The 5th Annual Symposium on Applications Of Contextual Integrity

The report was compiled by Yan Shvartzshnaider based on the notes taken by Arthur Borem, Benjamin Laufer, Collins W. Munyendo, Dorothy Ko, Elleen Pan Errol Francis II, James Bailie, Johanna Gunawan, John Paul Schnabel, Kyra Milan Abrams Lauren Ministero, Lukas Seiling, Madison Pickering, Nikita Samarin. Sohyeon Hwang,Tamalika Mukherjee

Special thanks to our spnsors: National Science Foundation, Lassonde School of Engineering, York University, Microsoft and Digital Life Initiative

Symposium Chairs

Marshini Chetty (University of Chicago)
Helen Nissenbaum (Cornell Tech)
Yan Shvartzshnaider (York University)

Program Committee

Noah Apthorpe (Colgate University)
Louise Barkhuus (The IT University of Copenhagen)
Sebastian Benthall (New York University)
Jorge Pereira Campos (Leiden University)
Ignacio Cofone (McGill University)
Rachel Cummings (Columbia University)
Cathy Dwyer (Pace Univertisy)
Serge Egelman (ICSI & UC, Berkeley)
Yafit Lev-Aretz (Zicklin School of Business, Baruch College)
Priya Kumar (Pennsylvania State University)
Kirsten Martin (University of Notre Dame)
Lee James McGuigan (University of North Carolina at Chapel Hill)
Mainack Mondal (IIT Kharagpur)
Madelyn Sanfilippo (University of Illinois at Urbana-Champaign)
Ido Sivan-Sevilla (University of Maryland)
Luke Stark (Western University)
Daniel Susser (Penn State University)
Eran Toch (Tel Aviv University)
Salomé Viljoen (University of Michigan Law School)
Jessica Vitak (University of Maryland)
Primal Wijesekera (ICSI)
Michael Zimmer (Marquette University)

Executive Summary

In late September 2023, the 5th symposium on application integrity took place at York University's Lassonde School of Engineering in Toronto, Canada. The CI symposium brought together students, postdoctoral fellows, faculty, and practitioners from Europe, the United States, and Canada to present and discuss early-stage and published work related to contextual integrity (CI), as well as to design, evaluate, and generate ideas.

The symposium included (9) paper and use-case discussion sessions, (2) CI Community Feedback sessions, and a mentors' lunch. The paper and use-case panel discussions covered topics related to investigating privacy challenges, exploring governance and norms, health, differential privacy, tracking, and ad tech. Session 1 discussed privacy challenges in VR classrooms and privacy concerns related to the use of technology by older Canadian adults. Session 2 delved into privacy concerns related to decentralized social media platforms and scholarly infrastructures. Following the first two sessions, the CI Community Feedback session focused on using CI to explore privacy in a crowdsourced gig work knowledge-sharing platform, blockchain technologies, and AI-powered metaverse. Session 3 examined using CI to model Perceived Privacy Risk with Contextual Integrity for Risk Communication and also included a discussion on the proposed expansion of the CI theory to include different types of information flows.

On the second day, Session 4 panels discussed a use case of a Human Rights-Based Approach to the Governance of Algorithmic, Biometric, and Smart City Technologies, as well as Reconciling Polycentric Governance via Data Protection Addenda. The second CI Community Feedback session discussed Digital Redlining; using Contextual Integrity to explore privacy perceptions of low-SES users; Choice Architecture and Contextual Integrity in Privacy Decision Making. Session 5 included presentations on Remote Healthcare Technology Use Cases and the Contextual Integrity of Older Adult User Privacy use case; Applying Contextual Integrity to the Development of Polypharmacy Indicators in Canada and measures to protect health data and promotion of health data for research purposes. Session 6 discussed CI applications and methods with works on Automating GKC-CI Privacy Policy Annotations with LLMs; Applying Contextual Integrity to Elicit Acceptance towards COVID Mitigation Mobile Applications in the US and an experiment of applying Contextual Integrity in the context of wearables. Session 7 included proposals on using CI for norm discovery and regulation. Session 8 looked at integrating CI with other privacy-preserving methods like Differential Privacy and the Five Safes framework. The final session focused on Tracking and AdTech with works on Using Contextual Integrity to Determine Permissible Secondary Data Uses and the work on First-Party vs Third-Party Privacy.

Session 1: CI and Privacy Challenges

Privacy Challenges in VR Classrooms: A CI Use Case

Karoline Brehm (Bauhaus-Universität Weimar), Yan Shvartzshnaider (York University), David Goedicke (Cornell Tech)

The paper addresses the question of understanding the privacy considerations arising in virtual – and virtual reality – classrooms. Virtual classrooms have particular privacy challenges since students (and teachers) in a virtual classroom are simultaneously located in multiple environments. Students are physically located in, for example, their home, while they are virtually located in the classroom. Since information flows between these multiple environments (for example, a student talking in class will also be heard by their housemates), there are novel privacy challenges associated with virtual classrooms.

This presentation aims to investigate the privacy norms that arise in the context of virtual, and virtual reality, classrooms. The theory of CI will be used to capture context-appropriate flows of information in situations where actors are simultaneously in multiple environments. The authors plan to visualise these information flows and use their results to inform the design, policy, and regulation of virtual classrooms.

Contextualizing Privacy for Older Adults in Canada

Paola Marmorato (Carleton University), Ruchi Swami (Carleton University), Sanchita Kamath (Manipal Academy of Higher Education), Nadila Asikaer (University of Waterloo), Elizabeth Stobert (Carleton University), Heather Molyneaux (National Research Council of Canada), Cosmin Munteanu (University of Waterloo)

As Canada's population ages and as technology becomes more ubiquitous in society, it is increasingly important to understand the privacy concerns of older adults in Canada. Support for this group is particularly needed because older adults tend to have higher privacy concerns while simultaneously having lower efficacy in privacy Management.

This paper contributes to addressing these issues by conducting a survey on older Canadian adults' privacy concerns relating to their use of technology. The theory of contextual integrity was used to design and analyse the survey. Survey respondents were also assigned a privacy attitude according to the Westin segmentation index. The authors found that older adults are more concerned that they may be exploited by private – rather than public – organizations.

There were difficulties in recruiting study participants, particularly from minority populations. In future work, the authors plan to collect more data by increasing their survey sample size and by augmenting the survey questions. In this way, they plan to obtain more in-depth responses on a set of key topics identified from the initial survey's results.

In the session's discussion, conference participants raised several pertinent points. Firstly, it is very difficult to design survey questions so that they are commonly understood and elicit true responses without being led. Secondly, the multitude of simultaneous environments, as induced by technology, have been posing privacy conundrums for a long time, even before the existence of virtual classrooms. For example, using a mobile phone in a coffee shop creates its own privacy issues.

Finally, there can be a clash between informational and legal norms, especially when the virtual and physical environments are governed by different jurisdictions. If, for example, an EU student wants to exercise their right to be forgotten, how does a Canadian lecturer balance this with the informational norms of their classroom? It is an open question of how to deal with clashes between these different types of norms.

Session 2: CI and Platforms

Whose Policy? Privacy Challenges of Decentralized Platforms

Sohyeon Hwang (Northwestern University), Priyanka Nanayakkara (Northwestern University), Yan Shvartzshnaider (York University)

In this paper session, the first was presented by Sohyeon Hwang for work conducted with Priyanka Nanayakkara and Yan Shvartzshnaider. This paper was motivated by the recent moves from Twitter/X to decentralized platforms and sought to understand how privacy and data flows between decentralized servers might be framed within CI (in particular, how self-governance in open-source, decentralized platforms might operate). Hwang sampled 803 Mastodon servers to find that over 80% use the same default policy template with the remaining ~20% using other cookie-cutter policies or vague policy text, and discussed some of the issues with contextual integrity violations at between-server interstices.

Some challenges encountered in this work include questions of how federated protocols span across servers with regards to CI, and whether this type of regulation is more of a data transmission or governance question. For future work, the authors plan to collect user perspectives through a recently approved IRB study and investigate policy and tooling designs further. In particular, they play to talk to server owners using policies as artifacts to discuss in the study.

CI + Scholar Infrastructure

Madiha Zahrah Choksi (Cornell Tech), Jake Chanenson (University of Chicago)

In the second paper, early-stage conceptual work regarding scholar infrastructures, Madiha Choksi presented work with Jake Chanenson. This draft aims to address concerns of scientific disillusionment and integrity/liability in research in the wake of AI and the prevalence of language-learning models. The authors discuss the challenges in trying to maintain the integrity of facts and accountability in knowledge-producing spaces, in particular highlighting the potential implications for all social spaces (e.g. including politics). With regards to CI, the authors plan to identify contextual parameters for data flows in this space, in particular, understanding what truth values might be and how CI might be implemented as a framework both to understand these truth values as well as protect them towards transparency and accountability in AI algorithms. As early-stage work, the authors are currently working on scoping the CI parameters and working to understand how verification mechanisms are developed, and continue asking (for future work) how AI systems might comply with CI norms, and how privacy and transparency might be balanced in AI tools (the authors suggest federated learning as one approach).

In Q&A, discussions centered around how end users might be able to understand (and trust) potential verification systems, how to truly conceptualize a truth value, and how to scope AI beyond LLMs and to other models or deployments.

CI Community Feedback Session #1

Privacy & contextual integrity in a crowdsourced gig work knowledge sharing platform

Arthur Borem (University of Chicago), Elleen Pan (University of Chicago)

This work focuses on the problem of preserving gig worker privacy and autonomy. The authors note that the gig economy is growing, with a rising number of workers playing an increasingly large role in the economy of highly developed nations. However, the Gig Economy platforms themselves are largely in control of workers' time and behavior.

The authors seek to build an information-sharing app to increase transparency among workers and to provide insights into how platform decisions are made. The app will be facilitated by workers uploading their personal data, as obtained through data access

requests. The app will answer questions like, "*Am I compensated fairly with respect to other employees?*", "*What does the platform know about me that I may be surprised by?*", and "*How am I at risk by participating in this platform?*". The authors seek to utilize CI theory to examine the data flows from gig workers to the platform, as well as from gig workers (whom the presenters hope to interview) to the presenters themselves.

That being said, the authors are not fully certain about how to allow for the creation of their app in a way that would not violate CI. In particular, the data flow between the presenters and participants may put the gig workers at risk of re-identification. And, the data flow between the participants and their platforms may result in malicious use by the platform. The majority of the Q&A touched on questions surrounding the data flows and key actors accordingly. Namely, the platforms of interest are all major gigwork platforms, including Lyft, Uber, and Taskrabbit, but excluding care work. The researcher's app was originally ideated to run in parallel to a gigworking app, but will most likely be moved to a dashboard due to practical constraints. The session closed with a comment that visualizing complex data flows is an unsolved issue, which may result in practical difficulties when constructing the dashboard.

Interaction between decentralization and contextual integrity: a focus on blockchain technologies

Fabien Lechevalier (Paris-Saclay University/Laval University)

The problems the authors focus on are that (1) there exists third-party centralization and (2) many of these third parties are forming contracts through approval of GDPR cookies, regardless of whether CI is being violated. The potential violation of CI raises issues of CI/data confidentiality and data sovereignty. To combat this issue, the authors look to blockchain. Blockchain can help capture information flows without paternalism; an individual has self-ownership of their data.

However, there are a number of issues raised when attempting to apply CI to blockchain. Namely, self-ownership may conflict with the contextual understanding of privacy. There are furthermore open issues relating to control in contextual and relational frameworks. For example,) it is unclear how CI factors into trust, which is a problem because it is difficult to apply CI in a trustless environment. Further, it is also unclear how to manage transparency while still providing cryptography with a large number of peers.

The Matrix of Privacy: Data Infrastructure in the AI-Powered Metaverse

Argyri Panezi (UNB Law), Nizan Geslevich Packin (CUNY), Leon Anidjar (IE Law School)

The talk is primarily focused on posing a question: how can we maintain current norms in Web 3.0/Metaverse? This is motivated by the observation that Web 3.0/semantic web/VR has drastically changed the ways humans interact with technology. Furthermore, there are more and more data-hungry technologies now in place. The presenters observe that data is no longer a commodity, but rather part of the infrastructure. Data serves as a fundamental building block for applications, a shared resource for seamless experiences, and a source of insights. The authors argue that this is a significant conceptual departure from previous iterations of the web. Furthermore, this innovation has similarly resulted in innovative harm. For example, likeness and identity are now more easily appropriated, resulting in more frequent identity theft.

Some questions raised where about how the existence of multiple metaverses plays into the above. The authors note that they consider only a hypothetical single, unifying metaverse. The authors further note that their motivation for using CI is to better understand complex interactions. In particular, the authors feel that the law in its current state may not be well-equipped to handle the rapidly changing nature of technology, but CI provides useful guidance for how the law may need to change to preserve individual privacy and interests

Session 3: CI and Theory Session

Modeling Perceived Privacy Risk with CI for Risk Communication

Lukas Seiling (Weizenbaum Institute)

The purpose of this talk is to understand how we should communicate privacy risks to data subjects. The main motivation behind modeling privacy risk is the issue of Informed Consent in the GDPR context. GDPR mandates that people are given appropriate notice of how their data is being used. But people must first read privacy notices which can be long and

cumbersome, and second, they must understand what the notice means. Thus there is a need for making helpful abstractions of privacy policy.

Risk communication is useful when risks are negative consequences resulting from specific events. These events or consequences happen within certain contexts. This talk combines the perceived risk model with CI and conducts expert interviews to understand how to link the consequences of the events. People experience consequences in a specific context. There are Tangible and Latent consequences that can be immaterial or material.

The authors propose to communicate the harms of these consequences to users to inform them of the risk. Methods include designing a browser plugin that gives understandable icons that convey privacy policy to end users.

Expanding Contextual Integrity: Three Types of Information Flows

Cara Bloom (MITRE), Lauren Ministero (MITRE)

The talk argues for adding new information flows to the classical definition of CI -

1. Information Generation (Subject, Data Type, Generator, Generation Principle)
2. Information Derivation (Deriver, Subject, Derivation Principle, Lower Order/Higher Order Data Type)

The new CI norms would include: Data Subject, Actors (Generator, Sender, Receiver, Deriver), Data Types, and Transmission Principles

The motivation behind the proposal is based on the notion that there are implicit norms about data generation and data derivation that are not captured by the classical definition, this is why the definition of CI should be expanded. Additionally, this is easier to communicate with people not well-versed in CI.

Session 4: CI and Governance

The talks in this session centered broadly around governance, highlighting how governance is context-dependent and often administered through decentralized points of decision-making, rather than solely through a centralized source.

Future-Proofing the City: A Human Rights-Based Approach to the Governance of Algorithmic, Biometric, and Smart City Technologies

Anna Artyushina (Toronto Metropolitan University)

In the first talk ("Future-Proofing the City: A Human Rights-Based Approach to the Governance of Algorithmic, Biometric, and Smart City Technologies"), Anna Artyushina (Toronto Metropolitan University) and Alina Wernick (University of Helsinki) present a use case that argues how governance frameworks do not work universally and must be contextualized. Focusing on human rights in the smart city context, the authors underscore how frameworks like the Human Rights-Based Approach framework fail at community-level implementations due to variations in norms and values. The HRBA framework also has distinct limitations, such as neglecting collective dimensions and structural inequalities that contribute to issues. As a result, it's important to co-evolve our frameworks and approaches with technologies, while working towards value-driven technologies (such as having HRBA by design in technologies). There are many strategies like leveraging existing civic and institutional tools or focusing on community needs and risks that can aid this effort.

Reconciling Polycentric Governance via Data Protection Addenda

Madelyn Sanfilippo (University of Illinois at Urbana-Champaign)

The talk examines policy documents to discuss how policy gets implemented in practice across various organizations and groups with different norms and values. Leveraging the concept of polycentricity (a system with many centers of decision-making, with overlapping spheres of influence), Sanfilippo notes how constellations of decision-making apparatuses interface, and how they can demonstrate redundancy, or sometimes yield conflicting decisions. This is examined in an analysis of institutional texts for data protection in the EdTech sector that leverages the GKC-CI framework described in earlier work to examine how institutions handle information flows. Some core takeaways of this work include the use of unenforceable language as well as the value of such texts to understand the institutional and legal contexts these organizations operate within and around. As a result, it's valuable to consider how privacy and data practices vary not only by time and place, but also by sector.

CI Community Feedback Session #2

The second CI community feedback session included a discussion on three early-stage works.

Digital Redlining: Redlining in a New Era

Kyra Abrams (University of Illinois at Urbana-Champaign)

The work examines how modern technologies contribute to discriminatory practices in areas such as health, finance, and housing. The presentation discussed a case study on the redlining practices of Facebook Ad platforms that allowed employment and housing agencies to discriminate against women and elderly workers. The focus of the study is on comparing the motivation behind modern digital redlining with the ones before the introduction of modern technologies. The authors would like to use CI to examine whether digital redlining violates contextual integrity.

Using Contextual Integrity to Explore Privacy Perceptions of low-SES users

Sana Maqsood (York University)

The work proposes to examine privacy perceptions among low Socioeconomic Status (SES) communities when using technologies in various social contexts, such as home and work. Addressing privacy in low-SES communities is crucial due to common practices of device sharing and heightened tensions resulting from increased government surveillance practices, where traditionally expected norms might not apply. To better understand the existing contextual norms and expectations governing the use of mobile devices in these communities, the authors plan to conduct CI-based semi-structured interviews with 205 low-SES individuals in Canada.

Choice Architecture and Contextual Integrity in Privacy Decision Making

Lauren Ministero (The MITRE Corporation), Jonas Ludwig (Tel Aviv University)

The work aims to investigate the CI-based methods to better understand how choices in the framing of questions might affect the resulting privacy norms when using vignette type surveys. The authors propose developing an experimental framework–choice experiments assess the language in composing survey questions used to capture the appropriateness of information flows. The proposed framework would also help quantify the effects of each CI parameter using statistical modeling.

Session 5: CI and Health Summary

The presentations in the session touched on privacy and ethical considerations in healthcare technology, the development of healthcare indicators, and the balance between data protection and data sharing in research, all within the framework of Contextual Integrity.

Remote Healthcare Technology Use Cases and the Contextual Integrity of Older Adult User Privacy:

Daniela Napoli & Sonia Chiasson

The authors discuss a health emergency scenario, like a seizure or heart attack, in which it would be helpful to contact emergency responders through the remote healthcare technology. This technology would always be on. Similarly if used in palliative care, the device would likely always be on. The authors investigate the existing norms around disclosing intimate details, but only in specific scenarios and would not expect disclosure of mild symptoms and episodic use. Various factors can impact privacy expectations in this scenario. The authors will work toward evaluating the different factors that impact privacy expectations going forward.

Applying Contextual Integrity to the Development of Polypharmacy Indicators in Canada

Pierre-Luc Déziel (Université Laval), Sylvain Auclair (Université Laval), Caroline Sirois (Université Laval), Richard Khoury (Université Laval)

The authors describe the need for developing polypharmacy indicators in a manner that patients feel are appropriate. Current indicators are heterogeneous and have limited usefulness.. The authors aim to develop indicators with contextual integrity in mind when evaluating whether certain types of data violate contextual integrity.

Future work includes using a vignette for CI assessment of polypharmacy indicators when evaluating personal information from government databases.

A Fair Balance: Health data protection and the promotion of health data for research purposes

Irith Kist (Netherlands Cancer Institute and Leiden University)

The paper aims to address the concern of cases of hospitalized individuals suffering in nursing homes due to the use of consent mechanism. Hospitalized individuals may not have the capacity to give traditional notice and consent. Europe's GDPR focus on notice and consent promulgates these issues. We overly focus on the legal basis of data sharing and with the focus on the value of data sharing. The work used CI to explore the normative values for giving consent to propose a solution focusing on the value of data sharing. Future work include the use of CI to instruct/reinterpret GDPR in determining when notice and consent should be utilized.

Session 6: CI, Applications and Methods

GKC-CI Annotations With Large Language Models

Jake Chanenson (University of Chicago), Madison Pickering (University of Chicago), Noah Apthorpe (Colgate University

The authors present work that addresses a gap in evaluating privacy policies – manual privacy policy annotation requires experts in the field, and is not scalable. To address scalability, they leverage LLMs to attempt to make privacy policy evaluation easier. They tested a variety of LLMs (open/closed source, different sizes, varying architectures) of 16 privacy policies with 4.2k training examples and 1.9k testing examples, benchmarked the results against each other, and qualitatively coded the results from the best-performing model. Models were asked to annotate the parameters of excerpts using the GKC-CI framework and were also asked to detect the presence of parameters. They found that larger models performed better than smaller models and newer models also performed better than older models. GPT-3.5 turbo was able to correctly identify correct GKC-CI parameters, including lack of presence, 99.6% of the time and identified the exact same text as a human annotator 77.3% of the time. Some of these errors are attributed to the phrase being ambiguous, semantically the same, the expert labeling being incorrect, or the model over-labeling text.

The next step in this work is to perform large-scale analyses on privacy policies using this system; prior literature has only studied a maximum of 16 privacy policies, whereas this system can be used to label millions. This lends itself to interesting comparisons across industries to identify which are following GKC-CI parameters, or how specifications change in response to regulatory events such as GDPR.

In response to a question of how evaluations were done, the authors specified that they evaluated models in comparison to a dataset hand-labeled as a ground truth then compared the different models and different resulting annotations against this ground truth. They also did not tweak hyperparameters due to it taking a long time and not being transferable to larger models. In response to why they did not use a legal text-specific model, the authors did a test on these types of models such as Legal Bird but found that it was older and had a

worse baseline performance.

Applying Contextual Integrity to Elicit Acceptance toward COVID Mitigation Mobile Applications in the US

Yuanyuan Feng (University of Vermont), Brad Stenger (University of Vermont)

The focus of Feng and Stenge's work is modeling people's privacy attitudes with respect to recently developed public health technologies sparked by the COVID-19 pandemic. They argue that, while many countries have effectively declared the end of the pandemic, this is a good time to consider how to improve the adoption and sustained use of these technologies for the next public health crisis. Additionally, some of these technologies are here to stay, such as contact tracing and vaccination record keeping and sharing. In order to define the factors impacting people's adoption of COVID mitigation technologies, the authors ran a survey where participants were shown 10 randomly selected samples from a corpus of 60 vignettes that varied on three of the framework's parameters, the transmission principle (information sharing in a commercial context or information sharing in a public health focused context), the data type (vaccination records, contact tracing history, test results), and the recipient (government, restaurant, etc.), while fixing the sender and data subject (the same individual using the technologies).

Through multilevel ordered logistic regression, the authors highlighted two key findings: the strong influence of specific parameters in participants' acceptance of the data flow and how interactions between parameters impacted this acceptance. The recipient in the vignettes was the strongest predictor of acceptance among participants. Healthcare recipients in particular encouraged higher acceptance while non-essential stores were less likely to elicit a similar response. The recipient and transmission principle sometimes influenced participants' willingness. Participants were much more accepting of sharing vaccination records for the purposes of entering an establishment if the recipient was an airline company over a healthcare provider, for example.

This work has a concrete goal of informing public health policies and decisions based on people's privacy attitudes and behaviors. Future CI work with similar goals in different contexts (e.g., tech regulation) could use similar methodologies and frame results in similar ways so that the audience of this research (government and organizations) can apply these findings as seamlessly as possible. Questions during the Q&A and discussion focused on challenges of applied CI research such as this. Feng noted that as the amount of variable parameters increases, the complexity (statistical and otherwise) of the study increases significantly. This tradeoff was necessary to manage the scope of this work.

Contextual Integrity in the Context of wearables: an experiment

August Bourgeus (imec-SMIT-VUB), Laurens Vandercruysse (imec-SMIT-VUB), Nanouk Verhulst (imec-SMIT-VUB)

Wearables are increasingly used by people across the world and have privacy implications due to their collecting sensitive data, so this work sought to understand how people understood information flows in wearables. They conducted a factorial vignette survey with demographically representative participants from northern Belgium and asked them to rate 18 scenarios in which each scenario varied in information types, senders, and transmission principles. They found that people were most accepting of the transmission principles where their data was used for developing new technologies or for research purposes. In contrast, people were least accepting when their data was used for advertising. They also found that there was no significant difference in the acceptance of advertisements when senders (the application) used the user's weight data and inferred heart disease. This indicates that people tend to feel powerless over personalized advertisements. The author identified three main themes that are important in developing a smart default: the role of context, individual preferences, and agency.

In response to how the author would want to influence Europe when these results were presented, they mention that Belgium has a lot of government and has data stores on people and that this has the implication that there is a lack of ability to consent. The author spoke about needing to move past the consent model. Another question asked the author to clarify the recipients, but it was complex to add recipients because each transmission principle had differing implicit recipients. In response to their challenges and compromises made for a full factorial survey, the project was explorative and had to move quickly, and this research and results are useful for a larger-scale design.

Session 7: CI and Norms

Privacy Mini-Publics: A Deliberative Democratic Approach to Understanding Informational Norms

Daniel Susser (Cornell University), Matteo Bonotti (Monash University)

The early stages of a paper by Daniel Susser and Matteo Bonotti, call out the curious misalignment between the concept of privacy, defined by Contextual Integrity (CI), and the methods traditionally employed to study it. According to CI, privacy should be understood in terms of collective social norms rather than individual preferences or expectations. However, the commonly used methodology involves asking people about their preferences, which may not accurately capture the essence of privacy. The authors emphasize the distinction between social norms and individual, majority, or average preferences. They propose using Deliberative Mini-Publics to study privacy better to bridge this gap. This procedure entails

convening forums of randomly selected citizens, led by a panel of experts, to deliberate on policy issues collectively. Susser and Bonotti hope that the findings of these mini-publics will influence public policy by providing a more holistic understanding of privacy based on deliberative democracy principles.

There are several critical stages in the Deliberative Mini-Publics process. In Stage 1, participants are recruited through stratified random sampling to ensure inclusivity. Stage 2 focuses on equipping participants with expertise and information through expert briefings from academia, government, and industry. Stage 3 entails deliberation and facilitation, where trained facilitators guide participants through discussions involving contrasting viewpoints, identifying preferred outcomes, and considering trade-offs. The goal is to reach an agreement on existing social norms, determine whether technology has disrupted these norms, and determine whether interventions and regulations are required to address these changes. While deliberative mini-publics have advantages such as obtaining more information and evidence, a broader range of perspectives, and improved problem-solving abilities, they also have disadvantages, particularly in cost and scalability, compared to survey and vignette-based studies. As a result, the goal is to use a hybrid approach combining elements of both methods to address the complex nuances of privacy regulation effectively.

Adaptively Regulating Privacy as Contextual Integrity

Sebastian Benthall (New York University), Ido Sivan-Sevilla (University of Maryland)

Presenting the early stages of a paper titled "Adaptively Regulating Privacy as Contextual Integrity," Sebastian Benthall and co-author Ido Sivan-Sevilla introduce a novel approach to privacy regulation based on Contextual Integrity (CI). The duo started by outlining several challenges in implementing CI within the privacy regulatory process. The first requirement is to operationalize "social goods" as legitimated factors for information flows. Measuring and quantifying these social goods, critical components of effective CI-based regulation, are specifically called into question. The second challenge is the complex, multi-directional information flows in real-world scenarios that are typically hidden from regulators, making traditional regulatory instruments challenging to keep up with. The authors proposed a shift toward adaptive regulation, emphasizing the importance of iterative learning cycles for identifying new privacy risks, enabling real-time monitoring of violation indicators, and validating existing regulatory measures. This proactive, data-driven approach would involve building measurement tools and instruments in collaboration with social scientists, privacy risk professionals, and community leaders. Furthermore, Benthall and Sivan-Sevilla emphasize the importance of more specific and formal modeling of privacy regulation using CI. These models, once established, must be calibrated using empirical data to assess information flows, norms, and their applicability in context.

Session 8: CI, DP, and other Privacy Methods

When PETs Misbehave: A Contextual Integrity Analysis

Ero Balsa (Cornell Tech), Yan Shvartzshnaider (York University)

In this paper, the authors point out to a gap between trust as understood by the public and trust as understood by cryptographers (and also by privacy researchers/practitioners), and most of the privacy enhancing technologies (PETs) such as Google's Privacy Sandbox focus on a narrow definition of privacy (i.e., confidentiality or control). The work aims to bridge the gap between practical applications and theoretical considerations when implementing Confidentiality Impact (CI) in various PETs. His research addressed several key questions:

1. Can CI offer more comprehensive insights into the issues associated with these PET deployments?
2. Does CI possess sufficient explanatory capability to address these concerns effectively?
3. What valuable lessons can we derive from CI analysis?

To illustrate the challenges of CI analysis, Balsa presented three use cases: client-side scanning, anomaly credentials, and federated learning. For instance, anonymous credentials exhibit certain issues, particularly in their underlying infrastructure, which are not easily remedied through CI analysis alone.

The Five Safes as a Privacy Context

James Bailie (Harvard University), Ruobin Gong (Rutgers University)

The work aims to address the challenges in implement Differential Privacy with different parameters in the context of CI.

The authors propose to use Five Saves (FS) framework to parametrize the Contextual Integrity in a situation where the information flow is a statistical dissemination (Safe People, S Projects, S Settings, S Data, S Outputs). FS provides a context for DP as a framework for controlling the disclosure risk of statistical determination; and context for setting DP parameters in implementation. Also, FS describes factors you should consider before releasing a dataset to a broader public. FS is a solution concept for implementing DP in a way that respects contextual integrity. This represents an initial theoretical framework, which will be expanded upon through empirical exploration in future research.

Session 9: CI, Tracking and AdTech

Data Minimization in AdTech: Using Contextual Integrity to Determine Permissible Secondary Data Uses (use-case)

Sara Geoghegan (Electronic Privacy Information Center)

There are many privacy laws in the US including ADPPA, HIPAA, FCRA, and COPPA. While all achieve some form of privacy protection, they have been unable to keep up with the latest technology trends. Further, every state has its own laws and these tend to vary drastically across states. Coupled with the FTC's approach of enforcing laws on a case by case basis, it has become very challenging for all these laws to be properly enforced. Thus, the existing legal framework still needs lots of work to adequately protect consumers' privacy.

The presenter argues that the data minimization principle limits all processing (collection, use, disclosure and retention) of personal information to the processing necessary for the given processing purposes. It is proposed that contextual integrity can help identify any secondary use of individuals' data for other purposes (like harmful surveillance advertising) as out of context processing. For future directions, the presenter is exploring how contextual integrity can be incorporated into both policy and regulation, at best through a positive rights approach.

The Q&A session elaborated on data minimization by discussing to which extent data minimisation intersects with the CI concept of "appropriate flow". While potential discrepancies were attributed to practical limitations, there was consensus on the goal of aligning both concepts as much as possible within these limits.

First-Party vs Third-Party Privacy

Benjamin Laufer (Cornell Tech), Ahana Datta (University College London)

In privacy policies negative reference to processing of personal data by "third parties" is mostly made as an assurance towards data subjects. This work questions this assurance by contrasting such processing operations with the processing by "first parties", pointing out that the ambiguity of both terms lends itself well for the creation of confusion and thus for the usage in privacy policies. This is exemplified by the attempt to assign the aforementioned roles in the case of using Google Chrome on an Apple MacBook to access Facebook in order to speak to a friend.

The presenter argues that, to limit or remove existing ambiguities, contextual integrity can

be helpful as it provides a framework for the specification of the parties involved in the processing of personal data. The consideration of contextually appropriate flows can shed light on the circumstances under which third parties might be appropriate recipients of data. It also highlights that first party processing is also not unproblematic as data can be used for a variety of purposes. Apple's expansion of its own on-platform advertising business after having curtailed third party tracking on its mobile phones is interpreted as a sign of consolidation: while data assurances against third parties weaken, the first parties are continuously engaging in tracking, strengthening their position. During the Q&A it was noted that while such consolidation processes might lead to more efficient antitrust enforcement, a monopoly break-up would not address the fundamental unclarity about the granularity on which CI analysis should be applied to data sharing practices (between companies, within companies, within teams).

The panel discussion, at the end of the session, linked data minimisation and consolidation of monopolistic structures as data minimisation also includes ending a majority of current third party access. The resulting consolidation moves the resulting revenue from the third parties to the first party, with no positive effect for data subject privacy. Thus, while the existence of data brokers under data minimisation is questionable, the question of what functionality is rightfully the user's when they interact with the web requires expanding the privacy lens with other approaches. Current antitrust structures fail to ensure such rightful functionality as they allow very large only platforms to claim a variety of ill-defined purposes, effectively side-stepping the data minimisation principle. While neither antitrust nor privacy approaches individually were considered sufficient to address the power imbalance resulting from this lack of conceptual clarity, the presenters indicated optimism about a potential merger of the two.