DEPARTMENT OF COMMERCE
National Telecommunications and Information Administration
Docket no. 180821780-8780*91
Developing the Administration's Approach to Consumer Privacy
Request for Comment

Presented by:
Dr. Yan Shvartzshnaider
Faculty Fellow, New York University
Visiting Associate Research Scholar Center for Information Technology Policy, Princeton University

The Department of Commerce has requested input on its proposed approach to consumer privacy. As a member of an academic group working on technologies and privacy frameworks at the intersection of privacy, technology policy, and the integrity of social life, I submit these comments on my own behalf.

In NTIA's Request for Comment, the Department asks for feedback on whether it has identified the "core privacy outcomes that consumers can expect" from the businesses and organizations that collect and further use and repackage their personal data. In this comment, I will outline additional privacy outcomes which I believe that NTIA and the Administration should consider. In addition, at the end, I provide supporting papers, together with an offer to meet with NTIA to fully explore the privacy outcomes we recommend.

Overview:

The Privacy Outcomes which NTIA outlined in this Request for Comment are important. However, to foster real change, companies and organizations must take a giant step further. Specifically, in addition to current outcomes --- *transparency* for how companies handle users' information; *controls* to restrict data collection; employing *security* measures, etc --- companies should adopt a socially meaningful conception of privacy, one that meets people's expectations, and is ethically and legally legitimate. This approach entails a) adopting the definition of privacy as contextual integrity b) discovering contextual norms and relevant consumer's (user's) privacy expectations and c) devising mechanisms that ensure that the (user's) privacy expectations are respected. Further, NTIA should add three additional privacy outcomes as critical to any recommendations for approaches to consumer privacy that NTIA develops:

A. Additional Privacy Outcome 1: Adopting Privacy as Contextual Integrity

In order to achieve ethical and legitimate privacy outcomes, companies must recognize and understand the norms of the social domains in which they operate. The theory of Contextual Integrity serves as an important framework in achieving this and informing the rest of the privacy outcomes that NTIA has published. Formally, contextual Integrity (CI) defines privacy as an appropriate information flow, where appropriateness, in turn, is defined as conformance with legitimate, informational norms. *Protecting privacy becomes about ensuring the appropriateness of informational flows. To achieve that, companies and organization must first discover contextual norms, identify users' privacy expectations and then ensure that they are respected by their systems.*

B. Additional Privacy Outcome 2: Discovery of Consumer's (user's) Privacy Expectations

Privacy for future "smart homes" in the Internet of Things, for example, must depend not only on privacy choices a company may choose to provide, but on the expectations that homeowners, as a group, bring: the expectations of privacy that users bring to their homes and their relationships with future "smart home" platform and service providers.

Homeowners, as all people, learn and adopt implicit and explicit privacy norms from their families, friends, and communities. Consequently, companies who create platforms and services as intricately involved in our personal and familial lives as "smart homes" must be mindful of relevant societal norms. Law and regulation, handbooks, and current "best practices" may lag behind the current state of privacy expectations of the majority of people, consumers, and users.

Companies must not merely dictate privacy control; rather they must study, research, explore and understand their customer's privacy expectations and the social norms in which they are introducing new technologies into people's lives, family interactions and homes.

C. Additional Privacy Outcome 3: Respecting Consumer's (User's) Privacy Expectations

After the company has taken steps to understand its customers' privacy expectations, it must make substantial efforts to respect them. Nothing could be more fundamental to a modern society than to respect the collective and common privacy expectations of its consumers and communities. Yet, today this often does not take place and privacy is a casualty of complicated information flows between platform providers and their users. Many of the information systems upon which we rely, and which are deeply embedded in the fabric of our daily life, do not incorporate the necessary mechanisms to ensure their handling of our personal data is consistent with our privacy norms and expectations.

Accordingly, to have a socially meaningful and legally legitimate concept of privacy, companies must create privacy policies and privacy systems which respect the privacy norms they learn from their customers and communities. Companies must incorporate personal data processing collection, processing, use, storage and disclosure mechanisms which respect community privacy norms, societal expectations and behavior.

This is a critical outcome which must be added to any principle-based approach to privacy which NTIA may write and the Administration might pursue.

D. Considerable work has been done on the three additional "privacy outcome" principles proposed above.

As requested by the Request for Comment, below I included links to the papers that describe our ongoing work on devising ways to discover norms and societal expectations of privacy, and mechanisms that ensure that these norms and expectations are respected in the development of complex technical systems.

Finally, attached is also a copy of the final report that summarizes the discussions in the recent symposium on application of contextual integrity that took place in Princeton University in September (http://privaci.info/ci_symposium/program.html) The symposium brought together academic scholars and industry practitioners to discuss a wide range of mature and ongoing work on applications of contextual integrity. The topics included discussions on an application of CI to build privacy preserving smart homes and IoT systems, improving Human-Computer Interaction (HCI), discovering users' privacy expectation and other societal privacy issues involving new technologies.

Papers:

Analyzing Privacy Policies Using Contextual Integrity Annotations
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=324487
Yan Shvartzshnaider, Noah Apthorpe, Nick Feamster (Princeton University), Helen Nissenbaum, SSRN Preprint. 2018

Discovering Smart Home Internet of Things Privacy Norms Using Contextual Integrity
https://export.arxiv.org/pdf/1805.06031
Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, Nick Feamster in Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT/UbiComp), 2018

Learning Privacy Expectations by Crowdsourcing Contextual Informational Norms
https://privaci.github.io/papers/hcomp_paper.pdf
Yan Shvartzshnaider, Schrasing Tong, Thomas Wies, Paula Klift, Helen Nissenbaum, Lakshminarayanan Subramanian, Prateek Mittal in Proceedings of the Fourth AAAI Conference on Human Computation and Crowdsourcing (HCOMP), 2016

"It Takes a Village": A Community Based Participatory Framework for Privacy Design
https://networkedprivacy2018.files.wordpress.com/2018/04/mir.pdf
Darakhshan Mir, Yan Shvartzshnaider and Mark Latonero in Proceedings of the International Workshop on Privacy Engineering (IWPE), 2018

Conclusion:

NTIA has embarked on an important inquiry, and I am pleased to submit these comments. I look forward to continuing the conversation with NTIA, and would like to participate in future meetings and discussion that will take place. Thank you for this opportunity to comment.

Dr. Yan Shvartzshnaider
Faculty Fellow, New York University
Visiting Associate Research Scholar Center for Information Technology Policy, Princeton University

SYMPOSIUM ON APPLICATIONS OF CONTEXTUAL INTEGRITY

September 13-14, 2018

Princeton University

REPORT

Compiled by Noah Apthorpe

WITH CONTRIBUTIONS BY

Karla Badillo-Urquiola

Erica Du

Jaulie Goe

Priya Kumar

Marijn Sax

HOSTED BY

Princeton University Center for Information Technology Policy

Cornell Tech Digital Life Initiative

SUPPORTED BY

A Gift from the Microsoft Foundation

Contents

Executive Summary

Noah Apthorpe

The Princeton University Center for Information Technology (CITP) and the Cornell Tech Digital Life Initiative (DLI) hosted the **Symposium on Applications of Contextual Integrity** on September 13-14, 2018. The event brought together faculty, postdoctoral researchers, graduate students, undergraduates, and industry partners to present research using the theory of contextual integrity as a framework to reason about, design and evaluate, craft regulation for, and generate formal logics for privacy.

Contextual integrity (CI) was first proposed by Helen Nissenbaum in 2004 as a new framework for reasoning about privacy. CI focuses on societal norms which govern the appropriateness of information flows in defined contexts. Information flows within a context that do not abide by existing norms are perceived as privacy violations. For example, it may appropriate for a smartphone users' location information to be sent to a website in order to provide recommendations for nearby restaurants. However, the specific details of this context are essential. It may be unacceptable for the same location information to be transferred to the same website for a different purpose, such as advertising.

Researchers in disciplines ranging from law to computer science to sociology have adopted and applied the theory of contextual integrity to their work. Presentations at the Symposium included research applying CI to smart homes and IoT, human-computer interaction, system design, discovering users' privacy opinions, and many social issues involving privacy and technology.

CI and Society

These sessions included presentations about seven projects using CI to investigate societal issues. The first presentation proposed combining CI with the Governing Knowledge Commons Framework with examples from examples from online social movements. The second project predicted the privacy behavior of active social media users, finding that communication, literacy, and demographic antecedents have strong effects on privacy self-efficacy. The third project introduced situated information flow theory to describe cases where information flows across contexts. The fourth project proposed a study examining whether compelled surveillance in religious communities has any effect on privacy norms and perceptions. The fifth project used sentiment analysis to evaluate articles published in several countries about the Cambridge Analytica scandal and extract social norms that were violated by Facebook's actions. The sixth presentation described a method, amenable to crowdsourcing, of annotating privacy policies with CI flow parameters in order to simplify analysis. The seventh presentation investigated how unexpected increases in exposure to nominally public online content triggered norm violations.

PrivaCI Challenge

Symposium attendees worked together in the "PrivaCI Challenge" to map CI concepts to real world case studies. Each table of attendees was assigned a case study about online privacy, such as "Saint Louis University will put 2,300 Echo Dots in student residences" with corresponding news articles. The goal was to determine norms relevant to each situation and identify CI information flow parameters (attribute, subject, etc.) involved in each case study. The difficulty of the PrivaCI Challenge revealed a need for additional work bridging CI theory to application, especially as real world scenarios do not always involve clear-cut information flows and may involve value judgements unrelated to information transfer.

CI in Smart Homes and IoT

This session explored privacy concerns regarding Internet of Things devices. The first project used CI to frame interviews and surveys to determine how privacy opinions and practices of IoT device owners change over time. The second project discussed designing AI agents to obey social norms. The third project explored how to define, apply, and communicate norms for passively listening IoT devices when existing privacy controls are insufficient.

CI in HCI

This session explored the use of CI in human-computer interaction (HCI) research. The first presentation investigated children's and parents' understanding and norms about online information sharing. The second project analyzed the importance and influence of implicit and explicit emotions on privacy-related opinions and decision making. The third presentation described a literature review of articles in HCI venues containing the keyphrase "contextual integrity."

CI Based Systems Design

This session focused on using CI to inform the design of privacy preservation systems. The first two studies used CI to evaluate how users make decisions about application permissions. This data informed the design of new privacy settings interfaces that take contextual norms into account. The third study described the development of a data manager that programmatically enforces a data handling policy by intermediating third party application data fetches and adapting to context changes.

Discovering Users' Privacy Expectation

This session focused on methods and challenges of ascertaining people's privacy preferences. Each of the projects presented in this session involved connected devices. The first used CI to generate interview questions for patients employing IoT devices to assist aging in place. This work suggested that CI could be used as a replacement or supplement to fair information practice principles (FIPPs). The second project used CI to frame interviews of IoT home device users and

discover whether the information collection and storage practices of these devices violated user norms. The third project generated questions about CI information flows to query patients' privacy opinions about tele-monitoring devices in use during and after heart failure treatment.

Keynote: Understanding Privacy and Contextual Integrity: A Personal Journey
Anupam Datta (CMU) described several projects relating U.S. legislation and tech company privacy policies to contextual integrity. He and his collaborators translated HIPAA into first-order temporal logic formalized from the descriptive component of CI, discovering limitations in both the specificity of the law and the expressivity of the CI framework. Anupam also applied CI to build automated systems to help engineers bootstrap privacy compliance in big data systems. He raised several interesting questions at the conclusion of the keynote, including how to handle data "types" in CI, what it means to "use" a type of data, and how we should enforce "purpose" restrictions in privacy policies and legislation.

Conclusion
The symposium concluded with the hope that contextual integrity will continue to gain support and enthusiasm from researchers, industry actors, and policymakers. Attendees noted that technology companies are paying increased attention to privacy issues, especially in response to the European General Data Privacy Regulation (GDPR). This provides an opportunity for the notion of *context* espoused by contextual integrity to play a fundamental role in the design of new privacy-preserving features.

Takeaways and Future Work
The symposium demonstrated significant excitement for incorporating CI into a variety of technical research areas in need of improved privacy frameworks. However, it also highlighted the diversity of interpretations of CI (e.g. inconsistent sets of information flow parameters), as well as difficulties adapting the strict definitions of CI to real world situations and documents. The symposium also showed that projects involving privacy and any notion of "context" sometimes co-opt the CI label without necessarily incorporating ideas of information flow appropriateness or contextual social norms essential to the framework.

These points suggest the need for an updated version of CI, a CI 2.0, that would better define how the core elements of CI should be applied to real-world vagaries. This CI 2.0 could, among other improvements, incorporate cross-context flows, prevent "transmission principle" from becoming a catch-all parameter, and further clarify the ideas of norms, information flows, and contexts. Discussions and development of this CI 2.0 would be appropriate for future symposia, as input from individuals from a variety of backgrounds will needed to ensure that the framework is both theoretically rigorous and practically useful across disciplinary boundaries.

CI and Society #1

Chair: Jake Goldenfein
Notetakers: Karla Badillo-Urquiola, Jaulie Goe

Contextual Integrity as Commons Governance in Online Political Organizing
Madelyn Rose Sanfilippo and Katherine Strandburg

Madelyn Sanfilippo presented her work on "Contextual Integrity as Commons Governance in Online Political Organizing." This work focused on integrating the Contextual Integrity Framework with the Governing Knowledge Commons Framework to expand existing questions regarding Background, Attributes, Governance, Patterns, and Use. She and her collaborator studied empirical cases of online social movements: Women's March, March for Science, and A Day Without Immigrants. They conducted a survey and interviews with the participants, and studied internal documents (scraped from social media).

Q&A
Madelyn was asked the following questions about this research:

- Can you say something about the multi-dimensionality of your approach?
    - *Response*: The authors found that a lot of the norms were intertwined behind the resources. Privacy would shape who was visible but not actually what was going on.
- To what extent were the norms around the knowledge resources coming from the institutional organizations and this new action?
    - *Response*: There weren't general advocacy groups. March for science leaned on existing structures, while Women's March focused more on independent actors and developed their own resources

Knowing and believing: Privacy literacy, privacy self-efficacy and context in privacy-protecting behaviors
Dmitry Epstein and Kelly Quinn

Dmitry Epstein presented his work on "Knowing and believing: Privacy literacy, privacy self-efficacy and context in privacy-protecting behaviors." This work tries to answer the following questions:

- What are the factors that go into how people interact with privacy?
- What effect do those factors have on people expectations and behavior?
- What is the social structure governing how people interact with media?

- What are the predictors of simple/complex privacy behavior?
- What are predictors of vertical/horizontal privacy behavior?

Dimitry's research expanded Helen Nissenbaum's approach by incorporating psychology of people's interactions with privacy. The basic model used for his research was:

- privacy antecedents (age, education, gender, history of awareness)
- privacy concerns (interested in dimensionality of this part)
- privacy outcomes.

The model assumes that concerns are precursors/predictors to privacy behavior. He also examined context as a factor of privacy outcomes. For his experiment, he recruited 600 active social media users, matched to age/income/gender of ACS. He then used principal component analysis and logistic regression to predict high/low levels of privacy behavior. He found that there is a hierarchy of privacy-protecting behaviors:

- activities (such as blocking people)
- disclosure (social measure such as untagging photos)
- use of higher measures (alias, disguise)
- advanced technical measures (using encryption, TOR)

He also found that context of communication is actually a very important measure of privacy activity, and that self-advocacy has limited influence on privacy behavior. Finally, he found that literacy has significant influence on vertical privacy. While some antecedents (age, disposition, literacy) are good predictors of vertical privacy behavior, context is not.

Q&A

Dmitry was asked the following questions about his research:

- What explains the variance of your results?
    - *Response*: The model varies with regard to privacy behavior.
- People have collective (privacy) behavior: Did you see any evidence of this in your survey?
    - *Response*: Network privacy is an important dimension, but we did not ask about protecting others' privacy. We focused on the individual.
- How did you embed, explain, or capture vertical and horizontal behaviors in your data?
    - *Response*: There were questions that asked participants about privacy concerns and behaviors (e.g., protecting yourself from peers versus institutional behaviors)
- How would you evaluate self-advocacy/literacy of a population, and do you feel the lack of role in the result speaks to the role of CI privacy?
    - *Response*: We looked at the relationship between literacy/advocacy and privacy behaviors. We approached the problem from a knowledge standpoint and found an inverse relationship between privacy literacy and age. Also, as literacy increased, higher level activity decreased.
- How did you indicate context?

- *Response*: We asked about 47 ways that people employ social media which resulted in 4 primary mechanisms: communication, entertainment, companion, and information sharing. There were different impacts on what kinds of behaviors were observed in each category.

Situated Information Flow
Sebastian Benthall

Sebastian Benthall presented his work on the theoretical gap in contextual integrity on context clashes. He first introduced the idea of privacy violations due to data reuse, such as when social media profiles are used to develop psychographic profiles of individuals for political ad targeting. Information gets its meaning from the context where it is using, according to contextual integrity theory. The theoretical gap results when data "flows across contexts" with contradictory norms. An important definition was that of a situated information flow. An information flow is a message or signal from which something can be learned due to its nomic associations. A situated information flow is a causal flow situated in the context of other causal relations. Dr. Benthall introduced the idea of using Bayesian networks as a formalism for representing the relationships between random events. Situated information flow theory raises questions about the nature of probability and causality because of the disconnect between real causal relations and people's beliefs about causal structure. This can help determine which cases are appropriate for prevention of data transfer, an issue that comes up during the implementation of omnibus data protection laws.

Dr. Benthall hopes that in the future, more people will be asking *how* data was collected, getting at the generative process of data collection. This could fundamentally change how data is viewed and used. He also hopes to study the differences in data protection laws and explore the circumstances under which general prevention of data transfer is appropriate.

Q&A
The discussion following the presentation focused on:

- Whether the mobilization of data and how it's used as a practice fits into the provided framework. Data's meaning is its use; however, if regulators only restrict data collection, they are making the system vulnerable to exposure threats.
- Does situated information flow theory transform data's "sticky" nature, wherein it attracts other information to itself? If data is split in the generative process, it cannot be recombined, so the model makes data less sticky.

- Possibility theory from differential privacy looks as knowledge not as new info but a statistical fact that can be derived from. Could information flows be modeled in a strict way?

Privacy and Religious Views
Madelyn Rose Sanfilippo and Yafit Lev-Aretz

Madelyn Rose Sanphilippo and Yafit Lev-Aretz examined whether compelled surveillance in religious communities has any effect on privacy norms and perceptions. Religious communities come with a pre-provided context because religion shapes people's privacy norms. The panelists were interested in whether those norms translate to opinions and norms about corporate surveillance. The panelists presented previous literature on surveillance and compelled disclosure in religious communities, including a Pew Religious Landscape Study from the Pew Religion & Public Life division. Relating to privacy, there were some relevant insights:

- People who are religious are likely to turn to prayer to make decisions
- Highly religious people are much more comfortable sharing information with religious leaders rather than trained professionals
- There were differences in perception around information flows among people of various religions
- People were generally skeptical of big government and government surveillance
- Interestingly, stigmatized groups were most comfortable with pervasive government oversight

Q&A
The panelists welcomed discussion from the audience and the conversation that followed touched on:

- Whether there are religions that do not compel surveillance and instead encourage a notion of privacy. Ms. Lev-Aretz mentioned that Judaism does have communities that believe God's omnipresence is misguided. The panelists also discussed how various religions differ in what roles God and community play in information sharing.
- Which religions should be included in the study. The panelists acknowledged that a different approach for each community may be necessary.
- Whether belief in a higher being was a relevant antecedent to privacy. The Pew study captures that beliefs are antecedents to public life of religious people. This is something worth investigating in the future.

CI and Society #2

Chair: Ben Zevenbergen
Notetaker: Priya Kumar

Applying Contextual Integrity to the Cambridge Analytica Case
Catherine Dwyer

Problem: Cambridge Analytica (CA), a political data firm, used an extensive set of Facebook data for predictive models that targeted political ads during the 2016 U.S. Presidential election and the Brexit vote in the U.K. CA case was widely perceived as a violation of privacy norms for social media. How can we use CI to identify in a granular specific way, privacy norms for social media?

How/Why CI used: We have a hard time saying what those norms are for social media. This is related to the privacy paradox – people say they want privacy but disclose information anyway. The way people use social media makes the privacy norms regarding it very vague. This research uses the CA case, which seemed to be a big break/violation of norms, to study what privacy norms are in the context of social media. They apply four CI components (actor, attribute, context, transmission principles) to news accounts of the CA case.

Progress/Results: The researchers used the BYU now corpus of news on the web (corpus.byu.edu/now). They searched for the keywords "Cambridge Analytica" occurring from 3/17/18 (date of first articles in NYT and Guardian) through 5/02/2018 (day CA declared bankruptcy, ceased operations). They recorded URLs and retrieved articles published this date range containing the keywords. This resulted in a dataset of 2777 articles from 520 publications in 19 countries.

They then applied descriptive stats and NLTK sentiment analysis. They showed a publication timeline. There were spikes in the number of articles published right after the news of the CA case broke (10-fold increase in the first week), after the announcement that Mark Zuckerberg would appear before US congress (early April), the day Zuckerberg appeared in front of congress (mid April), and the day CA declared bankruptcy (early May).

The amount of negative sentiment was highest in Bangladesh (14 articles) and Jamaica (9 articles), likely due to small sample size effects. Sentiments of articles from the US (343 articles), UK (255 articles), and India (471 articles) are all pretty negative.

They then used the NLTK collocation tool to find words that commonly co-occur. They took they top 200 co-located terms (e.g., CA, social media, Donald Trump) and tried to map them to CI parameters. The top 5 co-located words for each parameter are as follows:

- Attributes: user data, personal info, personal data, FB data, FB profiles
- Contexts: social media, social network, election campaign, consent decree (because there are rules around how info should be treated), 2016 election
- Actors: CA, Trump, Nixon, Zuckerberg, info commissioner
- Transmission Principles: Improperly obtained, Data breach, built models, secretly recorded gained access.

They performed close readings of sentences in the articles to pull out norms. For example: "*Permission from FB* [transmission principle] to harvest *profiles* [attribute] in large quantities was specifically restricted to *academic use* [context]**.**" The enormous scope of data collected and words that implied power dynamic (exploit, improper) stood out.

In future work, they intend to continue mapping to CI, perform an India case study – high number of articles, publish the dataset or version of it on Github, use CI to specifically and granularly describe social media norms for use by policymakers and legislators to strengthen privacy protections.

Q&A

The following questions were asked about this work:

- Fascinating work. One asterisk. You're looking at the elite discourse and how elites talk about privacy. We should differentiate. That's another context. Elites vs. non-elites who use social media. When you ask people to describe privacy, that differs from how academics report them. There also may be cultural differences to look for.
    - *Response:* Good point. We haven't looked at the sources to see if they present different viewpoints
- How would it change your analysis to push the sample back to the election or before the election? Some media conversation on CA occurred right after election, c.f. Dec. 2016 NYT.
    - *Response*: Did talk about Ted Cruz, which happened before the election. However, the CA case seemed like the social media privacy big bang. It got a lot of people's attention. I do think that in terms of understanding norms, you can look at the vocabulary used in this context and compare it to another.
- I want to understand the logic of the study. The aim is to extract norms. You get the articles and zoom in to identify sentences that identify CI parameters.
    - *Response:* They describe what they were upset about. Is there a way I could restate that sentence through CI?

- *Audience Response:* Yan will talk about this in the next presentation, including the use of Mechanical Turk. It's interesting to triangulate content analysis versus surveys to see how things emerge that show commonalities. When courts need to establish reasonable expectations of privacy, norms can be a strong reference point.

Analyzing Privacy Policies Using Contextual Integrity Annotations
Yan Shvartzshnaider, Noah Apthorpe, Nick Feamster and Helen Nissenbaum

Your inbox on May 25 was probably full of emails about GDPR updates. However, privacy policies re lengthy, hard to parse, written with legal lingo. Case in point: Dima Yarovinski's I AGREE artwork. It's hard to rigorously analyze changes between privacy policy versions.

This is not a new problem – usablepriavacy.org, Terms of Service; Didn't Read, and other research has used ML and NLP to analyze privacy policy text via lexical or semantic analysis to find relevant or interesting paragraphs, track changes in TOS, and crowdsourcing ranking of privacy statements.

We felt something was missing: the framework was missing. Privacy policy analysis needs a theoretical framing to help you analyze, think, and assess implications of what privacy policies actually say.

Method: We used CI as the framework to do this. The idea of appropriate information flows maps nicely onto privacy statements that discuss information handling. We take privacy statements and annotate words associated with senders, recipients, attributes, transmission principles, and subjects (almost always you, the consumer)

Goal (not exhaustive, but what they focused on now):

- Compare CI parameters between privacy policies
- Identify incomplete info flows (missing one or more parameters)
- Identify info flows suffer from "CI parameter bloating" (multiple CI parameters of the same type in the same flow), so when you go to study the implications of the flow, you can't figure out which flows made by the cross-product of the multiple CI parameters are actually occurring.
- Identify vague and ambiguous flows that can, might, in general, sometimes, etc., happen

Facebook Case Study: Used methodology to annotate and analyze previous and updated versions of Facebook's privacy policy. Saw increase in the description of number of flows. But more

flows doesn't mean more clarity! The updated policy increases the number of flows, senders, many more unique attributes, more recipients, more transmission principles. 45% of flows were incomplete in the previous policy compared to 68% in the updated policy. Failing to specify parameters introduces ambiguity, leaving consumers uninformed about company behaviors. Literature shows that when you don't specify the conditions, users project the models they have in their head onto the policy. Gave example of CI parameter bloating. Cited Bhatia et al, 2016 work that also found vague or ambiguous flows.

Crowdsourcing Annotations: constructed CI annotations on Mechanical Turk. Way to annotate more policies. 99 of 143 turkers made it through the 3 screener questions, analyzed 48 policy excerpts, took majority view. Turkers were quite good, but there were a variety of errors. The most common error was skipped parameters, but there were also errors due to ambiguous parameters, overlapping parameters (software didn't let them overlap), and mislabelings.

Future Goal: produce large corpus of privacy policy annotations to discover trends within and across industries.

Q&A

- Aren't Florencia Marotta-Wurgler et al. also doing something similar?
    - *Response*: Yes, they analyzed changes in privacy policies. The biggest contribution of our work is that we bring the CI framework into it. Plenty of other work also analyzes paragraphs, sentences, etc.
- Could these firms employ a consistent language and approach? Could you standardize to make the privacy policies more concise using the vocabulary of CI.
    - *Response* Yes, we want to work with HCI people to do this. Our motivation was to write better privacy policies. Some good takeaways: 1) Policies with missing CI parameters allow users to interpolate from their expectations. 2) CI doesn't cover everything in privacy policies. CI is just about information flow. Privacy policies also discuss user controls etc. that other frameworks may handle better.
- I authored the original privacy policy for Windows. It's important to distinguish the audience. Policies are written both for end users and for experts to define the definitive behavior. Legal people will say don't write separate policies for these different audiences because you end up declaring two different types of behaviors. Instead you end up with a single complicated policy. It's awesome to have CI as a framework for the definitive behaviors for the system so you don't miss the flows and so you have completeness. This might be best done not for the human end users but for automating the system analysis.
- Regarding applying NLP – if multiple experts reading a policy can't come to consensus, how can we have any ground truth when trying to build models?

- *Response:* This work isn't based on NLP. We used manual annotations from experts and crowdsourcing, but that is a general problem of privacy policies being confusing. Our method can be used to improve the text and make it clearer. One of our final goals is to create a corpus of these annotations to feed into algorithms and trained to retrieve info. This is a start.

Enforcing Contextual Integrity With Exposure Control
Mainack Mondal and Blase Ur

There are a variety of definitions of privacy (Warren and Brandeis 1890, Solove 2008, Nissenbaum). Contextual Integrity is a framework to argue about privacy violations with a normative model about privacy as appropriate flows of information. Conceptions of privacy are based on dynamic ethical concerns, so might change over time. CI also explains the "what" of privacy, what people understand. A subsequent step is to build privacy preserving mechanisms in systems. The state of art to do this is via access control mechanisms. For example, whenever you want to upload content on Facebook, you can specify access control lists (ACLs) of who can view content. Privacy violations happen when someone who is not in the list views the content.

The access control model is useful to enforce contextual integrity when all five parameters are explicitly expressed via ACLs. Consider the of a drunken photo. A freshmen posts it as public on Facebook but only expects that people at the party would see it. However, someone then posts it on Reddit, so more people saw it. The privacy violation happened due to increased accessibility; someone who the user did not expect viewed the content. In CI terms, both recipient and transmission principles in the flow are changed, resulting in a norm violation. Access control is inadequate to capture these violations. This resonates with privacy in public that CI moves us toward.

Example 1: Facebook news feed. Facebook pushes your content automatically into a feed that people can see. People saw this as a privacy violation. CI was violated because the expected set of recipients is increased. Facebook argued (correctly) that access control didn't change before/after news feed.

Example 2: In 2012, Facebook Timeline started indexing content by upload date. Anyone can search content by time. After Timeline, old content became easily searchable and more accessible. Users felt their privacy was violated. But again, access control didn't change. However, transmission principle changed; before you would have to scroll all the way through content; now you can click a button.

Example 3: Spokeo. Service that aggregates public data from the web. Other users see this data. Hiring managers used this data to make hiring decisions. After aggregation, inferring nonpublic data became easier. Users complained that it was a privacy violations. Contextual integrity violated, but access control did not change.

Takeaway: In each case, CI was violated. But access control was not violated. Therefore, access control is inadequate to capture use intention / expectation and thus to enforce CI.

New model: exposure. See universe of all users and subset of people who can access. Know prominence of data at time t. Also exists an exposure set of all who will eventually know data (time t to infinity). However, users are bad at inferring how many people can see content (Bernstein, et al.). There may be a feeling of privacy violation when actual exposure differs from expected exposure. In the Facebook news feed, the exposure of content changed. Same for Facebook Timeline and Spokeo.

Takeaway 2: Exposure control extends access control and accounts for privacy violation in CI.

Challenges:

- How to estimate exposure for content. Might be possible to infer based on ways that sites infer popularity of content. They have enough data from previous content.
- How to present exposure to people? Show actual exposure, estimated exposure, list the people, etc.?
- How to allow users more control – new norms. Could take content offline if more than N people see it. Recalls Genie project plus another work on understanding retrospective privacy management pressures; do you really want your new friends to see your old FB content?

Q&A

- I think this is really interesting. See examples in email systems, e.g. alert “do you realize this would go to 500 people?” You could instrument this sort of alert in other contexts if you have enough knowledge. This is an open research topic for future.
    - *Response:* Consider the case of viral content and public shaming. This can be easily detected because it looks like viral content. There could be a trip that says, for example, “you don’t usually get 100 comments on post, so we hid this, but you can open it again.” Even more simply, you could look into users’ past behaviors to infer whether they want to share old content with new friends or not.
- See research from 4-5 years ago with Yang Wang at CMU where they applied behavioral economics and privacy nudging. The most successful was the nudge that said “this will be seen by all your friends and everyone on Facebook.” Snapchat gives users alerts when

others screenshot their content. However, when content hops from Facebook to Reddit, an individual site will have a hard time making you aware of that exposure. That's where we see a lot of bad virality.
    - *Response:* model will infer the hopping, and the goal is to stop that.
- Nice talk. It seems like lots of violations can spur from bad actors. Are there other weaknesses?
    - *Response:* Exposure does not protect against privacy violation stemming from something the user hasn't done before
- We know people are bad at deciding who they want to share content with. We also don't see people going back and grouping Facebook friends into groups etc. How would you incorporate that?
    - *Response:* What exposure control brings in is implicit signals – if you're talking to a set of 40 people a lot, liking their photos, messaging, etc., it's possible they're close to you and should see your updates. On the other hand, someone who you befriended 5 years ago but you never talk to may not be suitable to receive your updates. We can leverage these signals, because users may not know who they want to share with.

PrivaCI Challenge: Context Matters

Yan Shvartzshnaider, Marshini Chetty, Helen Nissenbaum

What Is the PrivaCI Challenge?

The PrivaCI challenge is designed for evaluating information technologies and to discuss legitimate responses. It puts into practice the approach formulated by the theory of Contextual Integrity for providing "a rigorous, substantive account of factors determining when people will perceive new information technologies and system as threats to privacy" (Nissenbaum, H., 2009).

In the symposium, we used the challenge to discuss and evaluate recent-privacy relevant events. The challenge included 8 teams and 4 contextual scenarios. Each team was presented with a use case/context scenario which then they discussed using the theory of CI. This way each contextual scenario was discussed by a couple of teams.

To facilitate a structured discussion we asked the group to fill in the following template:

Context Scenario: The template included a brief summary of a context scenario which in our case was based on one of the four privacy news related stories with a link to the original story.

Contextual Informational Norms and privacy expectations: During the discussion, the teams had to identify the relevant contextual information norms and privacy expectations and provide examples of information flows violating these norms.

Example of flows violating the norms: We asked each flow to be broken down into relevant CI Params, i.e., Identify the actors involved (senders, receivers, subjects), Attributes, Transmission Principle.

Possible solutions: Finally, the teams were asked to think of possible solutions to the problem which incorporates previous or ongoing research projects of your teammates.

Context scenario St. Louis Uber driver has put video of hundreds of passengers online. Most have no idea. Their driver was streaming a live video of them to the internet, and comments from viewers were pouring in. https://www.stltoday.com/news/local/metro/st-louis-uber-driver-has-put-video-of-hundreds-of/article_9060fd2f-f683-5321-8c67-ebba5559c753.html	**Norm 1**	**Norm 2**
What are contextual information norms/privacy expectations?	I don't expect to be recorded in a cab	Recording for security purposes
Example of flows violating norms:	**Flow**	**Flow**
Break down the flow into CI terms: **Actors:** (Identify the actors involved (senders, receivers, subjects)	Subject: passenger, Sender: cab driver, Recipient: Anyone with access to the Internet	Subject: passenger, Sender: cab driver, Recipient: Anyone with access to the Internet
Attributes: (What type of information being transferred)?	Your video	Your video
Transmission Principle: (What are the constraints?)	Public	Public and For entertainment purposes
Possible solution: (e.g., change in policy, GUI, technology, regulation, etc)	a) An option to opt out service	a) Make it an option to opt out service and b) explain why you introduce the service

What Were The Privacy-Related Scenarios Discussed?

We briefly summarize the four case studies/privacy-related scenarios and discuss some of the takeaways here from the group discussions.

1. St. Louis Uber driver has put a video of hundreds of his passengers online without letting them knowing.

https://www.stltoday.com/news/local/metro/st-louis-uber-driver-has-put-video-of-hundreds-of/article_9060fd2f-f683-5321-8c67-ebba5559c753.html

2. "Saint Louis University will put 2,300 Echo Dots in student residences. The school has unveiled plans to provide all 2,300 student residences on campus (both dorms and apartments)." https://www.engadget.com/2018/08/16/saint-louis-university-to-install-2300-echo-dots/

3. Google tracks your movements even if users set the settings to prevent it. https://apnews.com/828aefab64d4411bac257a07c1af0ecb

4. Facebook asked large U.S. banks to share financial information on their customers. https://www.wsj.com/articles/facebook-to-banks-give-us-your-data-well-give-you-our-users-1533564049

Identifying Governing Norms

Much of the discussion focused on the relevant governing norms. For some groups, identifying norms was a relatively straightforward task. For example, in the Uber driver scenario, a group listed: "We do not expect to be filmed in private (?) spaces like Uber/Lyft vehicles." In the Facebook case, one of the groups articulated a norm as "Financial information should only be shared between financial institutions and individuals, by default, AND Facebook is a social space where personal financial information is not shared."

Other groups, could not always identify norms that were violated. For example, in the same "Google tracks your movements, like it or not" scenario, one of the teams could not formulate what norms were breached. Nevertheless, they felt uncomfortable with the overall notion of being tracked. Similarly, a group analyzing the scenario where "Facebook has asked large U.S. banks to share detailed financial information about their customers" found that the notion of an information flow traversing between social and financial spheres unacceptable. Nevertheless, they were not sure about the governing norms.

The unfolded discussion included whether norms usually correspond to "best" practice, due diligence. It might be even possible for Facebook to claim that it is all legal and no laws were breached in the process, but this by itself does not mean there was no violation of a norm.

We emphasized the fact that norms are not always grounded in law. An information flow can still violate a norm, despite being specified in a privacy policy or even if it is considered legal, or a "best" practice. Norms are influenced by many other factors. If we feel uneasy about an

information flow, it probably violates some deeper norm that we might not be consciously aware of. This requires a deeper analysis.

Norms and privacy expectations vary among members of groups and across groups.

The challenge showcases the norms and privacy expectations may vary. Some members of the group, and across groups, had different privacy expectations for the same context scenario. For example, in the Uber scenario, some members of the group, expected drivers to film their passengers for security purposes, while others did not expect to be filmed at all. In this case, we followed the CI decision heuristic which "recommends assessing [alternative flows'] respective merits as a function of the of their meaning and significance in relation to the aims, purposes, and values of the context." It was interesting to see how by explaining the values of a "violating" information flows, it was possible to get the members of the team to consider their validity in a certain context under very specific conditions. For example, it might be acceptable for a taxi driver to record their passengers onto a secure server (without Internet access) for safety reasons.

Contextual Integrity offers a framework to capture contextual information norms.

The challenge revealed additional aspects regarding the way groups approach the norm identification task. Two separate teams listed the following statement as norms: "Consistency between presentation of service and actual functioning," and "Privacy controls actually do something." These outline general expectations and fall under the deceptive practice of the Federal Trade Commission (FTC) act; nevertheless these expectations are difficult to capture and asses using the CI framework because they do not articulate in terms of appropriate information flows. This also might be a limitation of the task itself, due to time limitation, the groups were asked to articulate the norms in general sentences, rather than specify them using the five CI parameters.

Norm violating information flows

Once norms were identified, the groups were asked to specify possible information flows that violate them. It was encouraging to see that most teams were able to articulate the violating information flows in a correct manner, i.e., specifying the parameters that correspond to the flow. A team working on the Google's location tracking scenario could pinpoint the violating information flow: Google should not generate flow without users' awareness or consent , i.e., the flow can happen under specific conditions. Similar violations identified in other scenarios. For example, in the case, where an Uber driver was streaming live videos of his passengers onto the internet site. Here also the change in transmission principle and the recipient prompted a feeling of privacy violation among the group.

Finally, we asked the groups to propose possible solutions to mitigate the problem. Most of the solutions included asking users for permissions, notifying or designing an opt-in only system. The most critical takeaway from the discussion on the fact that norms and users' privacy expectation evolve as new information flows are introduced, their merits need to be discussed in terms of the functions they serve.

Summary

The PrivaCI Challenge was a success! It served as an icebreaker for the participants to know each other a little better and also offered a structured way to brainstorm and discuss specific cases. The goal of the challenge exercise was to introduce a systematic way of using the CI framework to evaluate a system in a given scenario. We believe similar challenges can be used as a methodology to introduce and discuss Contextual Integrity in an educational setting or even possibly during the design stage of a product to reveal possible privacy violations.

Additional material and resources

You can access the challenge description and the template here:
http://privaci.info/ci_symposium/challenge

The symposium program is available here.

To learn more about the theory of Contextual Integrity (CI and how it differs from other existing privacy frameworks we recommend reading "Privacy in Context: Technology, Policy, and the Integrity of Social Life" by Helen Nissenbaum.

To participate in the discussion on CI, follow us on Twitter @privaci_way
Visit the website: http://privaci.info
Join the CI mailing list

https://www.zotero.org/groups/2228317/privaci/

References

Nissenbaum, H., 2009. Privacy in context: Technology, policy, and the integrity of social life. Stanford University Press.

CI in Smart Homes and IoT

Chair: Güneş Acar
Notetaker: Erica Du

Disentangling Privacy in Smart Homes
Martin J Kraemer and Ivan Flechais

Problem: How do privacy practices in smart homes change over time?

Approach: Longitudinal panel study with interventions to investigate processes of formation and development of privacy practices in the home

Why CI?: We perceive homes as our most private spaces. Homes are quite involved with respect to socio-cultural context: social order, moral order, social settings, cohabitation, etc. New technology is being introduced into this space, which poses new challenges for living. Every individual has specific ideas of what privacy is and holds different expectations.

How CI?: Discovering contexts within the smart homes. Discovering teleological structures. Understanding relevant contextual values of home practices to inform arguments of moral legitimacy. How do we accept new technical systems? Many systems will violate this contextual integrity, what is the underlying value being provided to users? Learn how to talk to users and educate them about contexts and normative values.

Project Progress, Future, Challenges:
Exploring social order and device usage:

- Varying perceptions of responsibility and care, as well as tensions as a result of traditional gender roles and biases (i.e. home domain traditionally associated with females, where as tech traditionally associated with males)
- Dichotomy of usage expectations and provided functionality. When the device is not providing service expected, some users find workaround whereas some stop using the device
- Challenges in self agency in light of anticipated and unanticipated behavior (i.e. it doesn't provide the benefit I thought it would do)
- Example of changed behavior: wife moves banking activities to different room to avoid surveillance camera installed by husband to keep property safe
- Assumed or articulated consent: when there is more than one person living in a household, you would hope that the use of devices would be agreed upon (since the device affects all people!), but sometimes there isn't that kind of consent

How do these practices change over time?

- Practices are a result of social order formed through negotiation and users' expectations, norms, and purposes
- Practices evolve due to technology and users' mental models
- Practices are shaped by attitudes/preferences established through privacy debates and available controls

Disentangling Privacy:

Methodology combining diary studies with interviews and interventions, combining elements of observation with inquiry to produce qualitative data and enrich with quantitative data

- Device appropriation: tech intervention
    - First, establish baseline. What devices already exist? What is existing understanding of technology?
    - Start by giving them a new device: Alexa, smart TV stick. There are also arguments around what the characteristics of these devices should be
- Mental Models: Privacy Intervention
    - Highlight privacy controls and how to use them
    - Observe if and how that influences behavior. How does it change behavior? How does it change the way users think?
- Choice architectures and options
    - Look at who and what structures users choose to share data with X company and/or share post on social media with Y group
- Participants: 6 households with focus on UK

Q&A

- Why only 6 households?
    - *Response:* Feasibility, not enough resources
- Do you have any anecdotes to illustrate?
    - *Response:* A family installed a smart light system before Christmas so they can turn on their Christmas lights easily. They kept the smart system long after Christmas. The thing is, you need to have the house light switches flipped on all the time in order for the system to work. However, the other person in household would accidentally turn it off, and family members would need to message person to turn system on from work.
- Are you making this operational for IOT developers? Can they take findings and do something in design?
    - *Response:* Third study would be to get together with product designers and software developers and see what they mean to do and then do a comparison.

What do they expect and what do they want? Where does it fit in GDPR? Is it a good fit or should we improve the way its being implemented and used?

- Probe on teleological structures. How do people approach these systems? What other considerations need to be added to contextual integrity model when people are in decision-making process?
    - *Response:* People use devices if the devices further goals, and might ignore side effects they're not aware of. We hope to flesh out this situation.
- Can you speculate on normative exploration, because it's less usual for people to go into that direction? You're looking at specific context, the household. What kinds of values are being disrupted by these practices? For example, the wife doing household accounts somewhere else. What are you looking for in the situation?
    - *Response:* We see concerns about Alexa use, but smart meters are not perceived in the same way. People perceive energy savings as a good thing, and are therefore are not bothered by smart meters. As another direction, we talked to a superuser who's never owned an Alexa and thinks they're spying on everything and doing the same thing as CCTV.

On Engineering AI Agents for Privacy
Rafa Gálvez and Seda Gürses

Problem: Can an AI agent respect contextual integrity? What kind of influence does the engineering process have on the behavior of the agent? We look at interplay of knowledge base methodology and machine learning methodology to empower engineers to think about both of these methodologies to solve complex problems

Why CI?: Engineers think about social norms. Engineers can also think about informational norms. Let's help engineers think about and understand how their software may violate their user's norms. Recommendations are very dependent on information flows, and not all may be appropriate. An AI agent performs intelligent tasks; the model and code both influence behavior

Challenges:

- What is an AI agent? Is there a definition?
- Is the engineering process and the behavior of the agent in a single social context?
- Is it appropriate to use a proxy feature to get access to an inappropriate feature?
- What about composition of contexts?

Future:

- Unified AI engineering methodology

- Relate steps to corresponding contexts

Q&A

- What makes reinforcement learning special?
    - *Response:* Time. Learning from interactions w/ users. So there is another dynamic parameter involved: planning.
- Should we as a society not allow this? Or should individual owners of these devices say we don't want our Alexa to integrate this factor about our kid.
    - *Response:* Individual norms? Even if engineer took into account in design, agent might still be able to use info in incorrect context
- Family is context. But every family is different. For example, in some families, the genders are treated very differently while in others they are treated similarly. There are different norm profiles in every family. Can you learn about the context? Or would you select the norm? Is it user inputted or learned?
    - *Response:* There is a paper on social networks, authors write about implicit contextual integrity, agent detecting norms by how people interact, and then enforcing it. If agent learns you are complaining, then agent will learn about the appropriate norm. We want to empower engineers to ask these kinds of questions.
- Take issue with a venn diagram in the presentation. The field of machine learning is much broader, while reinforcement learning is a subset.
    - *Response:* Agree, that's why it's a challenge. How do we incorporate knowledge based in reinforcement learning.

Use Case: Passively Listening Personal Assistants
Nathan Malkin, Primal Wijesekera, Serge Egelman, David Wagner

Problem: Privacy protections for always listening devices like Google Clips as they become valuable enough that people will actually buy them. Can we still get benefits but prevent use cases we're worried about?

Why CI?: Existing solutions on privacy controls are probably not good enough – think of the standard iOS/mobile permission control popup that asks to "Allow use of microphone." Perhaps can use CI to help devise privacy protections.

Challenges:

- What do we mean by context? How do we define context?

    - "Home" is very broad. Maybe there are multiple contexts, so how do we infer which contexts you're in? It's hard for a human, are we going to ask a computer to do it for us?
    - Even if we have a context, how do we learn norms?
- How do we apply norms? How do we operationalize from both normative and engineering perspectives? How do we deal with the fact that users might have different preferences even in same household?
- How do we communicate this with users?
    - This is exacerbated because the user interfaces are becoming more and more minimal. Example: no screen, just a speaker

Q&A

- How do these shift when devices become mobile? Like google clip moves!
- What direction are you going in? Qualitative studies? Develop new interfaces? Develop new framework to think about?
    - *Response:* Starting with qualitative. Trying to figure out norms we might want to apply. When we do design, we'll know what to look for.
- When you're looking at how to build this methodology to evaluate these systems, are you doing top/bottom or bottom/top? Take specific contexts or more broadly apply?
    - *Response:* So far we think bottom up. Think of specific use cases, using these use cases, what are the features it would need, when would it need to listen, what are the limitations we would need to put in place, then, how to generalize?
- One factor to infer is the context you might want to consider and what value the device gives back to the user. Google Clip value adds the natural photos of user and child. What does the device do for the user? This will change the context, because the device is here to serve me rather than surveil me.
    - *Response:* That's why bottom up is more useful. If you start with when is it appropriate for device to listen, you'll get never! But when you start from use case you can see that you need to make trade offs
- What's the data you're going to use to learn norms? How do you account for the fact that it's an incredibly political decision?
    - *Response:* That's one of the questions I hope to get feedback on. I have some qualitative techniques but formulating survey questions is very hard w/o building bias.
- Every system must embody some norms. Everyone is going to make assumptions. I want us to be aware that we're not just talking about camera and kids, we're thinking about intermediaries in the loop, any of these design questions, and what is the alternative to whatever approach we're looking at? Each one of these brings the baggage you point to.

CI and HCI

Chair: Michael Byrne
Notetakers: Marijn Sax, Jaulie Goe

Contextual Integrity as a Conceptual, Analytical, and Educational Tool for Research
Priya Kumar

In this presentation, Kumar discusses how she has worked with the theory of Contextual Integrity ('CI') in three different research projects.

The **first project** where she used CI dealt with unexpected social media flows.[1] More specifically, the project examined the phenomenon of parents posting pictures of their children online and the privacy implications of this practice. In order to unpack the norms around this practice, the blog 'STFU, Parents'[2] was studied. This blog addresses 'parental oversharing' in a snarky, humorous fashion by sharing screenshots of (alleged) parental oversharing on social media and commenting on them in a snarky tone. In unpacking the norms that are propagated by this blog, Kumar was struck by the fact that privacy did not come out as a prominent source of (snarky) critique. Applying the CI framework to this case, it became clear that (somewhat ironically) the blog *itself* was taking content out of one context (i.e., a social online context where pictures of kids are intended to be shared with friends and family) to reproduce it in a different context (i.e., a context of humor and social criticism where the parent where made an example of parental oversharing). This could be construed as an unexpected flow of information from the perspective of the parents. Moreover, the blog appears to prioritize the content and the way in which it can be used to convey a particular message. The blog does *not* seem to prioritize the subjects, i.e. the persons in the pictures. Kumar concludes by explaining how CI served as a useful meta-analytical tool in this project; CI helped to see the data in a different light.

The **second project** focused on how privacy expectations can change over time. Interviews with Fitbit users were conducted in order to ask them about their perspective on the information flows that are occurring, or can occur, when using a wearable device such as a Fitbit. These interviews were conducted in 2017. The results were compared with similar interview data from a 2013 by Heather Patterson.[3] A comparison was made between 5 context were health information can

[1] Kumar, P. (2018). Emerging Norms and Privacy Implications of Parental Online Sharing: The Perspective of the STFU, Parents Blog. Presented at the 68th Annual Conference of the International Communication Association (Prage, Czech Republic, 2018), 1–30.
[2] http://www.stfuparentsblog.com/
[3] Patterson, H. (2013). Contextual Expectations of Privacy in Self-Generated Health Information Flows. *TPRC 41: The 41st Research Conference on Communication, Information and Internet Policy.* Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2242144

(potentially) be shared, namely with one's (1) doctor, (2) insurer, (3) employer, and with (4) law enforcement, and (5) on social media. The comparison brought out the following results. In 2013 sharing self-generated (Fitbit) health data with one's doctor was considered to be largely appropriate, while this was questioned (even) more in 2017. In the insurance case, respondents were split on whether sharing was appropriate in 2013. In 2017, however, respondents were more willing to consider sharing self-generated health data with their insurer (for instance in order to receive benefits such as lower premiums). The employer, law enforcement, and social media context all showed a similar development: in 2013 respondents clearly indicated that the self-generated health data should *not* flow to these actors. In 2017, respondents were willing to consider such flows of information. Kumar explains that for this study, CI served as both the conceptual framework that inspired the study and as the analytical tool through which the data were interpreted.

The **third project** focused on children's understanding of privacy online. Family interviews were conducted for this project. As it turned out, children typically had a general understanding of how different actors and attributes of data can affect privacy online. Children under the age of 10 did not discuss any transmission principles.[4] Besides these descriptive findings, Kumar also discussed some prescriptive implications. Rather than providing children with a range of explicit do's and don'ts, Kumar argues[5] that children should be equipped with 'privacy decision-making skills' that help them to make informed choices online. CI could form the basis of an educational tool aimed at helping children develop privacy decision-making skills to navigate privacy online.

Q&A

- The first question is a suggestion for a possible collaboration. Earlier research has found that parents do not really take context into account when they are parenting their children. The often provide specific do's and don'ts to children. The suggestion is to maybe use the proposed educational tool to not only teach children, but also parents.
    - *Response:* Kumar answers that she likes that idea. In the family interviews, she learnt that parents often say that privacy does not matter *right now* and that they will get to (the teaching of) privacy when their children are older. The proposed CI educational tool could indeed help parents to teach their children privacy decision skills from an early age. The goal would be to teach children a more

[4] Kumar, P. et al. (2017). "No Telling Passcodes Out Because They're Private": Understanding Children's Mental Models of Privacy and Security Online. *Proceedings of the ACM on Human-Computer Interaction. 1*, CSCW (Dec. 2017), 1–21 . https://doi.org/10.1145/3134699.

[5] Kumar, P. et al. (2018). Co-Designing Online Privacy-Related Games and Stories with Children. *Proceedings of the 2018 Conference on Interaction Design and Children* (Trondheim, Norway, 2018). https://doi.org/10.1145/3202185.3202735.

intentional stance towards their privacy, allowing them to think more explicitly and critically about their own privacy behaviors.

- The second question addresses the second project and the found differences between 2013 and 2017. These differences show that norms are indeed changing. It is a little sad, however, to see *this particular* change, where respondents seem to be more willing to share self-generated health data for potential individual gains. A more 'social' perspective which also address the social changes that come with these kinds of norm changes around the sharing of health data would be welcome.
    - *Response:* Kumar agrees, and also emphasizes that (of course) not all respondents reasoned in a first personal manner and based on (just) individual benefits of sharing. She also agrees that it would be interesting to look more closely into the different motivations of people.

The Emotional Context of Information Privacy
Luke Stark

Stark explains that this presentation is based on a paper that is already published (henceforth 'the 2016 article').[6] That paper was inspired by another article published in 2013 by Bruce Schneier,[7] which sparked Stark's interested into thinking about the ways in which our (personal) privacy thinking is diffused with emotions. For this particular presentation Stark was also inspired by an article that was published a few weeks ago: 'Welcome to the Age of Privacy Nihilism.'[8]

Stark discusses 5 key points from his article from 2016 in order to critically reflect on them 2 years later. This will allow him to explain what he omitted, overlooked, and/or got wrong in 2016. Based on this analysis, he can suggest directions for further research.

Key point #1 from the 2016 article is that emotion is an *implicit* element within the broader theory of contextual integrity; the contexts for our feelings shape what we intuit as 'appropriate' in terms of privacy, and vice versa. Stark mentions as example the phenomenon of mood tracking, which often occurs with the help of apps. More specifically, he mentions the mood tracking app *Mood Panda*.[9] This app is very attuned to the social and emotional elements of tracking, which may explain why people like to use it.

[6] Stark, L. (2016). The Emotional Context of Information Privacy. *The Information Society, 31*(1), 14-27.
[7] https://www.cnn.com/2013/03/16/opinion/schneier-internet-surveillance/index.html
[8] https://www.theatlantic.com/technology/archive/2018/08/the-age-of-privacy-nihilism-is-here/568198/
[9] http://moodpanda.com/

Key point #2 is that today's digital media scramble and subvert the ways we perceive, understand, and express the emotional nuances of our privacy preferences. Stark explains that emotions can functions as signals, or heuristic devices, to signal our attention towards practices we may or may not condone of. This function of emotions can also help explain why people have been so upset about Facebook experiments which misuse our emotions, such as the emotional contagion study.[10]

Key point #3 is that our affective and emotional connection to the hardware and interfaces of our devices is what prompts us to be less conscious of our data and metadata. In other words: we can be *distracted* from the abstract nature of data and data flows. This insight can, of course, also be used (/exploited) by hardware and software designers.

Key point #4 is that much of the time, our embodied interactions with our device seem, in not entirely personal and private, then at least adequately 'contextual.' When using technology, we tend not to consider very blatant and/or outrageous privacy violations: "they can't be *that* bad." An example such as the Facebook experiment aimed at identifying (in real time) the emotional state of Australian adolescents for advertising purposes[11] serves as a good reminder that, for instance, social media platforms can in fact engage in practices far beyond of what we consider "adequately contextual."

Key point #5 is that both corporations and governments work hard to keep the emotional impact of data they possess about us at a minimum. This is a clear of example of where the 2016 article got it wrong. Consider, for example, the amounts of data Tinder keeps (and uses) of its users, as was shown by the journalist from *The Guardian* who requested all of her data.[12]

Stark continues by taking a closer look at the (implicit) role emotions play in CI. He references a passage in Nissenbaum (2015)[13] where she writes that "[the] social domain is the only

---

[10] See https://www.washingtonpost.com/news/the-intersect/wp/2014/07/01/9-answers-about-facebooks-creepy-emotional-manipulation-experiment/?utm_term=.ea95dfcf7b27 and https://www.pcworld.com/article/2450900/privacy-group-files-ftc-complaint-over-facebooks-emotional-contagion-study.html and https://www.theguardian.com/science/head-quarters/2014/jul/01/facebook-cornell-study-emotional-contagion-ethics-breach and http://www.wbur.org/cognoscenti/2014/07/15/experiment-facebook-steven-brykman.

[11] See, e.g., https://www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens and https://arstechnica.com/information-technology/2017/05/facebook-helped-advertisers-target-teens-who-feel-worthless/

[12] https://www.theguardian.com/technology/2017/sep/26/tinder-personal-data-dating-app-messages-hacked-sold

[13] Nissenbaum, H. (2015). Respecting Context to Protect Privacy: Why Meaning Matters. *Science and Engineering Ethics*. Available at: https://link.springer.com/article/10.1007%2Fs11948-015-9674-9

interpretation of contexts that marks a meaningful departure from business as usual." Referring to Hochschild,[14] Stark emphasized that emotions are fundamentally social and, therefore, also fit into the analysis in Nissenbaum (2015). Emotions can thus be used as important signals to understand where practices within a practices are no longer 'business as usual' and thus warrant critical analysis. This is further supported by the fact that Nissenbaum (2015) writes that "the theory of contextual integrity holds the source of […] anxiety to be neither in control not secrecy, but appropriateness."

Trying to connect emotions to specific elements from the CI framework, Stark proposes we can think of the emotions themselves as information types. The relevant (inter)actors can, next, be interpreted to be the 'emoters' which are those with whom we 'emote'. Lastly, transmission principles can be seen as 'emotings', which are the ways in which we 'emote.'

Stark discusses three main omissions/errors in the 2016 article. First, he was too sanguine about propaganda and manipulation. The Cambridge Analytics scandal clearly shows that propaganda and manipulation worries are real and urgent matters. Second, he was too silent on issues pertaining to gender, race, class, and identity. He acknowledges that thinking about emotions and privacy must also incorporate these dimensions. The recent discussion around Serena Williams' behavior during the final of the US Open is a good example.[15] It has been suggested that race and gender played an important role in this final, as Serena Williams' race and gender seemed to constrain the ways in which she was allowed to show and act upon her emotions. Third, he was too optimistic about design and the role it could and/or would play in addressing privacy issues. Stark discusses how both Google and Apple have recently announced the introduction of software (dashboard-like) tools aimed at helping consumers gain a better insight into how they are using their devices. These new tools are, according to Stark, not exactly the game changers that one would hope for.

Looking at potential future research, Stark mentions the concept of 'intuitive data toxicology.' Much like gas companies add an odor to odorless gas for it to be detectable by humans, one could think of possibilities of leveraging emotions and emotional responses to alert people to abstract data flows and data uses, rather than designing technologies that distract users from these already abstract data flows.

[14] Hochschild, A. R. 1(983). *The Managed Heart: Commercialization of Human Feeling*. Berkeley, CA: University of California Press.

[15] See, e.g., https://www.nytimes.com/2018/09/09/sports/tennis/serena-williams-sexism-us-open-.html

Q&A

- The first question is a suggestion to consider a cultural phenomenon that could be interesting for Stark's project. The (norms around) civil inattention which make it impolite to look at the one's neighbor's screen when he/she is texting could be an interesting source of inspiration.
    - *Response:* Stark agrees and will look into this.
- The second question is whether Stark had considered looking in the significance of devices as points of interaction. For example, when thinking, speaking, and writing about Internet of Things systems, people tend to refer to the connected app (with an interface) on one's device, rather than to the entirety of the IoT system. This could be an interesting perspective to take into account when discussing how our emotional dealings with technology can distract us from important features and information flows.
    - *Response:* Stark agrees and will look into this.
- The third question emphasizes how in the CI framework, discussions often start after a *felt* privacy violation. These experiences of violation are often emotional. This suggests that emotions can also be used to signal our attention to violations.
    - *Response:* Stark agrees and explains how this is indeed part of the paper. He also refers to Nissenbaum's PhD dissertation for an interesting earlier treatment of emotions.[16]

Examining Gaps and Opportunities for Engaging with Contextual Integrity in Human-computer Interaction

Karla Badillo-Urquiola, Xinru Page, Pamela Wisniewski

Page starts by explaining why CI is interesting for HCI scholars and why, therefore, it is worth exploring the use of CI in HCI in a systematic manner. Existing theories that classify different information types according to their sensitivities or focus on awareness and control over information do not do a very good job at *really* explaining privacy behavior. As a result, these theories frequently give rise to 'privacy paradoxes.' Page argues that norm-based theories that focus on norm violations (CI being the most prominent one) are better able to help us understand privacy behaviors.

To explore the use of CI in HCI literature, 3 research questions are answered:

1. Within which technology contexts do HCI researchers apply CI?
2. How deeply do HCI researchers engage with CI?

[16] Nissenbaum, H. (1985). *Emotions and Focus.* Chicago: The University of Chicago Press.

3. What types of studies do HCI researchers conduct when applying CI?

The following search criteria were used in order to create a sample of articles in the HCI literature:

1. The article must be published in last 10 years (i.e., 2008-2017)
2. The article must be peer-reviewed
3. The article must be published in one of the top HCI journals/conferences/venues based on GoogleScholar and Microsoft Academic Research rankings (this generated a list of 32 top HCI venues from which articles were selected).

To search the 32 top HCI venues the keyword "contextual integrity" (case insensitive) was used. This resulted in 24 hits. After inspecting these hits, not all articles met the search criteria. In the end, 15 articles were found. In order to code these articles, the following coding dimensions were used:

1. Technology context: the type of technology studied
2. CI engagement: to what extent CI was used in the research
3. Type of HCI research conducted: whether the study was a formative or summative evaluation and the type of scholarly contribution made by the research (e.g., design implications, framework, new technology).

For the *technology used*, the results were as follows:

1. 7 out of 15 studies dealt with Internet of Things
2. 5 out of 15 studies dealt with social media
3. 3 out of 5 studies dealt with mobile devices.

The Internet of Things studies tended to focus on the uncertainties this new technology introduces vis-à-vis existing and new (privacy) norms. Next, the social media studies tended to integrate CI by examining users' privacy perceptions, expectations, and norms as well as their willingness to share information. Lastly, the mobile device studies focused on studying design solutions for privacy concerns related to data leakage, surveillance, and bystander privacy intrusion.

For the *level of engagement with CI* the results were as follows:

1. 9 out of 15 studies referenced CI primarily in background literature
    a. 2 out of these 9 only cited CI
    b. 7 out of these 9 also explained (in some form) the CI framework
2. 2 out of 15 used CI as the guiding framework for understanding privacy challenges
3. 4 out of 14 used CI to inform the study design and data analysis
    a. 2 out of these 4 used CI to inform the codebook

    b. 2 other out of these 4 integrated CI into the actual study design.

For *the types of research that apply CI* the results were as follows:

1. In 8 out of 15 studies, a user study was performed
    a. 7 out of these 8 were formative evaluations
    b. 1 out of these 8 was a summative evaluations
2. 7 out of 15 studies detailed design implications and avenues for future research

In the user studies that were formative evaluations the CI framework primarily informed the design of new technologies by asking users about their perceptions through surveys and interviews. The summative evaluation used aspects of CI to assess a new system.

Based on this overview of the HCI literature that uses CI in some way, the authors provide the following recommendations:

1. Future studies could benefit from operationalizing the framework of CI as a robust multi-dimensional construct. This would allow scholars to with a convenient mechanism to evaluate (entire) systems in terms of its contextual integrity.
2. Future research should focus on summative evaluations of systems that instantiate CI in a meaningful way. Many of the studies provide useful insight into the *design* of future technologies. However, there was no *evaluation* of new technology solutions.
3. There are multiple open-ended questions that are worth discussing, such as:
    a. Different ways to systematically scope the literature (e.g., a forward reference search of Nissenbaum's work; this yielded 79 additional articles).
    b. The qualitative coding (in this case domain, depth of the level of engagement with CI, and the types of research that apply).
    c. How to distinguish between CS and HCI in a systematic fashion

Q&A

- The first question is whether keyword used to search for articles could be refined, for instance by also trying different variations that include, for instance, a hyphen between phrases.
    - *Response:* Page answers that that is a good suggestion.
- The second question is a suggestion that applications to other fields or milieus would be interesting, such as law.
    - *Response:* Page agrees with this suggestion.
- The third question asks about the codebook: is there a codebook available?
    - *Response:* Page answers that there is a codebook available and that Badillo-Urquiola knows (much) more about the codebook. It is also asked how the codebook was constructed. Page answers that only the different context (i.e.,

Internet of Things, social media, mobile devices) emerged during coding and that the rest was determined beforehand.

- The fourth question is whether the researchers have considered a GoogleScholar keyword search to generate a list of HCI articles that use CI.
    - *Response:* Page answers that this was indeed considered, but that (for now) the use of a venue list was most convenient to ensure a proper focus on HCI literature.
- The fifth question asks whether the authors could (in next presentations or publications) speak more about the filtering process. Some of these venues alone have over 3000 publications, meaning that just mentioning that there were 15 hits does not say much. Taking one venue with 3000 publications as an example, it would be interesting to know how many publications in that value are, for instance, about privacy, how many are about security, how many are about privacy and security, and how many of each of these sets use CI.
    - *Response:* Page answers that this is indeed something to consider.
- The sixth question is the suggestion to look into snowballing methods, starting from influential starting points. This would have the benefit of allowing for an analysis of *where* (i.e., in which literatures) CI ends up being used.
    - *Response:* Page agrees and suggest that this would be a good way to identify various ‘CI communities.’

CI based Systems Design

Chair: Darakhshan Mir
Notetaker: Erica Du

Contextual Permission Models for Better Privacy Protection
Primal Wijesekera, Joel Reardon, Irwin Reyes, Lynn Tsai, Jung-Wei Chen, Nathan Good, David Wagner, Konstantin Beznosov, and Serge Egelman

Problem: Reducing the gap between privacy expectations and how current systems are actually protecting privacy.

Goals:

- Create a permission model that allows application to access data when it is expected by the user
- Discover how users expect their privacy to be protected

Specifically, focusing on 20 permission types, what is the best way to ask for permissions on mobile?

Why CI?: CI provides an excellent framework to understand user expectations

Current Progress:

- Retrospective experience sampling
    - Procedure
        - Every time we saw a sensitive background request, we took a screenshot
        - Then we show screenshot
        - Then we ask what you would do
        - Do users want to have option to make decision to block requests?
    - Found that users wanted to vary their privacy decisions based on the requesting app's visibility -- the knowledge that the application is running. Services making request in "invisible" applications likely to be denied
    - Number of requests is way too high to always ask user. For example, Facebook accesses location 4 times per second (240 times per minute)
    - Platforms make decisions on behalf of user, now they want to make contextual decisions, but system has to then make predictions on context
- Collecting Ground Truth
    - Whenever we saw a request, we prompted the user

    - Visibility is the most important factor. Foreground application is also important. Application that is requesting the application is not significant, but application visibility is -- does the user know that the application is running?
    - Role played by front actors also matters – ie. medical data when you're a patient is different than medical data in other contexts
- Implementation in Android
    - Made a predictive model using SVMs (Support Vector Machine), yielded lower error rates that "ask on first use"
    - 37 users, 5.4m requests, 1k privacy decisions. Prompted users 4x a day.
    - Punchline: contextual model works – 5% error vs. 20% error for ask on first use
    - Also implemented new permissions
- Learned that context is applied similarly for everyone
- Future: What's the impact of the purpose of a permission request?

Q&A

- Could you predict what users expect? How consistent are these expectations with what contextual integrity thinks about norms? Might what users expect not be social norms? How would you bridge bridge these differences?
    - *Response:* We don't claim that visibility or foreground, moving forward we want to understand why foreground and visibility are important
- Have you looked at time? Have you looked at the language?
    - *Response:* Most people ignore prompt at least once, but people look at prompt again. We briefly looked at time they spent and decision. In second study, even if say we will deny the request, we will not actually deny request. Third study changed same wording that by denying you are actually denying resource.

Privacy Heuristics: How Users Manage Complex Contexts when Evaluating Mobile App Requests

Amit Rechavi, Eran Toch, Jason Hong

Problem: We are making contextual decisions on what kind of info our apps can access. Permission interfaces have changed over the years, for the better. Our aim is to understand how users make these decisions when they have complex set of context elements, especially when focus on purpose of decision. How do we create an interface to enable users to create social norms around transmission?

Why CI?: Contextual integrity as a descriptive theory for how decisions are made. Users make decisions differently between extreme and normal cases. Context is managed dynamically.

How CI?: First, need to understand what matters to users. We know context matters to users, but we need to understand how users form normative decisions around complex contexts. We don't ask "what are the norms?", we ask "how do you use the phone?".

Progress/Results:

- Using an existing 2012 dataset of 800 Amazon Mechanical Turk surveys about 1,200 applications
- Participants were presented with a scenario that included actual apps and permissions
- Results: all aspects of the context matters
    - Actor dependent norms
    - Application category
    - Comfort score
    - Some actor agnostic norms

There seems to be a model for normative decision-making:

- Evaluating request → is this an exceptional case?
    - Yes → conforms to general norms?
    - No → conforms to contextual norms?
    - yes → accept, no → reject

Q&A

- How are you thinking formally of what a context is? Is it just the purpose? Is it the category? Is it a combination?
    - *Response:* When you look at process of decision-making, we are not sure how users form context. Users use combinations of things, and our job is to understand this
- Conversation, not question -- I'm in favor of that. Concerned about the assumption in people's minds as context in a special case, like "in a context", or the idea of context as constituted by values of various parameters, not that people ask what context we are in, but rather that this is my teacher and we're talking about grades I got
- Have you touched on the challenge of how we operationalize context?
    - *Response:* We see there's a feedback loop. Perhaps some social norms associated with information transmission will help people define what context is, we don't necessarily think about context in advance of transmission

Privacy-Aware Programming for Microscale Data
Jason Waterman, Eugene Bagdasaryan, Matthew Griffith,
Griffin Berlstein, Nate Foster, Eleanor Birrell

Problem: We define microscale data as personal and sensitive data that we generate by living our lives in the digital age. There exists this tension between rich data and privacy. You can't just bolt on security and privacy at the end when building a product, you must develop it with that principle from the ground up. We need better tools for developing these applications.

Progress: ANCILE is the in-between manager of a user's data and a third party application. All requests for a user's data goes through ANCILE and the micro data core. This forces applications to think about they are going to use and the transformations on the data. They must also think about how they're going to use data and present those requests to the user. There is a policy that interactions must adhere to, and ANCILE will programmatically enforce that by taking care of data fetching for 3rd party applications. As system runs, the context changes. This change is handled by the event manager, which changes the policy.

Prototype applications:

- Roaming office hours
    - Data shared: location only in proper context
- Spontaneous study group
    - Location only in proper context
- Selective study participation
    - Anonymized user or nothing

Future Work:

- Implementation of more complex use cases, deployment outside of our lab
- Providing tools for visualizing and creating more complex policies
- Moving from single instance to a federated version of ANCILE, allowing for scaling to large deployments
- Could also go other direction, and move towards a more personal ANCILE

Q&A

- Are these going beyond access control policies in anyway? Can you say more about expressivity of policy language and how it adheres to CI?
    - *Response:* The data you get, a lot of the time it's all or nothing, you allow access to data or not determined by OAuth. Instead of getting access to entire gmail

account, ANCILE can specify filters on operations. It's done inside of core, none of that data gets out of system except for end when you have explicitly authorized it. Language: temporal events, can dependent on arbitrary data, focused on location based apps right now, context based on location and time, could incorporate based on other types of data

- Follow-up worth mentioning, responding to Anupam, re: "location only in proper context." Can you be more precise, is there a way for system to know who your students are, include in recipient category? Can you specify it's your students, do you care what slack is getting?
    - *Response:* Raises very good point about design parameters. We are making this end to end argument to put as little functionality, and bake in as few assumptions as possible. This concept of users, and how does system handle that, is it something that is built into the core? Ancile does have a core concept of users. However, as far as particular groups, we did not bake into system, slack comes into play here bc slack already has a concept of a group
- What about policy frameworks? As you dove deeper, did you hit any walls? Are there any new technical challenges? Is there a technical contribution as well?
    - *Response:* ANCILE is a general purpose framework, meant to support things like contextual integrity, represented as state transition. We tried hard to design touch points to system that are fairly narrow, because we wanted to replace one policy framework with another. There are only a couple of places in system you'd need to modify. As far as challenges, when you add another layer of indirection, we'd like to reduce and mitigate overhead that is allowed, like when context changes, you change policy. We'd like to separate existing policy that gets enacted, and the policies that change.

Lunch and Conversation with Susanne Wetzel, Program Director SaTC, NSF

Notetaker: Noah Apthorpe

Introduction:

- From Stevens Institute of Tech
- Been at NSF for 1.5 years
- Get to experience the process from the "other side," meet young faculty

Talk will Focus on SaTC (Secure and trustworthy cyberspace) program

- SaTC spans 5 NSF disciplines
    - CISE, SBE, EHR, ENG, MPS
    - Interdisciplinary program that makes funding decisions across disciplines
    - Encouraging new collaborations across fields, especially between computing and social sciences
- Spans a broad range of topics in cybersecurity (including privacy and useability)
- 2017 SaTC program: 139 awards
    - 11 education, 6 large, 35 medium, 67 small, 20 career
    - Budget year ends Sept 30, next year round about to begin
    - "Awards" means individual awards. Large projects may win multiple awards
    - Small: up to $500K over 3 years
        - "Sky is the limit" as to the topic
    - Medium, up to $1.2M over 4 years
        - usually multi-investigator or multi-institutional
        - Want to see something coming out of the award jointly from all collaborators, not just institutions working independently
    - Large: up to 3M over 5 years, Frontier: up to 10M over 5 years
        - Include several institutions. represent a specific focus that SaTC or NSF wants a specific focus or to move the needle in
    - Education: up to $500K over 3 years
- Other Opportunities
    - TPP (in conjunction with small/medium)
    - CRII and early CAREER awards
    - Partnerships with industry (e.g. Intel, STARSS)
    - EAGERs (early stage high-risk, high-reward)
    - Dear Colleague Letters (DCLs)
- For SaTC 2018
    - No large/frontier competition, all small mediums with a TPP option

    - No deadlines, proposals accepted as of October 1, 2018
    - Limit on number of proposals per PI/Co-PI
        - 1 small/medium
        - 1 TPP
        - 1 EDU
    - Broadening Participation in Computing (BPC)
        - Requirement that you do a BPC for all medium plans, will trickle down to all small plans eventually
        - Requires a “culture change”

SNF CyberCorps: Scholarship for Service (SFS) program

- Capacity building activities
- Examples: GenCyber and Women in Cyber Security

Discovering Users' Privacy Expectation

Chair: Daniel Susser
Notetaker: Priya Kumar

Context Matters: Guidance for Applying the Fair Information Practice Principles in the Internet of Things
Paula Bruening and Heather Patterson

Problem: How to apply longstanding principles of FIPPs in complex environments like IoT. FIPPS are a recognized, relied upon, and trusted guidance for data protection and basis for privacy law, "a common language for privacy." FIPPS are often used as a check box exercise, e.g., "we provide notice, choice, etc." Are they actually being used to support better privacy outcomes?

Intel's "Rethink Privacy" project proposed considering how FIPPs could be used as levers to be pushed and pulled and applied in a weighted fashion to address risks and circumstances. Many in Washington believed that the FIPPs should be abandoned because emerging technologies and data processing methods challenged the ability of companies to apply them. Intel believed that they remain relevant, that they continue to reflect prevailing values about data protection, and they facilitate interoperability between data protection regimes.

Research Question: Can Contextual Integrity be used to apply the FIPPs in emerging, complex, data rich tech environments.

How and why CI:

- Inserts discipline and predictability about how FIPPs should be applied.
- CI promotes decisions that align with prevailing social norms
- Is sufficiently flexible to adapt to changed norms when they evolved organically
- Motivates and accounts for questions in social settings, etc.
- Combine CI as a way to identify sources of unease, strain on social norms and FIPPs as concrete guidance to alleviate the unease and concerns around potential norm change

Current progress and results:

- Applied it to case study of how this would work in a given environment. Chose aging in place as supported in IoT. Didn't do empirical research; drew conclusions from reading mainstream literature. The point wasn't to understand this particular environment but to understand how to create a system/methodology.

- Aging in place is keeping adults at home, comfortable, safe etc. as long as possible, regardless of income or support systems. Most adults prefer this. 90 percent of adults over 65 say they want to stay in their homes as long as possible. Intel was working on IoT solutions. We saw how it could support aging in place, modifications to existing tech, creation of new tech, and analysis of new data. It also results in some new information flows (caregivers, family members, government agencies that support seniors, researchers, etc.)
- Create a template of questions an implementer would ask when thinking of how to implement the FIPPS.
    - Define the social context. Hybrid medical care and home care. Different but related kinds of systems.
    - Identify human values involved. In both contexts, the elderly want to feel confident they can divulge intimate details about health and life and confident that this information will be used responsibly, kept confidential, and not used inappropriately
    - Type of information, actors, and transmission principles at issue. Information includes medical info, blood pressure, pulse, other vital signs, meds, etc.
    - Other personal habit data including sleep patterns, exercise habits, hygiene habits, movement, taking care of small tasks
    - Community: who's coming to see the elderly person. What's that interaction like?
    - Actors: caregivers, family members, health care workers, people at the company level installing the system, government agencies, researchers
    - Transmission principles: patients feel comfortable sharing information but will withhold sometimes. They will disclose to some but not others. Patients felt they should have the discretion to do that.
- Due to IoT, more granular information is collected more often. The friction of a doctor's visit goes away with more ubiquitous sensing
- The IoT is also more pervasive, such that the ability to decide what to disclose or not begins to disappear.
    - What is the normative impact? More autonomy is possible. People could stay in home longer, get support they needed for longer, or delay moves to support facility. However, it means there's more observation, which can change the way people operated within home.
    - So what do we do with this info if we're going to apply the FIPPs?
- Collection limitation principle – what are the prevailing expectations of data collection? Where is the data coming from and what's being collected?
- Found that what created unease wasn't data collection, but the fact that it was collected by novel actors and potentially used in new ways.

Response: If collection isn't the issue, then it's important to shift the balance to other FIPPs to enhance openness, provide more notice, limit use of data, and enhance data security.

Challenges & Lessons:

- Multiple IoT systems in an environment may be deployed and operated by different entities. How should we deal with competing norms?
- There is always going to be a profit motivation for companies. How do you address this? Can you keep the company accountable? Can accountability principle in FIPPs help that?
- There is a complex constellation of actors. Who are the FIPPs protecting? This is an interesting question when trying to balance competing interests.

Future work:

- Need to test the approach in operational environment where sensors are actually deployed
- Similar case studies across other environment, education, home, library, workplace
- How should you approach CI based decision making when environment presents conflicting norms? How do you parse those different norms
- What are the practical tech and design questions that emerge when you try to do this? How can they help you do this implementation?
    - When we talk about privacy, we talk about privacy by design. Can you build something into systems?

Q&A

- How do you think about the relationship between CI and FIPPS? CI can be a way to implement FIPPS, but it also seems like an alternative.
    - *Response:* CI is an alternative to the FIPPS. From a public policy standpoint, discussions in Washington, Brussels, APEC still see FIPPS as an anchor for data governance. However, they don't look at the social side. They focus on the legal economic perspective. They don't look at the social norms. Using CI can help us insert some of the norms stuff in. Companies are starting to look more at things like ethics. What are the general expectations people have about the environment? Can you build that in?
- I'd like to hear more about what motivates Intel to do this study? We understand why we can make money off data, but why would they try to build those privacy social norms into the system?
    - *Response:* A lot of tech companies are trying to establish trust. They want people to feel comfortable using tech, having their data used. If they don't build in some privacy protections, some guardrails, they'll lose that trust, and people won't want to share data or use the technologies. Intel wants to get ahead of this. I don't work for them now, but they want to figure out creative ways that allow data

governance that allow them to innovate and use data robustly but also maintain trust of users.

Studying User Expectations about Data Collection and Use by In-Home Smart Devices
Julia Bernd, Serge Egelman, Maritza Johnson, Nathan Malkin, Franziska Roesner, Madiha Tabassum, and Primal Wijesekera

Problem: The goal with this talk is to pitch study ideas my group has been thinking about and solicit feedback. We've been looking at CI broadly to observe different data flows and do user studies to ask people what their expectations were. Once we know their expectations, it's a matter of deconstructing the flows into the different parameters so we see how those parameters affect expectations. Inferring context on mobile device is hard since we only have a view of what's going on on the mobile phone. So we use proxies. We might know where the data is going, allowing us to infer the purpose. If it's going to an ad company, probably used for an ad. We also look at what the app does, what's happening on the device, so we can infer the purpose of the data flow and reason people's expectations.

Progress and Results: We started looking at IoT devices with a large scale survey of smart TV owners. We investigated how they felt about different modalities, voice commands, gestures, where they think the data processing happens, who accesses data, etc. We found a huge amount of variance in what people expect. A lot of people are concerned about data begin used for other purposes, which makes sense in CI; context changes so their expectations change. They're also cynical; they're concerned, but they expect things to be happening regardless. Many people believe legal protections prevent egregious examples, but these protections don't exist.

Previously we did factorial vignette studies. We randomize different CI parameters, e.g., sender x shared data y with recipient z. This lets us figure out different ways these parameters matter to people and affect their perceptions about appropriateness. We looked at mobile devices and app permissions. We also looked at this for potential future wearable devices.

Those studies told us that people are a lot more ok sharing data with computers (e.g. for processing), rather than a human looking at it. People also care about multimedia (audio, video, photo) in home. They care less about data that's more or less publically available, e.g. demographics. This could be be because these attributes are already inferable by looking at them. So how does this apply to IoT devices? The continuous sensing devices that Nathan talked about are constantly monitoring audio, eventually maybe video, in order to services to the user.

Mapping this to CI is more complex. Sender is device. We know initially if we have tools on the local networks, that allows us to infer some things. Once it goes to a third party, who knows what happens. That's a policy issue. What's the data subject for in home devices? Not just device owner. It's also different for mobile device. It could be me, my wife, kid, or guest. A guest might also have different expectations than owner and might not be aware of data attributes or what kind of data is being collected.

On phone, there's somewhat strong data types and APIs that regulate location. How do you determine what data type is being shared with who? What transmission principles apply when this data is captured in the home? Many fewer proxies exist in this setting to determine context, especially because cloud processing is happening. Audio is recorded on device but sent to a remote server to determine what the request is about. Only after data goes to remote service can it determine whether info was sensitive and shouldn't have been recorded.

Planned Studies:

- What are contextual social norms?
    - People's expectations? How do these align with what the devices already do?
    - Planning to look at current users to see when users expect the device to be recording/not recording them, then seeing what's actually happening. Is it happening inappropriately? Is it not a big deal? What happens when the data leaves the device?
    - Both popular devices (Google Home and Amazon Alexa) have tools to let users see previous queries and you can also delete the data if you want.
        - Plan to develop Chrome and Firefox extensions to scrape these and then ask people about previous recordings. Did you know? Did you expect? How did you think it would be shared? How long do they think the company keeps this? Do they want to delete it?
    - First, ask what do they think happens to the audio? Is it deleted immediately, saved temporarily, gets saved indefinitely, or I don't know?
        - Then we're going to tell them the answer is indefinitely.
    - Then will play recordings and ask them who the recording is.
        - Will follow by asking whether the person speaking addressed the device, was the recording an accident, and do you remember making that recording, etc.
    - Then we'll use the extensions to let them delete if they want to
    - Will ask them what they think an ideal retention policy is
        - Is it acceptable to be analyzed by computer or human?
        - Is it more acceptable if audio is being analyzed or if the transcript is being analyzed?

- Sensitive data
    - Trying to build a corpus of annotated conversations to pull out features and train a classifier to detect future conversations where the device should shut off. If you were the speaker, would you expect this to be shared with various entities or could this be beneficial for the speaker based on what people label as sensitive?
    - Students doing this have found a bunch of existing datasets, especially from the speech community, NLP, and speech processing. To get more natural conversations across contexts, thought about pulling transcripts from reality TV. CBS all access allows you to download the archive of conversations. One idea is to have Turkers annotate the conversations then do future validation studies after building a classifier
- Other work is based on self reports about what users think they think they would do in reaction to proposed device behavior. However, we know this is different than people's actual behavior. Our goal is to give people devices and see how they behave. We know when the device is recording, and could do experience sampling using phones with a pop up asking about recent experience
- We could also give people a dummy device that says it might be recording, but it really isn't, and see how long it takes for them to unplug it.

Q&A

- We were really thinking about the five parameters. It's important to have them all. How do you ask these questions to make sure they're all included? Once you miss a parameter, the user starts to impose what's in their own head
    - *Response:* Yes, we need to think about this. Simply asking these questions in a vacuum, e.g., "do you care if the data is shared," results in most people caring. Need to specify the exact circumstances and brainstorm how to ask these questions correctly.
- How are parents consenting for their children to use these devices?
    - *Response:* We've been thinking about we would do studies in home and get informed consent from all parties in the house. The question is easier for kids, because parents are there and they can provide parental consent on behalf of children. The more difficult question is consent for transient people like guests. Are you going to ask them for consent every time they come over? What's the notice?
- As an educator, I often run a poll and ask students to pick one of five options. I show individual outcomes. Then I tell them to talk to each other and then rerun the poll to see how the answers changed. In the same way, one of the concerns I have is that these questions are typically asked of individuals, whereas expectations and norms are socially

constructed. You're asking people about decisions they make individually, but then they're talking about interactions with that device and others
    - *Response:* One of the fundamental problems is that norms shift. Certainly people's expectations are also informed by external forces. So asking how do you feel before they've had time to reflect is an issue. It's important to get a diverse set of participants, but also to do longitudinal work.
- We've talked about our work with Kirsten. We spent days trying to figure out how to ask the question. I have an idea for a methodology. So many surveys, many on MTurk, always ask individuals. We don't want to necessarily ask the individuals how they feel because that gets at their tastes or interests. Sometimes we also want to get a sense of the norm. I may not like something, but I can acknowledge that it is a norm. I only collaborate with social scientists who are familiar with this method. But is there a way to get a collaborative answer
    - *Response:* Focus groups. We conducted some focus groups.
- I'm also wondering about how the group sense of the norms are indexed to their understanding of the group. I wonder if our norms about privacy are indexed to friction around norms. Does that change as norms shift?
    - *Response:* That's why I brought up smart TV study. People's expectations are driven by not understanding tech and what it's capable of, but also by not understanding the protections. They think "these bad things could be happening, but there are laws to prevent that," but there are no laws on this.

Perceiving Patient Privacy in the Context of Heart-Failure Telemonitoring: Adapting the Contextual Integrity Framework to Gauge Patients' Privacy Perspectives
Martin French and Heather Patterson

Problem: This paper came out of collaborations with Northwestern and Berkeley. They were working on how effective mobile phones could be to monitor people who had been hospitalized for heart failure and discharged. Their study gave patients a mobile device and tracked accelerometer and GPS data to see when they're entering periods of low activity. They then could intervene to see how they're doing so they aren't re-hospitalized. We joined to see how patients might perceive the privacy implications of this tele-monitoring.

We reviewed literature on surveys of people's perceptions of health info and medical privacy. Classic examples include a Westin Harris Equifax survey from 1993 which asks 4 questions, e.g. "It concerns me that my medical info is being seeing today by many organization beyond those that I go to for health care services. Do you agree or disagree?" 13 percent of people are high sensitivity. 45 percent medium, and 42 percent low sensitivity.

This is a great starting point. But what does it mean to have high sensitivity to questions of medical privacy? Here's where we thought CI would be useful.

- How comfortable are you with your doctor getting regular updates about your weight. Are you comfortable or uncomfortable? Replace doctor with insurance company.
- How comfortable are you with your doctor getting regular updates about your alcohol and tobacco use?

We spent a lot of time thinking about conceptualizing the different types of information transmitted via tele-monitoring to potential recipients under different transmission principles. We Came up with lots of variables inspired by factorial vignette studies. We came up with 48 questions and integrated them into a pre/post survey that would be administered to patients. We struggled to come up with what configurations would work.

- Pilot tested various configurations to check for face validity. We spoke with health care team. They said no way, the cognitive load is too high. You're working with patients who have been hospitalized, they might be heavily medicated, in pain, etc. There were lots of challenges we had to deal with to come up with questionnaire.
- Came up with series of 35 questions involving doctors and nurses who treat your heart failure getting updates on your weight, alcohol, etc.

Takeaways: it's difficult to ask chronic heart failure patients about contextual integrity of their health info. How should health care systems care for the CI of their patients' info? Are hospitals doing enough to provide info counseling, info advocacy, ombudsperson to patients who might be asked to participate in new tech programs to care for their health but that raise questions about privacy?

Q&A

- Did you think about the attitudes of patients in terms of transferring information being an expression of their relationship with the nurses and doctors rather than their concerns about privacy or their feelings about being so ill?
    - *Response:* No, our instrument wouldn't allow us to tease that out without in-depth interviews. Hard to disentangle views given trust or distrust for their particular doctor.
- Do you think this kind of system could apply to patients with other types of chronic illnesses? Could it also provide educational opportunity for patients to better understand how to facilitate their care?
    - *Response:* Yes, I hope so. We could think about using an instrument like this more broadly for other health information flows. While working with Helen, we did some work with colleagues at UIUC on decision support systems for data

segmentation for decision support, an inquiry into whether health information exchanges could exchange health records in a way that protects privacy by suppressing certain elements of those records. It's hard to sit down and discuss them with patients. If you suppress this part of the record and share with this recipient, there's X probability that recipient can infer the missing data. If we just hold patients responsible for making those decisions, that's not OK.

- Nice inclusion of granularity. It seems like when we're trying to understand certain transmission principles, how amount of information vs. the type of information isn't something we talk about as much. I wouldn't mind my doctor having enough information as they need. I don't want them to have more information than they need. Did you test people on how much information they wanted to give?
    - We did. We also had to drop some of the levels of granularity when going from 48 to 35 questions. We would have to look at the data to see what we found. When speaking with tech designers, folks who were programming the software, it was interesting to think about them getting so much data they can't handle it. We had conversations about purpose - granularity is one way to operationalize transmission principles, but there are others.

KEYNOTE: Understanding Privacy and Contextual Integrity: A Personal Journey

Speaker: Anupam Datta
Notetaker: Noah Apthorpe

Personal history of thinking about contextual integrity and privacy:

- Began with PORTIA (Privacy, Obligations, and Rights in Technologies of Information Assessment) (large NSF funded center with Helen as one of the lead PIs)
- Anupam was finishing a Ph.D. at Stanford at the time, and Helen's talk was a pivotal moment in his research career
- In original Washington Law Review article, CI had 2 types of norms, this was eventually reduced to 1 type of norm for the CI book and later iterations
- 2006 IEEE S&P paper Barth et al. was first attempt to bring CI from law to CS
- Originally focused on descriptive component of contextual integrity (i.e. norms of flow)
    - Wanted to see to what extent this structure gets reflected in privacy regulation
        - GLB Act does talk about senders, recipients, attributes, subjects, and TPs
    - Also formalized the descriptive component of CI in formal (first-order temporal) logic
        - demonstrated that sample clauses from the regulation lined up with the formal logic specification
- Enforcing Privacy:
    - contextual integrity norms ← → U.S. sectorial privacy laws
    - U.S. sectorial privacy laws ← → formalized privacy policies
    - A monitoring engine that could take the formalized privacy laws and programs/logs/datasets and see whether there was a violation
- Can we specify the entirety of these laws using the formal logic?
    - At CMU, Anupam's research group worked on translating HIPAA into the CI formal logic.
        - Found that the structure of the law largely follows CI flow descriptions
        - However, restrictions of use of personal information for specific purposes were outside the scope of the CI flow-based norms
            - Does this mean that CI needs to be extended to cover use restrictions?
            - Are non-flow use restrictions/cases really privacy harms? Can they instead be recast as flows (e.g. flows with a different TP)?
            - Use restrictions also come into fairness concerns in addition to privacy concerns
        - Some other clauses from HIPAA were difficult to translate directly

        - References to “purposes” and “beliefs”
            - These draw a distinction between black-and-white concepts and grey concepts (the latter of which are less amenable to enforcement)
- Next looked at policy auditing over incomplete logs
    - Moved to first order logic (instead of propositional logic as before)
    - Had to deal with incompleteness of the logs
    - Reduce: an iterative algorithm for taking the log and the policy and output a residual log that could be determined whether it satisfied the policy only when additional information is revealed
        - Allows incremental auditing, but still required an “oracle” to answer some questions, e.g. did the doctors “believe” some information
- Bootstrapping privacy compliance in big data systems
    - Came out of conversations and student internships at Microsoft Research
    - Microsoft was facing the problem that privacy policies are written in English, but applications are written in code. In order to check that the code was matching the privacy policy, they needed experts embedded in teams. Wanted to make this much more automated
    - Created a toolchain with 2 technical components
        - Legal*ease* - a formal policy specification language that could be used by the legal team that was formally well defined for use and automated checking
        - Grok - a data inventory with policy labels
            - Marks data with “attribute” labels - important because there was a wide gap between attributes as described in the English policy and the actual data types and variable names in the code.
            - Annotations performed automatically using data points of entry. Once manual annotations performed on roughly 20% of entry pipes, you could propagate these annotations to nearly all of the code/database.
        - Grok and Legal*ease* work together to identify places in the code that the audit team should manually check to verify compliance
    - However, the privacy policies in use may be quite far from what CI norms would expect
        - This is an opportunity for further research

Questions relevant to CI

- What is the “type” (or topic) of a piece of data?
    - Some are simplistic - e.g. email address is an email address

    - However, it is hard to say whether data does **not** contain information about a topic, especially given progress with ML inference
- Is it useful to have incomplete enforcement?
    - Whenever privacy policies refer to a type of information, it is very hard to do so in a way that is *complete.*
    - Should we remove all dependence on semantics of data types?
        - This is how other theories of privacy deal with this, e.g. differential privacy does not consider the type of information, nor does origin privacy
        - It is possible to do this partially with HIPAA, i.e. remove all information types and replace with just a data origin.
        - It would completely fail for something like GDPR which has a lot of dependence on data types

Use Privacy in Data-Driven Systems

- Example: the Target pregnancy inference case (2012), Google sleep apnea case (2013-14)
- Again, use restrictions are a different category of statements outside of the CI flow-based norms

Summary

- Contextual Integrity is an immensely important piece of the privacy puzzle

Challenges and opportunities

- What is the "type" or topic of a piece of data
    - is it useful to have an incomplete statement?
    - should we remove all dependence on semantics of data types?
- What does it mean to "use" a type of data?
    - Normative theory of use privacy (in addition to epistemic flow-based privacy)
    - Operationalizing use privacy for data-driven systems
        - Especially important for systems using machine learning models on data
- What does "purpose" mean and how do we enforce purpose restrictions?
    - Could show that an action is part of a plan for achieving a purpose
        - "action" and "plan" are both formalized in ML planning research
    - initial work in Tschantz et al.
- Deploy in production systems
    - Due to externalities like GDPR, we have the interest of industry, making this a good time to work on these problems

Q&A

- Although it removes type information, differential privacy does not protect against semantic meaning encoded in the dataset itself or derived by ML on the data (even if it wouldn't change much by including a participant or not)
    - *Response:* Agreed, but removing type information still makes it easier to get "complete" information flows, even if this is not possible for all applications or for all policies
- Could you encode use restrictions as transmission principles and therefore changes in use cases could potentially create new "previous" flows that were not consented to
    - *Response:* It's certainly possible to construct a formal logic based on CI this way, the question is whether it's in keeping with the philosophy of the framework
- If an automated approach to limiting data flows or data of a certain type is incomplete, would it be possible to use a societal method, e.g. airgapping to provide the needed guarantees?

Wrap-Up Discussion

Marshini Chetty, Helen Nissenbaum, Yan Shvartzshnaider
Notetaker: Noah Apthorpe

Comments from audience:

- Would like collaboration to see if there is a way to use CI to set the epsilon parameter for differential privacy
- Thought that the symposium was a unique valuable experience compared to other privacy conferences. Something like this should happen again!
- Privacy in the home has become much more central, recalling the original Washington Law Review article that used public/private distinction as a major example. Perhaps this is a topic area to extend/focus on for future meetings
- The policy/privacy group at UC Irvine has a shared folder for slides, teaching materials, etc. Something similar would potentially be useful
- What did the organizers have in mind when they created the event?
    - *Response:* Essentially this! The primary goal was to bring people all working on CI together and use discussion to push CI and privacy forward
- It was a great event! Would be nice to include more people from industry. There is a huge demand for solutions that automate all parts of this puzzle. The work that Anupam discussed was very valuable for Microsoft. With the forcing function of GDPR, a lot of companies will be attempting solutions. It would be a shame if they didn't involve context in some form. Since people are building it anyway, we should include industry even if we're still treating CI as a work in progress
    - *Response:* A lot of CI is the merge of roles in society that have many functions that have developed over thousands of years (e.g. physicians). We now have companies that are sitting on massive amounts of data. A lot of CI comes not just from the meaning of the data, but also the intention of the actors. In the upcoming years/centuries, we need to focus on purpose/use because the status quo thinking is inadequate
    - *Response:* We would love to have more industry people involved. This symposium came together as a result of the advertising and connections in place. Will do more advertising outreach to companies in the future
- It was nice to see some undergraduate students here! Would be nice to see more detailed/structured discussion of specific panels of presentations beyond just the short Q&A after each talk.
    - *Response:* It would be great to have a poster session. This would help address the previous point

Appendices

Symposium Chairs

- Marshini Chetty (Princeton University)
- Helen Nissenbaum (Cornell Tech)
- Yan Shvartzshnaider (NYU and Princeton University)

Program Committee

- Sebastian Benthall (University of California at Berkeley, Cornell Tech)
- Anupam Datta (CMU)
- Serge Egelman (University of California, Berkeley)
- Nick Feamster (Princeton University)
- Seda Gürses (University of Leuven)
- Darakhshan Mir (Bucknell University)
- Julia Powles (Cornell Tech and NYU)
- Xinru Page (University of California, Irvine)
- Jessica Vitak (University of Maryland)
- Pam Wisnieski (University of Central Florida)
- Thomas Wies (NYU)

Contact

- privaci.research@gmail.com

Attendees

- Alison Cooper, Cornell University
- Arunesh Mathur, Princeton University
- Noah Apthorpe, Princeton University
- Ben Zevenbergen, Princeton University
- Catherine Dwyer, Pace University
- Darakhshan Mir, Bucknell University
- Anupam Datta, CMU
- Deni Chan, Cornell Tech
- Dennis Redeker, NYU (CEMS)
- Dmitry Epstein, Jerusalem University
- Dylan Rogers, Bucknell University
- Ed Felten, Princeton University
- Edo Roth, University of Pennsylvania
- Serge Egelman, UC Berkeley / ICSI
- Eleanor Birrell, Pomona College
- Eran Toch, Tel Aviv University
- Elizabeth O'Neill, Cornell Tech / Eindhoven University of Technology
- Erica Du, Cornell Tech
- Fabian Okeke, Cornell Tech
- George Pappachen, NYU Stern
- Griffin Berlstein, Vassar College
- Hans Klein, Princeton University / Georgia Tech
- Helen Nissenbaum, Cornell Tech
- Jacopo Arpetti, University of Rome Tor Vergata
- Jason Waterman, Vassar College
- Julia Bernd, UC Berkeley / ICSI
- Jeffrey Friedberg, Microsoft
- Jaulie Goe, NYU
- Jake Goldenfein, Cornell Tech
- Jessica Vitak, University of Maryland
- Karla Badillo-Urquiola, University of Central Florida
- Kelly Quinn, University of Illinois at Chicago
- Yafit Lev-Aretz, CUNY Baruch College
- Lauren van Haaften-Schick, Cornell University
- Laura Garcia, University of Ottawa
- Luke Stark, Microsoft Research Montreal
- Mainack Mondal, University of Chicago
- Marshini Chetty, Princeton University
- Martin French, Concordia University
- Martin Kraemer, University of Oxford
- Michael Sobolev, Cornell Tech
- Michael Byrne, Cornell Tech
- Madelyn Sanfilippo, NYU
- Nathan Malkin, UC Berkeley
- Nora McDonald, Drexel University
- Pamela Wisniewski, University of Central Florida
- Patrycja Sleboda, Cornell Tech
- Paula Bruening, Sequel Technology / IP Law LLC
- Priya Kumar, University of Maryland
- Primal Wijesekera, UC Berkeley / ICSI
- Rachel Cummings, Georgia Tech
- Rafa Gálvez, KU Leuven
- Marijn Sax, Cornell Tech / University of Amsterdam
- Sebastian Benthall, NYU
- Susanne Wetzel, National Science Foundation
- Nirvan Tyagi, Cornell Tech
- Xinru Page, Bentley University
- Yuan Wang, Cornell Tech
- Yan Shvartzshnaider, NYU / Princeton University
- Danny Huang, Princeton University
- Germana Volpi, Italian Communications Regulatory Authority
- Kathryn Kleiman, Princeton University
- Annette Zimmermann, Princeton University
- Daniel Susser, Penn State University.
- Michael Sobolev, Cornell Tech
- Patrycja Sleboda, Cornell Tech

Program

Thursday, September 13

2:00pm **Registration**
2:30pm **Welcome**

2:45-3:45pm **CI and Society #1**, *Chair: Jake Goldenfein*

- Contextual Integrity as Commons Governance in Online Political Organizing, *Madelyn Rose Sanfilippo and Katherine Strandburg*
- Knowing and believing: Privacy literacy, privacy self-efficacy and context in privacy-protecting behaviors, *Dmitry Epstein and Kelly Quinn*
- Situated Information Flow, *Sebastian Benthall*
- Privacy and Religious Views, *Madelyn Rose Sanfilippo and Yafit Lev-Aretz*

4:00-5:00pm **CI and Society #2,** *Chair: Ben Zevenbergen*

- Applying Contextual Integrity to the Cambridge Analytica Case, *Catherine Dwyer*
- Analyzing Privacy Policies Using Contextual Integrity Annotations, *Yan Shvartzshnaider, Noah Apthorpe, Nick Feamster and Helen Nissenbaum*
- Enforcing Contextual Integrity With Exposure Control, *Mainack Mondal and Blase Ur*

5:15-6:00pm **PrivaCI Challenge**
6:30pm **Discussion over Dinner**

Friday, September 14

9:00am **Coffee and Refreshments**

9:30-10:20am **CI in Smart Homes and IoT,** *Chair: Güneş Acar*

- Disentangling Privacy in Smart Homes, *Martin J Kraemer and Ivan Flechais*
- On Engineering AI Agents for Privacy, *Rafa Gálvez and Seda Gürses*
- Use Case: Passively Listening Personal Assistants, *Nathan Malkin, Primal Wijesekera, Serge Egelman, David Wagner*

10:30-11:20am **CI and HCI,** *Chair: Michael Byrne*

- Contextual Integrity as a Conceptual, Analytical, and Educational Tool for Research, *Priya Kumar*

- The Emotional Context Of Information Privacy, *Luke Stark*
- Literature Review: Examining Contextual Integrity within Human-Computer Interaction, *Karla Badillo-Urquiola, Xinru Page, Pamela Wisniewski*

11:45-12:30pm **CI based Systems Design,** *Chair: Darakhshan Mir*

- Contextual Permission Models for Better Privacy Protection, *Primal Wijesekera, Joel Reardon, Irwin Reyes, Lynn Tsai, Jung-Wei Chen, Nathan Good, David Wagner, Konstantin Beznosov, and Serge Egelman*
- Privacy Heuristics: How Users Manage Complex Contexts when Evaluating Mobile App Requests, *Amit Rechavi, Eran Toch, Jason Hong*
- Privacy-Aware Programming for Microscale Data, *Jason Waterman, Eugene Bagdasaryan, Matthew Griffith, Griffin Berlstein, Nate Foster, Eleanor Birrell*

12:30-2pm **Lunch and Conversation with Susanne Wetzel, Program Director SaTC, NSF**

2:00-2:55pm **Discovering Users' Privacy Expectation,** *Chair: Daniel Susser*

- Context Matters: Guidance for Applying the Fair Information Practice Principles in the Internet of Things, *Paula Bruening and Heather Patterson*
- Studying User Expectations about Data Collection and Use by In-Home Smart Devices, *Julia Bernd, Serge Egelman, Maritza Johnson, Nathan Malkin, Franziska Roesner, Madiha Tabassum, and Primal Wijesekera*
- Perceiving Patient Privacy in the Context of Heart-Failure Telemonitoring: Adapting the Contextual Integrity Framework to Gauge Patients' Privacy Perspectives, *Martin French and Heather Patterson*

3:15-4:00pm **Understanding Privacy and Contextual Integrity: A Personal Journey**
Anupam Datta

4:00-5:00pm **Spillover, Discussion, and Wrap up,**
Marshini Chetty, Helen Nissenbaum, Yan Shvartzshnaider