Regulatory CI: Adaptively Regulating Privacy as Contextual Integrity

SEBASTIAN BENTHALL, New York University School of Law, USA
IDO SIVAN-SEVILLA*, University of Maryland, USA

The practice of regulating privacy, largely based on theories of privacy as control or secrecy, has come under scrutiny. The notice and consent paradigm has proven ineffective in the face of opaque technologies and managerialist reactions by the market. We propose an alternative regulatory model for privacy pivoted around the definition of privacy as Contextual Integrity (CI). Regulating according to CI involves operationalizing the social goods at stake and modeling how appropriate information flow promotes those goods. The social scientific modeling process is informed, deployed, and evaluated through agile regulatory processes – adaptive regulation – in three learning cycles: (a) the assessment of new risks, (b) real-time monitoring of existing threat actors, and (c) validity assessment of existing regulatory instruments. At the core of our proposal is Regulatory CI, a formalization of Contextual Integrity in which information flows are modeled and audited using Bayesian networks and causal game theory. We use the Cambridge Analytica scandal to demonstrate existing gaps in current regulatory paradigms and the novelty of our proposal.

1 INTRODUCTION

There are many well-documented problems and blind spots in privacy regulation today. Many regulators still hold to the paradigm of privacy as individual control over personal information [69], putting impossible burden on individuals through notice and consent requirements [4, 40, 57, 60]. Complex and opaque information flows make assessing the costs and consequences of privacy violations unclear. Regulated industries employ managerialist tactics to determine the nature and scope of their own compliance [17, 48, 66]. This paper proposes an alternative model of privacy regulation that addresses these known shortcomings.

Our proposal is centered on the theory of privacy as Contextual Integrity (CI), which defines privacy as the appropriate flow of information, placing front and center the social values that privacy promotes [45, 60]. We outline how regulators can design and employ rigorous models of information flows and their consequences for social outcomes. Learning and updating these models in an environment in which market actors are rapidly innovating in their collection and use of data [42, 71] involves agile regulatory processes. We imagine close and adaptive engagement between regulators, industry, and civil society in understanding the goals and needs of privacy.

We map out three main challenges for a CI-based regulatory model: (1) The struggle to operationalize the desired social ends and goals for social contexts in which information flows. This makes it difficult to systematically evaluate the legitimacy of information flows from new technological developments; (2) The challenge of monitoring commonly obscure and opaque information flows and dynamically assess their social validity; (3) We are also lacking learning processes to temporally assess the validity of regulatory instruments, which unfortunately, are often abused and fail to achieve their original intention [17].

To address these regulatory challenges, we adopt the novel conceptualization recently offered for adaptive regulation in previous work [61], taking into account agile regulatory principles discussed by various scholars [6, 13, 39, 68]. Three parallel learning cycles update the regulatory process when necessary through structured close engagement between regulators and regulated industries. These regulatory processes can enable near real-time questioning of information flow practices when social values are at stake and data inferences strikingly shift from agreed-upon norms.

*Both authors contributed equally to this research.

Authors' addresses: Sebastian Benthall, spb413@nyu.edu, New York University School of Law, USA; Ido Sivan-Sevilla, sevilla@umd.edu, University of Maryland, USA.

At the heart of this proposal is Regulatory CI, a formal framework for modeling information flows and the social goods at stake. This framework serves to structure what is learned through the adaptive regulation process. This framework explicitly models the actors, incentives, and information flows in the sociotechnical context of the information systems involved. It is an application of multi-agent influence diagrams, a form of causal game theoretic modeling used in computer science [22, 31]. We motivate this framework as a crystallization of CI that addresses some known gaps in the theory and affords some useful analysis of the privacy implications of information flows based on graphical criteria.

The first learning-cycle (LC #1) aims to cope with the challenge of assessing newly introduced information flows. We urge social scientists, privacy risk professionals, civil society organizations, and community leaders to build measurements and models for the social purposes that different informational contexts aim to promote. The information-gathering process for this learning cycle can flag and model new market practices that are currently unregulated and highlight new avenues for information flows that tilt socially desirable goals.

The second learning-cycle (LC #2) tackles the challenge of monitoring difficult-to-capture information flows in the market. We call regulators to team up with industry, researchers, and civil society actors and engage in real-time monitoring instruments of known privacy thresholds, empirically calibrating the model from the first learning cycle to check whether desired norms are actually followed. These tools can include, for example, the independent tracing of data collected by third parties in popular websites and mobile apps, or a computational comparison of changes in privacy policies of main digital service providers over time.

The third proposed learning-cycle (LC #3) verifies the validity of existing regulatory instruments and checks whether the instruments (e.g., notice and consent, end-to-end encryption, differential privacy techniques) serve the norms. Can markets meaningfully comply with privacy requirements? Do consumers really comprehend notices on data usage? Do regulated companies show meaningful compliance, or mostly engage in 'Regulatory Managerialism' - normalizing business interests and procedural compliance over public values [17]. In contrast to the previous two learning cycles that model and monitor new and existing privacy risks, this cycle assesses the relevancy and effectiveness of regulatory instruments over time. In case regulatory goals and requirements turned out to be too difficult to follow, regulatory re-design may be triggered.

We will use the Cambridge Analytica data scandal as an example of how regulators could adaptively regulate the social goods at stake through proper modeling of the associated information flows and learning cycles that could proactively prevent the next scandal. The affair was exposed in 2018 by whistleblower Christopher Wylie. He shared how despite privacy regulations, data belonging to 87 million Facebook users was collected without consent by the British consulting firm Cambridge Analytica [14]. The data were predominantly used for political advertising to assist the 2016 presidential campaigns of Ted Cruz and Donald Trump. Facebook allowed a third party app to collect personal information from survey respondents and their Facebook friends, which the Trump campaign then used to build psychographic profiles, determining users' personality traits based on their Facebook activity [53]. Those profiles enabled campaign managers to conduct micro-targeting techniques and display custom-made messages to different voters, covertly manipulating voters' political behavior and undermining their autonomy [62].

The paper is organized as follows. Section 2 addresses the problems of existing privacy regulation and motivates our proposed approach, which marries Contextual Integrity with Adaptive Regulation. Section 3 introduces Regulatory CI, a formalization of Contextual Integrity tailored to the purposes of regulation. Sections 4, 5, and 6 consider the three learning cycles of adaptive regulation, and discuss how hypothetcally Regulatory CI would have been employed in

the Cambridge Analytica scandal under such a regime. Section 7 discusses the limitations of our proposal and directions for future work.

2 REGULATING PRIVACY: FROM TRADITIONAL DATA REGULATION TO ADAPTIVE OVERSIGHT OF INFORMATION FLOWS

2.1 Contextual Integrity

Contextual Integrity (CI) [45] is an interdisciplinary theory of socially meaningful privacy. According to CI, privacy is *appropriate information flow*. An information flow is the transfer of information about an attribute of a data subject from a sender to receiver. Appropriateness is defined in terms of contextually grounded information norms. Contextual Integrity is concerned with the normative appropriateness of both positive information flows (such as the flow of information from a patient to a doctor) and negative restrictions on information flows (such as the confidentiality of that information with respect to third parties).

CI differs from, and complements, other theories of privacy. CI explicitly contradicts the notion of privacy as control over personal data [69], which has been the target of many critiques [16]. Whereas the former is perhaps best exemplified legally through the notice and consent regime of consumer privacy protection, CI is best represented legally by sectoral privacy laws such as HIPAA, GLBA, and FERPA. Contrary to approaches to privacy that emphasis its particularism and political contestedness [44], CI presents a unifying theory of how privacy expectations manifest in different contexts, and emphasized the shared social and political understandings that emerge in mature fields or spheres of society. CI is also different from cryptographic notions of privacy such as differential privacy [20] that focus entirely on the prevention or limiting of information flows, as opposed to the conditions under which information flows are appropriate.[1]

The theory of Contextual Integrity is inspired by Walzer [67]'s conceptualization of a just society. According to Walzer, society is composed of different spheres - market, political, healthcare, and military spheres for instance - and advantages accrued in one sphere, should not be translated to advantages in another, lest some members of society come to dominate over others. Extreme wealth from the market sphere, for example, should not be translated into access to better healthcare or greater political power. Nissenbaum [45] expands this view of justice to a theory about the legitimacy of information flows. In her work, spheres are elaborated into a specific view of a social *context*. [2] Contexts are defined in terms of their *purpose* in society, or why the context exists in society. For example, the context of health care has the purpose of preserving the health of people. Contexts are also defined in terms of the roles of agents operating in the context, and in terms of the kinds of relevant personal information that flow within the context. Each context forms norms about the flow of personal information which promote the purposes of the context, while balancing the ends of the actors. CI therefore argues that privacy is not so much an end in itself but rather a pattern of social behavior and expectations which supports the promotion of other social goods.

CI is a theory with both descriptive and normative components. To its credit, many of its descriptive claims, especially those about the way privacy expectations vary with context and can be parameterized into socially meaningful norms,

[1] Arguably, CI is a theory of data protection in its broad scope beyond privacy. Data protection is identified as a fundamental human right in the European Union, and this right is the basis of the General Data Protection Regulation among other laws. In the European context, privacy is narrower than data protection. In the United States, in scholarly discourse 'privacy' has expanded beyond its narrow definitions. See also the distinction between privacy enhancing technologeis (PETs) and protection optimizing technologies (POTs)[33].

[2] Crucially, a context is **not** defined as a spatial location or the vicinity of some technical system. Indeed, Benthall et al. [10] note two divergent ways of conceptualizing "context" in computer science: (1) the descriptive *situation* of the system, including its users and their location; and (2) the normative social *sphere* of social expectations in which norms are embedded. CI is primarily concerned with the latter spheres; Nissenbaum [46] clarifies that "Respective roles, activities, purposes, information types do not exist in a context; rather, these factors constitute a context."

have been supported by empirical findings [2, 38]. We also acknowledge that there are several known theoretical gaps in CI which are open problems to be addressed in future research and elaboration of the theory [10]. However, this paper takes as an assumption that CI as a theory of privacy is valid, and an improvement to prevailing theories of privacy as control and privacy as secrecy. One contribution of this paper, presented in Section 3, is a novel formalization of CI that addresses some of the known weaknesses in the original theory, aimed to support regulating privacy as CI.

2.2 Regulating privacy using CI

We first recognize various regulatory challenges for promoting the CI approach. Scholarship on regulation and governance defines regulation as the "design, monitoring, and enforcement of rules" [32, 34]. Applying the CI approach to privacy regulations includes challenges in each of those regulatory phases. The first challenge is to understand the object that needs to be regulated. Instead of regulating data or personal information, a CI-based approach regulates information flows, in which data is just one element. To do so, we need to properly model information flows, operationalize the social goods at stake, and enable or constrain the information flow accordingly. We use the Cambridge Analytica case to illustrate each of these facets.

Operationalizing social goods. For CI, information flows are legitimized by the contextual purposes, societal values, and the ends of actors. One obstacle for regulators interested in applying CI is the design of rules based on measurements of the social purposes and values that the framework aims to safeguard. We struggle to measure or quantify these social goods, which are emergent properties of each context and its information flows. Effective CI-based regulatory design depends on clearly operationalizing these social goods so that new technologies and privacy policies can be evaluated for the social purposes that privacy promotes.

In the Cambridge Analytica scandal, the ability of third-parties working with Facebook to gain personality insights from users' data and covertly impact their future behavior is a breach of users' autonomy and should be classified as a manipulative practice that regulators ought to prevent [62]. A successful operationalization of the threat to users' autonomy in this case would be enabled by a model that traces the information flows that might lead to such autonomy breach. Here, the political autonomy of voters is the social good at stake in social media settings.

Complex, opaque, and dynamic information flows. Second, complex and opaque information flow settings challenge the detection of privacy violations and the understanding of the impact of information flows. Even after careful regulatory design of CI rules based on modelled information flows and the social purposes they aim to promote, privacy regulators operate in complex and constantly changing technological environments. Regulators often lack the expertise and resources to comprehend information flows and detect privacy violations on the spot. [3]

In the Cambridge Analytica scandal, information flowed inappropriately between users, social media platforms, and political advertisers. Modeling and continuous monitoring of the relevant information flows would have kept regulators at pace with what is at stake.

Abused privacy regulatory instruments Third, the validity of regulatory instruments that are applied to tackle CI violations is under a constant threat. Past industry practices show how regulatory interventions such as notice and consent requirements are quickly turning to meaningless privacy measures [36, 57]. For properly regulating the

[3]Politically, by adopting a political economy view on data and inspecting the relations that our data materialize in society, it is hard to imagine a tilt in the balance of power between powerful tech companies, regulators, and consumers. Still, the recent Digital Markets Act in the EU that requires significant transparency from big tech platforms is an important step towards the ability of regulators to understand and model information flows of main data-driven market services. This might shift the business model and information flows suggested by companies, as recently seen in Facebook's decision to create a subscription-based, ad-free option, for EU citizens [54].

dynamic consumers' privacy environment, where business interests and policy intentions are often in conflict [17, 41], the validity of regulatory instruments needs to be constantly tested.

In the Cambridge Analytica scandal, notice and consent was the regulatory tool used to safeguard users' privacy upon social media usage. Facebook - the regulated company - was able to only partly notify its users on how it uses their data, while its users could not comprehend the risks associated with their consent decision. The consent provided could never anticipate risks such as illicit data access and use, or the type of inferences that Facebook's third party will extract from users' data. Given previous observations by researchers on how privacy policies are tricky and far from clear with regards to actual information practices [57], regulators should constantly ensure, through surveys and reviews, that their selected measures are effectively serving their purpose.

2.3 Adaptive regulation

The notion of adaptive regulation has been discussed by various scholars who have tried to cope with the challenge of slow and 'sticky' regulatory systems that cannot respond to the pace and uncertainty of domains like climate change, healthcare, and Internet of Things [13, 39]. In general, when policy change does occur, it tends to be crisis-driven, inspired by 'focusing events' [11] that successfully capture the limited attention of policymakers in boundedly rational policy arenas [35]. The occurrence of a crisis, however, usually indicates that change is already overdue. Regulations formulated in times of crisis, without preparation or sufficient analysis, may be excessive or ineffectually designed [6]. Importantly, adaptive regulatory processes are not strictly proactive or reactive, they aspire on-going engagements between regulators and industries to protect social goods. The goal of adaptive regulatory processes is to increase the responsiveness of the regulator to societal needs and industrial conditions. Regulators also commonly lack industry knowledge or may be captured by the industry, producing regulatory measures that are hard to follow or meaningfully comply with [39]. Adaptive regulatory processes then, should enable constant validity checks of regulatory instruments as well.

In the dynamic privacy regulatory arena, advances in technology change how personal information is collected, stored, shared, analyzed, and disseminated. Expectations and understandings of privacy are rapidly shifting in response to a wide spectrum of privacy-invasive technologies underlying applications in everyday use [49]. In addition, contemporary applications of data-driven technologies make it impossible for companies to decide and anticipate beforehand on the purpose of data processing, as the added value of data partly resides in the potential to uncover new purposes that may benefit the data collector [27]. Hence, any regulatory effort that is not adaptive runs the risk of obsolescence.

There are various instruments and techniques, currently applied across the patchwork of privacy regulations around that world, that could jointly enable effective implementation of adaptive privacy regulation. Those regulatory instruments will be discussed per each learning cycle in sections 4, 5, and 6. All implementation recommendations are based on a legally capable data protection agency, empowered towards strict privacy enforcement actions. Such agency would enjoy the mandate of operationalizing social goods affected by information flows and will be vested with the authority to apply significant sanctions, such as information-flow bans, or the posing of significant administrative fines. Data Protection Authorities in Europe post-GDPR already enjoy such a mandate. In the US context, the Federal Trade Commission (FTC) would need to be imbued with additional powers to supervise and sanction data processors and controllers.

3 MODELING APPROPRIATE FLOW WITH REGULATORY CI

Adaptively regulating privacy as CI will require rigorous operationalization of social goods, incorporating known information, and acknowledgement of uncertainty about unknown flows. These are all complex and interrelated operations. To manage this complexity successfully, regulators will need a systematic way to model contextualized, appropriate information flow that is applied consistently across adaptive learning cycles. In this section, we describe a formalization of CI for modeling and regulating appropriate information flow, which we call Regulatory CI.

3.1 Concerning formality

Regulatory CI is a formal method of modeling contexts and information flows that draws on probabilistic graphical modeling techniques from computer science. We are not the first to recognize the need to responsibly connect socially meaningful and computational definitions of privacy [49]. We are also not the first to formalize CI. Nissenbaum [45] did not provide a single formalism of CI, and instead chose to let computer scientists experiment with how to operationalize the theory through specific applications. For example, Barth et al. [5] formalize "some aspects" of CI in a way that is well motivated by the use case of designing audit mechanisms [18]. Others have extended CI's original framework to accommodate new applications [58]. Our formalization of CI is a novel contribution that is motivated by the concerns of data protection regulation. Namely, we contend that because the effects of an information flow can be the consequence of how that flow participates in a larger network of flows, a systematic technique is needed to map flows between multiple, strategically acting agents.

Notation	Meaning
P	Contextual purposes
R	Contextual roles
U_r	Contextual ends, by role
A	Contextual attributes
N	Norms, of a context
W	Information flows
l	Legitimacy, a function of P, R, U, N
e	Equilibrium outcomes, a function of R, U, N
a	Appropriateness, a function of N, W
$\hat{P}, \hat{U}, \hat{N}, \hat{W}, \ldots$	Estimates of P, U, etc. based on empiricial proxies

Table 1. Notation used in this paper.

We anticipate several objections to this formal technique. One objection is that we are offering a formalization of privacy that, on its surface, does not look like cryptographic formulations of privacy, such differential privacy [20]. However, these definitions have been subject to the critique that they do not capture what is socially meaningful about privacy [3, 49]. This motivates a broader mathematical formulation. It is anticipated that CI and cryptographic privacy concepts will eventually be integrated [8].

Beyond privacy, there are many known critiques of the use of formal or computational modeling to address the complexity of norms in sociotechnical systems [55]. However, we see a role for computing [1] in effective regulation. What is needed is a way of modeling sociotechnical systems in their complexity, including the social actors in the system's context, and their interests. Our approach uses causal modeling, which has been widely applied to problems in algorithmic fairness [15, 24, 37] and accountability [30, 50]. Rather than proposing regulators adopt one 'true' model, we are defining a space of models over which regulation can be negotiated and contested in an evidence-based and systematic way.

3.2 Context: spheres and norms

We posit the following formal expression of a context as structure within Regulatory CI.

Definition 3.1 (Context (Sphere)). A context is a defined as a tuple (P, R, U, A, N) where: P is a set of contextual purposes; R is a set of agent roles; $U = \{U_r | r \in R\}$, the agents *ends*, are utility functions for each role in R; A is a set of information attributes; and N is a set of information norms (see below).

Contexts are defined in terms of their purposes in society. The purpose is, at a high level, an explanation for why the context exists. An example of a context and its purpose is the context of health care, which exists to preserve the health of people. Contexts are also defined in terms of the roles of agents operating in the context, and in terms of the kinds of relevant personal information that flow within the context. The above definition is a simple translation of the analytic description of a context into a parameterized form.

The concept of *information norm* is connected to the concept of a context. Information norms are defined in terms of the properties of the context they adhere in. In Regulatory CI, a norm is defined as follows:

Definition 3.2 (Information norm). For a context (P, R, U, A, N) and given a set of transmission principles E, an information norm $n \in N$ is a tuple $(s, r, u, a, e) : s \in R$ is the sender of the information; $r \in R$ is the receiver of the information; $u \in R$ is the subject of the information; $a \in A$ is the information attribute; and $e \in E$ is a transmission principle.

Sender, receiver, and subject are all defined in terms of roles in a social context. For example, in the context of health care, agent roles include those of doctors, patients, and the guardians of patients. Attribute refers to the type of information being shared. Transmission principles are normative restrictions on information flows that fit the description of the other four parameters. We have introduced E to stand for a set of transmission principles that exists outside of any particular context. Example transmission principles include confidentially, reciprocity, notice and consent, "with for a warrant", and use restrictions.

3.3 Information flows

A key tension within CI is that while it endeavors to illuminate norms of information flows in socially meaningful terms, the true mechanics of information flow may elude social comprehension. This has become especially challenging as more and more statistical techniques and digital instrumentation have enabled higher-level inferences about people to be drawn from lower-level data that can be collected relatively innocuously [46]. An analyst or auditor determining if information flows are appropriate may need to have a thorough understanding of the flow than is broadly socially understood.

In particular, the so-called "social" understanding of information flows in CI suffers from what Reddy [51] calls the *conduit metaphor*: the information flow is one that *transfers* content, the specific attribute a about a data subject u, from sender to receiver. There are principled reasons to reject this notion of information flow. Prior work [47] has identified the instability of the attribute parameter, because what can be inferred from a piece of information can vary based on context and is not limited to its explicit or syntactic type. Moreover, the notion of a singular data subject is challenged by notions of relational [65, 70] and group [63] privacy, which highlight how data about one individual can support inferences about others who share relevant characteristic, especially when machine learning and statistical techniques are employed.

One formal solution to this problem of modeling information flows is *situated information flow theory* (SIFT) [7], according to which information flows are causal flows with regular associations due to a larger context of causal flows. In practice, this involves modeling information flows as taking place within a network of variables. Each node in the network can represent a member of one of the sphere's roles, or a technical device, or a variable of interest such as an information attribute. The nodes represent conditional probability distributions over possible states, as in a Bayesian network.

A Bayesian Network (BN) is a graphical model of the joint probability distribution of several random variables. Each variable is represented as a node in a directed acyclic graph (DAG). Each node has an associated probability distribution that is conditional on its parents. The joint probability distribution is the product of these conditional distributions. Bayesian network structure can then be used to identify possible inferences that can be made from data as it flows within the system by identifying *active paths*.

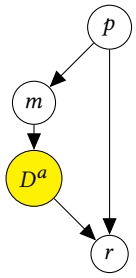


Fig. 1. A Bayesian network representing a social media based advertising strategy. The social media user has a personality p which influences both her social media activity m and her responsiveness to different kinds of advertising r. And advertiser with access to the social media activity is able to make a decision D^a. Because of the structural properties of this network, D^a decision is informed, via m, about p. D^a can use this information to control r.

Definition 3.3 (Active path). In a Bayesian network $G = (V, E)$ with observed nodes $O \subset V$, an *active path* is any path $P \subset V$ such that each node in P is connected, in sequence, by an edge, and for any v-structure $(a \rightarrow c \leftarrow b)$, either c or one of its descendants is in O, and no other nodes on the path are in E.

A well known result is that in a Bayesian network, variables that do not have an active path between them must be probabilistically independent. An active path between x and y indicates that x and y may carry information about each other. Figure 1 illustrates how the structural properties of a network of information flows can reveal what inferences are possible from data. An advertiser can make a decision D^a based on the observed social media activity m of a user. This social media activity conveys information about the user's personality, which is a factor in how she responds to advertising. In this diagram, m and r are connected by an active path, and so may not be independent from each other.[4]

Regulatory CI builds on and adopts SIFT. This allows adaptive regulators to employ the powerful principles and tools of graphical modeling in all of the regulatory learning cycles. Bayesian networks can be calibrated to empirical data and analyzed using known algorithms that are well understood in computational statistics and machine learning. Specifically, given a context (P, R, U, A, N) and a model of information flows W, represented as a Bayesian network, an auditor can begin to audit if the information flows are in compliance with contextual norms.

3.4 Modeling situations

While CI treats contexts as normatively understood spheres of activity, these are not the object of regulation. Rather, regulators are concerned with the real activities and outcomes within their jurisdictions. In this section, we introduce a method for modeling *situations* in Regulatory CI. This method combines the teleological elements of CI's understanding of contexts, with their purposes, roles, and ends, through the use of Bayesian networks to model information flows.

[4]Bayesian networks may include edges between nodes that are conditionally independent from each other, but this is not preferred. A Bayesian network in which only nodes that are conditionally dependent on each other are connected by edges is called a *faithful* network with respect to the represented probability distribution.

Formally, these elements are both present in Causal Influence Diagrams [19, 22, 23, 56] and Multi-Agent Causal Influence Diagrams [25, 31]. Multi-Agent Influence Diagrams build on Bayesian networks by introducing agents, their decisions, and their payoffs. They retain many of the useful properties of Bayesian networks (BN), and these properties can be used to better understand game-theoretic equilibrium outcomes of the model.

Definition 3.4 (Multi-Agent Influence Diagram). [25, 31] A multi-agent influence diagram (MAID) is a triple (N, V, E), where: $N = 1, 2, ..., n$ is a set of agents; and (V, E) is a directed acyclic graph (DAG) with a set of vertices V connected by directed edges $E \subseteq V \times V$. These vertices are partitioned into: $\mathcal{X}$, chance nodes; $\mathcal{U}_n$, utility nodes for each agent, such that $\mathcal{U} = \bigcup_N \mathcal{U}_n$; and $\mathcal{D}_n$, decision nodes for each agent, such that $\mathcal{D} = \bigcup_N \mathcal{D}_n$.

When a MAID is developed into a Multi-Agent Influence Model, the chance nodes are given conditional probability distributions governing their realization. Given a *strategy profile* σ which assigns a decision rule (conditional probability distribution) to each decision node $\mathcal{D}$, the MAID is induced into a BN. From this BN, it is straightforward to compute expected payoffs for each agent. The question then becomes what decision rules constitute the strategy profile, and are these in game-theoretic equilibrium with respect to each other and the expected payoffs.

The graphical structure of a MAID can be used to identify which decisions *strategically rely* on each other. Informally, a decision D_1 strategically relies on another D_2 if the optimal decision rule for D_1 depends on the rule chosen for D_2. These strategic reliance relations can then potentially be used to decompose the problem of computing equilibrium strategies into simpler subgames.

More formally:

Definition 3.5 (Strategic reliance). [31] Let D be a decision node in a MAID $\mathcal{M}$, δ be a decision rule for D , and σ be a strategy profile such that δ is optimal for σ. D *strategically relies* on a decision node D' in $\mathcal{M}$ if there is another strategy profile σ' such that δ' differs from σ only at D', but δ is not optimal for σ', and neither is any decision rule δ' that agrees with δ on all parent instantiations $\mathbf{pa} \in \mathrm{dom}(Pa(D))$ where $P_{\mathcal{M}[\sigma]}(\mathbf{pa}) > 0$.

The graphical criterion that is strongly associated with strategic reliance is *s-reachability*. For decision node D belonging to agent a, let $\mathcal{U}_D$ be the set of utility nodes in $\mathcal{U}_a$ that are descendants of D.

Definition 3.6 (S-reachability). [31] A node D' is *s-reachable* from a node D in a MAID $\mathcal{M}$ if there is some utility node $U \in \mathcal{U}_D$ such that if a new parent $\hat{D}'$ were added to D' , there would be an active path in $\mathcal{M}$ from $\hat{D}'$ to U given $Pa(D) \cup U$, where a path is active in a MAID if it is active in the same graph, viewed as a BN.

Koller and Milch [31] prove soundness and completeness results that show that given a graph structure, there is a MAID with that structure in which D strategically relies on D' if and only if D' is s-reachable from D.

3.5 Key questions for adaptive regulation of privacy

CI raises several key normative questions for privacy regulation. Using Regulatory CI, we will frame each of these questions using the constructs of the theory, including the sphere model (P, R, U, A, N) and the information flows W, and their empirically estimated values, $\hat{P}$, $\hat{U}$, $\hat{N}$ and $\hat{W}$.

Are the norms legitimate with respect to the contextual purposes? This is answered through an analysis of the situation model. The norms N are legitimized when equilibrium outcomes of agents following the norms $e(R, U, N)$ achieve the contextual purposes P; $l(P, R, U, A, N) = P(e(R, U, N))$. This is a focus of the first learning cycle, analyzing new risks

from information flows, discussed in section 4. The cycle is responsible for verifying **what the norms should be** given the purposes, the context, and the situation.

Are the information flows consistent with the norms? This is a function of norms and information flows, $a(N, \hat{W})$. This is a focus of the second learning cycle, real-time monitoring, discussed in section 5: checking to see if **the norms are actually followed**.

Are the observed social outcomes consistent with the model? If the model is making predictions that do not hold out in practice, that suggests the model must be correct. $\hat{P} \simeq P(e(R, \hat{U}, \hat{N}))$. This is a focus of the third learning cycle, discussed in Section 6: **do regulatory instruments serve the norms** or should the model be corrected?

In the following sections we detail how the three learning cycles in the adaptive regulation process enable (1) modeling or re-modeling of appropriate information flows; (2) monitoring of compliance by policy targets; and (3) assessing the validity of regulatory instruments. The cycles revolve around the continuous improving and updating of a model off the context and its appropriate flows.

4 LEARNING CYCLE #1 - ANALYZING NEW RISKS FROM INFORMATION FLOWS

The identification and analysis of new risks is crucial for designing regulations for new applications of technology in society. Forward-looking identification of new risks should involve inputs from a range of stakeholders - including industry, academics, civil society, privacy experts - for the purpose of ensuring the modelled information flows still hold and realize the social goods at stake. The established models are assessing the social risks following the introduction of new technologies or the expansion of existing ones. The cycle can be triggered by regulators, investigative journalists, or technology companies themselves when they start to operate or change their existing services in the market. Regulators are engaged with a domain of massive scale and pervasive digital instrumentation, and information gathering can involve data collection from on-line sources. The CI research community, and the privacy research community more broadly, has already developed many techniques for gathering and analyzing information from the wild [2, 21, 57].

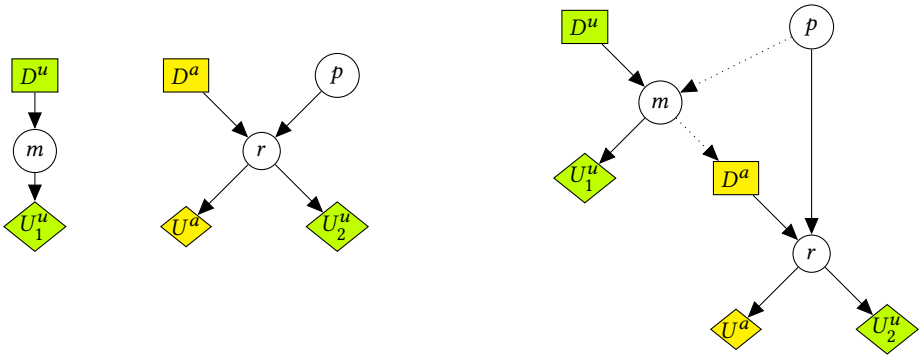
A number of policy techniques adapted from other domains would facilitate this learning cycle. Data processors could register themselves, just like data brokers in the state of Vermont in the US, in a dedicated repository, and provide regular updates on pertinent information. The registered data-driven market companies could become subjects for 'privacy hackathons,' where different privacy stakeholders from industry, academia, and government come together and aim to model information flows of popular digital services.[5] Once an information flow model has been established, it could be subject to comments from various stakeholders through a dedicated public comment period.

4.1 Example: Modeling social media and political manipulation

Imagine a committee of stakeholders gathered to determine regulations and procedures surrounding the relationship between social media and political advertising, using MAIDS as depicted in Figure 2. The committee first determines that the context of the problem is elections, which have the purpose of legitimately determining public representatives. Voter autonomy is essential to electoral legitimacy, and this means preventing voter manipulation.

Some amount of political advertising is normatively allowed in elections. But specific targeting of advertising based on psychographic profiling that may bypass rational decision-making is determined to be going too far. Advertisers can post ads D^a without knowledge of the specific personality p of the users it targets, and so must advertise in a more

[5]The White House recently engaged 'white-hat hackers' to hack commercial Large Language Models such as ChatGPT. The overarching tradition of government's reliance on 'white-hat hackers' could be diffused to the privacy landscape as well [43].



(a) The user's perspective in the Cambridge Analytica scandal, represented as a MAID. When the information flows from p to m and m to D_a are hidden, neither D_u nor D_a is s-reachable from the other.

(b) The advertiser's perspective in the Cambridge Analytica scandal, represented as a MAID. Because of covert flows $p \to m$ and $m \to D^a$.

Fig. 2. Two MAIDs representing the Cambridge Analytica scandal. Figure 2a depicts the point of view of the user, in which their decision to use social media does not strategically rely on the presence of the advertiser. Figure 2b shows how covert information flows allow the advertiser to influence outcomes for the user. Arguably, the social good of autonomy, which is important in the context of democratic voting, can be operationalized as the absence of the manipulation depicted here.

general way. Advertisers attempt to get a certain reaction r from the user, such as voting for a particular candidate. This reaction r is consequential for both the advertiser U^a and the user U_2^u.

Now the committee considers the introduction of social media. The user (u) can make decisions about whether and how to use social media D^u. This influences their social media activity m, and this activity affords them some utility U_1^u. When the social media usage is independnet from the political advertising, the situation can be modeled as in Figure 2a. In this world, the user's social media activity does not open up the possibility of manipulation. Formally, neither D_u nor D_a is s-reachable from each other, implying they do not strategically rely on each other. In this depicted game, the user at D_u will seek to maximize U_1^u without downstream consequences for U_2^u.

The regulators choose to operationalize the autonomy of voters as the absence of manipulation with respect to their participation in elections. Cambridge Analytica was scandalous because this autonomy was violated, and this was possible, in our analysis, because of covert information flows. The committee considers how, unbeknownst to the user, her social media activity m is also influenced by some unconscious aspects of her personality p, which also influence her reaction to advertising. Moreover, the social media activity m may be observable by the advertiser, because the advertiser can violate social media's terms of service to illegally access it. This situation corresponds to a different diagram, Figure 2b.

In this diagram, D_a and D_u are s-reachable from each other, and hence the two decisions strategically rely on each other. This means that the decision rules chosen *ex ante* by the agents would, in strategic equilibrium, be dependent on each other. If both the user and the advertiser are aware of all of the information flows depicted in the graph, then the user will choose to participate in social media not only in pursuit of utility U_1^u, but also in anticipation of how the advertiser will react D_a to their social media activity m, and the consequences of this reaction for their utility U_2^u. The advertiser may, in turn, anticipate the user's concerns and adopt a policy D_a that incentivizes the user to be more revealing at D_u.

When the information flows $p \rightarrow m$ and $m \rightarrow D^a$ are covert, this puts the user at a strategic disadvantage, as they are unable to anticipate the strategic response of the advertiser that will have influence over their outcomes. Hidden information flows, due to privacy policy violations or data "up and down the food chain" [46], can undermine consumer autonomy. Without visibility into these flows, consumers cannot be expected to make fully autonomous, strategic decisions on the market [9].

The diagrams used in the above scenario are simple and unrealistic, used here for presenting Regulatory CI within the limits of a scholarly paper. In practice, these models would be worked out by the regulator and stakeholders from the private sector and civil society during the first learning cycle. Stakeholders might each create their own models, representing their concerns, and these models can be merged into a model or set of models used for regulation. The modeling framework sensitizes the regulator to which information flows are threatening to these social goods, and suggests ways to monitor the market for problems. For example, if misuse of social media data can undermine voter autonomy, regulators might require social media companies to register their APIs as potential information leaks, and require political advertisers to disclose their information sources. Political advertising revenues can be disclosed as a proxy for their rewards. Automated surveys of consumer expectations can be established to verify that these are not out of step with situational information flows. The model is used to instrument the problem, and these instruments can be constantly updated during the following *real-time monitoring* and *validity of instruments* learning cycles.

5 LEARNING CYCLE #2 - REAL-TIME MONITORING OF VIOLATIONS.

Adaptive regulation calls for on-going and real-time monitoring to track performance indicators. When regulating privacy by CI, real-time monitoring is important for preventing harm and inspecting privacy behavior based on information flow models constructed in the first learning cycle. Because of the weakness of consumer consent with respect to protecting privacy [4], there is no reason to expect market forces to naturally correct for a systemic privacy violation. By inspecting data flows and privacy policies in the wild, regulators will be able to assess whether norms are actually followed and how modeled information flows behave in practice.

To facilitate the implementation of the second-learning cycle, dedicated research funding could be secured for sponsoring academics and privacy experts work among privacy regulatory agencies, creating non-residential fellowship programs within regulatory authorities. This could help bridge expertise gaps and create important connections for regulators with the greater community of privacy experts which would help design and apply monitoring tools of opaque information flows across different technologies.

5.1 Detecting and assessing noncompliance

Building on research into automated systems that detect inappropriate information flows in big data systems [18], regulators may turn to technology to monitor for noncompliance. Regulators may observe information flows $\hat{W}$ and use their model of norms to detect violations of appropriateness conditions $a(N, \hat{W})$. They might alternatively observe norms $\hat{N}$ as expressed in privacy policies and compare them to the legitimate norms N.

Regulators may have access to additional information once a company has come under regulatory scrutiny. Auditors with access to the company's internal records can reconstruct what might otherwise be a 'black box' in norms assessment models. In addition to providing new information to assess compliance, these information sources can be (in an anonymized way) reintegrated into the first learning cycle in order to improve regulation.

5.2 Using data in Regulatory CI

The models in Regulatory CI provide a rigorous basis for statistically integrating empirical data and testing for compliance. Recall that a MAID defines a dependency structure between variables, and that this structure is shared by a wide range of underlying probability distributes. We follow Everitt et al. [22]'s definition of a Multi-Agent Influence Model (MAIM) as a MAID that is also fully parameterized with a probability distribution. Combined with a strategy profile σ, the MAIM determines a BN. There is a large and well-established literature on estimating the parameters of a Bayesian Network from empirical data about variables represented within it. [26, 64] We guide the adaptive regulator to draw on these standard machine learning techniques when integrating real-time monitoring into privacy regulation.

Researchers have devised ways of tracking information flows directly, using computational methodologies. For example by logging the information flows conducted by browser cookies [21], or conducting dynamic analysis of how mobile apps handle personal data [52]. But for other flows, such as business to business flow, it can be harder to collect relevant data without participation from the private sector. There have been many studies guided by CI that use surveys to measure consumer expectations of privacy. These measurements may or may not agree with the flows indicated by privacy policies. These privacy policies are arguably more descriptive than normative, because consumers have poor comprehension of these policies, but they are written to prevent liability of corporations concerning their use of collected personal data.

In addition to tuning the parameters of the BN (and MAIM) to data, the fully specified BN can also be used to determine the likelihood of observed data. In some cases, the data may invalidate the model. For example, if two variables that are according to the model independent are in fact highly correlated, then that suggests that there is an information flow that is missing from the analysis.

5.2.1 Example: Cambridge Analytica Monitoring. In LC #1, the regulators determined that to preserve voter autonomy, there should be no information flow of social media activity to political advertisers. Social media companies register their APIs, and political advertisers report that they are not using any prohibited data. However, real-time monitoring detects a problem. Some select advertisers show significantly higher revenues U^a for segments of the population that are known to have much higher social media use m. Under the normative model, these variables should be independent, and so this triggers further scrutiny. Investigators learn that indeed the choice of advertising campaign D^a is also correlated with specific social media behavior. A formal investigation of noncompliance is initiated.

6 LEARNING CYCLE #3 - THE VALIDITY OF REGULATORY INSTRUMENTS

The validity of CI-based regulations should be regularly tested for as well [66]. On top of identifying new risks and monitoring for privacy violations, regulators should test for the validity of their own regulations and ensure that they serve their purpose. With constantly changing usages and applications of data, and a controversial history of tech companies tweaking the impact of existing privacy policies, regulators should be on top of understanding whether their designed rules promote meaningful compliance in dynamic information environments. To assist in the implementation of the third-learning cycle, the validity and compliance behavior of regulated industries could be evaluated through dedicated whistleblowers. Brought up in previous work, additional protections for whistleblowers could help regulators address corporate secrecy [12], learning how well-intentioned regulatory instruments turn into procedural checkboxes, far from promoting social purposes in the respective information context.

To test for the validity of existing regulatory instruments, regulators should assess to what extent the chosen transmission principle advances the purposes of a given context. For instance, they should run surveys among consumers

to check whether notice and consent is a valid transmission principle that respects the desired norms for the usage of social media and contribute to the social purpose for which the context exists. In the Cambridge Analytica Affair, such validity assessment of regulatory instruments was missing, as users who chose to participate in the personality survey were in fact very far from understanding the manipulation that their consent might allow. Moreover, in the wake of this scandal, important questions about automated and targeted political communications around elections have raised many empirical and normative questions [29]. In this cycle, the original operationalization of autonomy used by the regulators, as well as other model assumptions, would be assessed for validity based on ongoing scholarship and public debates.

7 LIMITATIONS AND DISCUSSION

Shifting the regulatory object from data to information flows requires modeling, empirical calibration, and consensus among stakeholders that are all hard to achieve in practice. Challenges include popularizing a CI-oriented mindset among regulators and empowering them legally to act accordingly. Regulating privacy as CI requires a will towards a positive, consensus view of what appropriate information flow means in each social sphere. Given the fraught political realities in many advanced democracies, we are conscious that what we are proposing is an ideal type of a regulatory model for privacy. Building context models with strategic equilibria and endogenous information flows, when these models are known to have large state spaces and high computational complexity, is something regulators are not used to doing. Such paradigm shift among privacy regulators will take time. The process of teaching regulator about CI may cause them to ask questions and recognize blind spots in their regulatory efforts.

A second challenge is the low level of transparency and openness of technology companies, who often experience conflict between their business interests and their willingness to meaningfully respect the privacy of their consumers [17]. Not all relevant data about active information flows can be easily detected. To properly monitor dynamic information flows, regulators will need to closely collaborate with researchers or adopt 'civic technologies' to assess how information norms are followed. We do see sporadic examples of these [59], but a more coordinated effort to monitor tech in real-time, backed by public budgets and cooperation between different stakeholders is needed. Enabling this crucial information gathering process may be the first substantive regulation to allow adaptive regulatory processes and signal the objects to be regulated.

Importantly, to be consistent with CI, all information flows about data subjects between companies and regulators must be *appropriate*. While gathering information to inform regulation in the public interest is perhaps broadly speaking a legitimizing basis for processing personal data, CI-informed regulators may want to take special care that personal data, if collected, is transferred according to the correct transmission principles. In some cases, it may be best for the regulator to use anonymized or otherwise privacy-enhanced data. Likewise, there may be cases where the information flow models produced by regulators shall be exposed to public scrutiny, but in other cases this information may be kept confidential. We leave this aspect of the regulatory design as an open problem.

Lastly, we have presented only a simple and speculative application of Regulatory CI that implicitly promotes a co-regulation model [28], but leaves much to the imagination. Our proposal raises many more technical questions than it answers. We leave further articulation of Regulatory CI and analysis of its feasibility to future work.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants No. 2131532 and No. 2131532. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors

and do not necessarily reflect the views of the sponsors. One of the authors of this article was supported by the New York University Information Law Institute's Fellows program, which is funded in part by Microsoft Corporation.

REFERENCES

[1] Rediet Abebe, Solon Barocas, Jon Kleinberg, Karen Levy, Manish Raghavan, and David G Robinson. 2020. Roles for computing in social change. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 252–260.

[2] Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, and Nick Feamster. 2018. Discovering smart home internet of things privacy norms using contextual integrity. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 2, 2 (2018), 1–23.

[3] Ero Balsa, Helen Nissenbaum, and Sunoo Park. 2022. Cryptography, Trust and Privacy: It's Complicated. In *Proceedings of the 2022 Symposium on Computer Science and Law*. 167–179.

[4] Solon Barocas and Helen Nissenbaum. 2014. Big data's end run around anonymity and consent. *Privacy, big data, and the public good: Frameworks for engagement* 1 (2014), 44–75.

[5] Adam Barth, Anupam Datta, John C Mitchell, and Helen Nissenbaum. 2006. Privacy and contextual integrity: Framework and applications. In *2006 IEEE symposium on security and privacy (S&P'06)*. IEEE, 15–pp.

[6] Lori S Bennear and Jonathan B Wiener. 2019. Adaptive regulation: instrument choice for policy learning over time. *Obtenido de Universidad de Harvard: https://www. hks. harvard. edu* (2019).

[7] Sebastian Benthall. 2019. Situated information flow theory. In *Proceedings of the 6th Annual Symposium on Hot Topics in the Science of Security*. 1–10.

[8] Sebastian Benthall and Rachel Cummings. 2024. Integrating Differential Privacy and Contextual Integrity. In *Proceedings of the 2024 ACM Symposium on Computer Science and Law*.

[9] Sebastian Benthall and Jake Goldenfein. 2021. Artificial intelligence and the purpose of social systems. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 3–12.

[10] Sebastian Benthall, Seda Gürses, Helen Nissenbaum, et al. 2017. *Contextual integrity through the lens of computer science*. Now Publishers.

[11] Thomas A Birkland. 1998. Focusing events, mobilization, and agenda setting. *Journal of public policy* 18, 1 (1998), 53–74.

[12] Hannah Bloch-Wehba. 2023. A Public Technology Option. *Law and Contemporary Problems* 86, 3 (2023), 222–255.

[13] Irina Brass and Jesse H Sowell. 2021. Adaptive governance for the Internet of Things: Coping with emerging security risks. *Regulation & Governance* 15, 4 (2021), 1092–1110.

[14] Carole Cadwalladr and Emma Graham-Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian.com* (March 2018). https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

[15] Alycia N Carey and Xintao Wu. 2022. The causal fairness field guide: Perspectives from social and formal sciences. *Frontiers in Big Data* 5 (2022), 892837.

[16] Julie E Cohen. 2019. Turning privacy inside out. *Theoretical inquiries in law* 20, 1 (2019), 1–31.

[17] Julie E Cohen and Waldman Ari Ezra. 2023. Introduction: Framing Regulatory Managerialism as an Object of Study and Strategic Displacement. *Law and Contemporary Problems* 86, 3 (2023).

[18] Anupam Datta, Jeremiah Blocki, Nicolas Christin, Henry DeYoung, Deepak Garg, Limin Jia, Dilsun Kaynar, and Arunesh Sinha. 2011. Understanding and protecting privacy: Formal semantics and principled audit mechanisms. (2011).

[19] A Philip Dawid. 2002. Influence diagrams for causal modelling and inference. *International Statistical Review* 70, 2 (2002), 161–189.

[20] Cynthia Dwork. 2006. Differential privacy. In *International colloquium on automata, languages, and programming*. Springer, 1–12.

[21] Steven Englehardt and Arvind Narayanan. 2016. Online Tracking: A 1-Million-Site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (Vienna, Austria) *(CCS '16)*. Association for Computing Machinery, New York, NY, USA, 1388–1401. https://doi.org/10.1145/2976749.2978313

[22] Tom Everitt, Ryan Carey, Eric Langlois, Pedro A Ortega, and Shane Legg. 2021. Agent incentives: A causal perspective. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence,(AAAI-21). Virtual. Forthcoming*.

[23] James Fox, Tom Everitt, Ryan Carey, Eric Langlois, Alessandro Abate, and Michael Wooldridge. 2021. PyCID: A Python Library for Causal Influence Diagrams.

[24] Naman Goel, Alfonso Amayuelas, Amit Deshpande, and Amit Sharma. 2021. The importance of modeling data missingness in algorithmic fairness: A causal perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 7564–7573.

[25] Lewis Hammond, James Fox, Tom Everitt, Alessandro Abate, and Michael Wooldridge. 2021. Equilibrium Refinements for Multi-Agent Influence Diagrams: Theory and Practice. *arXiv preprint arXiv:2102.05008* (2021).

[26] David Heckerman, Dan Geiger, and David M Chickering. 1995. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning* 20 (1995), 197–243.

[27] Mireille Hildebrandt. 2013. Slaves to Big Data, or Are We? *IDP Revista de Internet* (2013).

[28] Dennis D. Hirsch. 2011. The Law and Policy of Online Privacy: Regulation, Self-Regulation, or Co-Regulation. *SEATTLE U. L. REV* 34 (2011). Issue 3.

[29] Philip N Howard, Samuel Woolley, and Ryan Calo. 2018. Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *Journal of information technology & politics* 15, 2 (2018), 81–93.

[30] Severin Kacianka and Alexander Pretschner. 2021. Designing accountable systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 424–437.

[31] Daphne Koller and Brian Milch. 2003. Multi-agent influence diagrams for representing and solving games. *Games and economic behavior* 45, 1 (2003), 181–221.

[32] Christel Koop and Martin Lodge. 2017. What is regulation? An interdisciplinary concept analysis. *Regulation & Governance* 11, 1 (2017), 95–108.

[33] Bogdan Kulynych, Rebekah Overdorf, Carmela Troncoso, and Seda Gürses. 2020. POTs: protective optimization technologies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 177–188.

[34] David Levi-Faur. 2011. Regulation and regulatory governance. *Handbook on the Politics of Regulation* 1, 1 (2011), 1–25.

[35] Charles E Lindblom. 1959. The Science of "Muddling Through". *Public Administration Review* 19, 2 (1959), 79–88.

[36] Natasha Lomas. 2022. *Adtech's compliance theatre is headed to Europe's top court.* https://techcrunch.com/2022/09/07/iab-europe-tcf-gdpr-breach-appeal/

[37] David Madras, Elliot Creager, Toniann Pitassi, and Richard Zemel. 2019. Fairness through causal awareness: Learning causal latent-variable models for biased data. In *Proceedings of the conference on fairness, accountability, and transparency*. 349–358.

[38] Kirsten Martin. 2018. The penalty for privacy violations: How privacy violations impact trust online. *Journal of Business Research* 82 (2018), 103–116.

[39] Lawrence E McCray, Kenneth A Oye, and Arthur C Petersen. 2010. Planned adaptation in risk regulation: An initial survey of US environmental, health, and safety regulation. *Technological Forecasting and Social Change* 77, 6 (2010), 951–959.

[40] Aleecia M McDonald and Lorrie Faith Cranor. 2008. The cost of reading privacy policies. *ISJLP* 4 (2008), 543.

[41] Lee Mcguigan. 2023. *Selling the American People: Advertising, Optimization, and the Origins of Adtech.* MIT Press.

[42] Lee J McGuigan, Sarah Myers West, Ido Sivan-Sevilla, and Patrick Parham. 2023. The after party: Cynical resignation in Adtech's pivot to privacy. *Big Data Society* 10, 2 (2023). https://doi.org/10.1177/20539517231203665

[43] Alan Mislove. [n. d.]. Red-Teaming Large Language Models to Identify Novel AI Risks. *OSTP Blog* ([n. d.]). https://www.whitehouse.gov/ostp/news-updates/2023/08/29/red-teaming-large-language-models-to-identify-novel-ai-risks/

[44] Deirdre K Mulligan, Colin Koopman, and Nick Doty. 2016. Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374, 2083 (2016), 20160118.

[45] Helen Nissenbaum. 2009. *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford University Press.

[46] Helen Nissenbaum. 2019. Contextual integrity up and down the data food chain. *Theoretical Inquiries in Law* 20, 1 (2019), 221–256.

[47] Helen Nissenbaum, Sebastian Benthall, Anupam Datta, Michael C Tschantz, and Piot Mardziel. 2018. *Origin privacy: Protecting privacy in the big-data era.* Technical Report. NEW YORK UNIVERSITY.

[48] Helen Nissenbaum, Katherine Strandburg, and Salome Viljoen. 2024. The Great Regulatory Dodge. *Harvard Journal of Law and Technology* (2024).

[49] Kobbi Nissim and Alexandra Wood. 2018. Is privacy privacy? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2128 (2018), 20170358.

[50] Nikolaus Poechhacker and Severin Kacianka. 2021. Algorithmic accountability in context. Socio-technical perspectives on structural causal models. *frontiers in Big Data* 3 (2021), 519957.

[51] Michael Reddy. 1979. The conduit metaphor. *Metaphor and thought* 2 (1979), 285–324.

[52] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman. 2018. "Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale. *Proceedings on Privacy Enhancing Technologies* 2018 (06 2018), 63–83. https://doi.org/10.1515/popets-2018-0021

[53] Matthew Rosenberg, Nicholas Confessore, and Carole Cadwalladr. 2018. How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times* (March 2018). https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

[54] Adam Satariano and Christine Haiser. [n. d.]. In Europe, Meta Offers Ad-Free Versions of Facebook and Instagram for First Time. *The New York Times* ([n. d.]). https://www.nytimes.com/2023/10/30/technology/facebook-meta-subscription-europe.html

[55] Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*. 59–68.

[56] Ross D Shachter. 1988. Probabilistic inference and influence diagrams. *Operations research* 36, 4 (1988), 589–604.

[57] Yan Shvartzshnaider, Noah Apthorpe, Nick Feamster, and Helen Nissenbaum. 2019. Going against the (appropriate) flow: A contextual integrity approach to privacy policy analysis. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. 162–170.

[58] Yan Shvartzshnaider, Madelyn Rose Sanfilippo, and Noah Apthorpe. 2022. GKC-CI: A unifying framework for contextual norms and information governance. *Journal of the Association for Information Science and Technology* 73, 9 (2022), 1297–1313.

[59] Ido Sivan-Sevilla. 2023. To Save Society from Digital Tech, Enable Scrutiny of How Policies Are Implemented. *Issues in Science and Technology* 39, 4 (July 2023), 28–30. https://doi.org/10.58875/RARJ9814

[60] Ido Sivan-Sevilla, Helen Nissenbaum, and Patrick Parham. 2022. Public Comment for FTC's Commercial Surveillance ANPR. (Nov 2022). https://doi.org/10.31219/osf.io/wjr5z

[61] Ido Sivan-Sevilla and Lior Zalmanson. 2023. Information Systems in the Service of Adaptive Climate Regulatory Response. *Innovations in Management* 13 (2023), 56–61.

[62] Daniel Susser, Beate Roessler, and Helen Nissenbaum. 2019. Technology, autonomy, and manipulation. *Internet Policy Review* 8, 2 (2019). https://doi.org/10.14763/2019.2.1410

[63] Linnet Taylor, Luciano Floridi, and Bart Van der Sloot. 2016. *Group privacy: New challenges of data technologies*. Vol. 126. Springer.
[64] Simon Tong and Daphne Koller. 2000. Active learning for parameter estimation in Bayesian networks. *Advances in neural information processing systems* 13 (2000).
[65] Salome Viljoen. 2021. A relational theory of data governance. *Yale LJ* 131 (2021), 573.
[66] Ari Ezra Waldman. 2019. Privacy Law's False Promise. *Wash. UL Rev.* 97 (2019), 773.
[67] Michael Walzer. 1983. *Spheres of justice: A defense of pluralism and equality*. Basic Books.
[68] Barg S. Tyler S. Venema H. Tomar S. Bhadwal S. Nair S. Roy D. Drexhage J wanson, D. 2010. Seven tools for creating adaptive policies. Technological Forecasting and Social Change. *Technological Forecasting and Social Change* 77, 6 (2010), 924–939.
[69] Alan F Westin. 1968. Privacy and freedom. *Washington and Lee Law Review* 25, 1 (1968), 166.
[70] Wanrong Zhang, Olga Ohrimenko, and Rachel Cummings. 2020. Attribute Privacy: Framework and Mechanisms. *arXiv preprint arXiv:2009.04013* (2020).
[71] Shoshana Zuboff. 2023. The age of surveillance capitalism. In *Social Theory Re-Wired*. Routledge, 203–213.