, they may consent to tracking only their work commute. Others may consent only if they are assured that their location data is retained and used in an anonymized form.

We conducted a large-scale online vignette study to identify the contribution of different factors (such as the type of data, retention time, purpose of data collection, and location of data collection) in promoting or inhibiting individuals' self-professed comfort levels. We also studied the factors that trigger a desire for notifications about data collection. Our research identified which aspects of data

Context Label

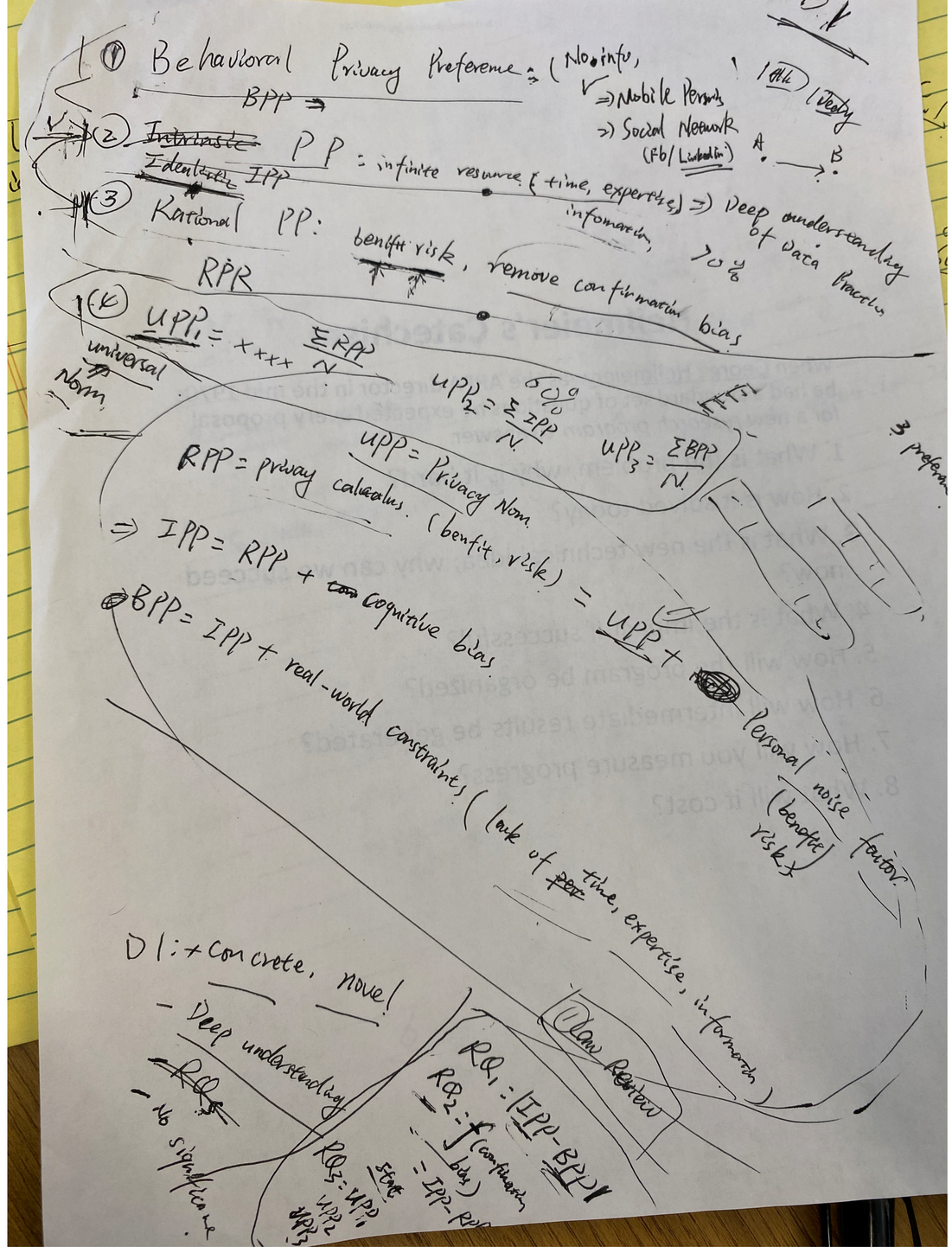
- Prove Rationality
- Context labels v.s. Categorical factors
- Longitudinal studies
- Focus on opinions

predict users' privacy concerns with an accuracy (73%) surpassing Privacy Segmentation Index (56%) and methods using categorical factors (59%).



Precision Privacy

- Behavioral Privacy Preferences (BPP)
- Idealistic Privacy Preferences (IPP)
- Rational Privacy Preferences (RPP)
- Universal Privacy Preferences (UPP)



Precision Privacy

- (Individual) Behavioral Privacy Preferences (BPP)
- (Individual) Idealistic Privacy Preferences (IPP)
- (Individual) Rational Privacy Preferences (RPP)
- (Group) Universal Privacy Preferences (UPP)

- RPP = Privacy Calculus (Benefits, Risks) = UPP + individual Risk/Benefit variances
- IPP = PP with infinite resources (time, expertise, information) = RPP + Cognitive Bias
- BPP = IPP + Real-world constraints

Questions

- Which Privacy Preferences should we respect?
  - Privacy norms or individual PP?
  - How to compute norms?
    - Universal Privacy Preferences (UPP) = Average of
      - Idealistic Privacy Preferences (IPP)?
      - Rational Privacy Preferences (RPP)?

What's the new technical idea? Why can we succeed now?

- Smaller problem scope. Focus on the mental preferences. No behaviors.
- New technology. Lean Privacy Review, LLM, Data Science.
- A CogSci inspired approach.