Deficiencies in Online Privacy Policies: Factors and Policy Recommendations

by

Jesse H. Sowell, II

B.S., Computer Science, Clemson University (2001)
M.S., Computer Science, Michigan State University (2005)
M.S., Criminal Justice, Michigan State University (2007)

Submitted to the Engineering Systems Division
in partial fulfillment of the requirements for the degree of

Master of Science in Technology and Policy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2010

Author .....
Engineering Systems Division
July 8, 2010

Certified by .....
Dr. David D. Clark
Senior Research Scientist, Computer Science and Artificial Intelligence Laboratory
Thesis Supervisor

Accepted by .....
Dr. Dava J. Newman
Professor of Aeronautics and Astronautics and Engineering Systems
Director, Technology and Policy Program

Deficiencies in Online Privacy Policies: Factors and Policy Recommendations

by

Jesse H. Sowell, II

Submitted to the Engineering Systems Division
on July 8, 2010, in partial fulfillment of the
requirements for the degree of
Master of Science in Technology and Policy

Abstract

Online service providers (OSPs) such as Google, Yahoo!, and Amazon provide customized features that do not behave as conventional experience goods. Absent familiar metaphors, unraveling the full scope and implications of attendant privacy hazards requires technical knowledge, creating information asymmetries for casual users. While a number of information asymmetries are proximately rooted in the substantive content of OSP privacy policies, the lack of countervailing standards guidelines can be traced to systemic failures on the part of privacy regulating institutions. In particular, the EU Data Protection Directive (EU-DPD) and the US Safe Harbor Agreement (US-SHA) are based on comprehensive norms, but do not provide pragmatic guidelines for addressing emerging privacy hazards in a timely manner. The dearth of substantive privacy standards for behavioral advertising and emerging location-based services highlight these gaps.

To explore this problem, the privacy policies of ten large OSPs were evaluated in terms of strategies for complying with the EU-DPD and US-SHA and in terms of their role as tools for enabling informed decision-making. Analysis of these policies shows that OSPs do little more than comply with the black letter of the EU-DPD and US-SHA. Tacit data collection is an illustrative instance. OSP privacy policies satisfice by acknowledging the nominal mechanisms behind tacit data collection supporting services that "enhance and customize the user experience," but these metaphors do not sufficiently elaborate the privacy implications necessary for the user to make informed choices. In contrast, privacy advocates prefer "privacy and surveillance" metaphors that draw users attention away from the immediate gratification of customized services. Although OSPs do bear some responsibility, neither the EU-DPD nor the US-SHA provide the guidance or incentives necessary to develop more substantive privacy standards.

In light of these deficiencies, this work identifies an alternative, collaborative approach to the design of privacy standards. OSPs often obscure emerging privacy hazards in favor of promoting innovative services. Privacy advocates err on the other side, giving primacy to "surveillance" metaphors and obscuring the utility of information based services. Rather than forcing users to unravel the conflicting metaphors, collaborative approaches focus on surfacing shared concerns. The collaborative approach presented here attempts to create a forum in which OSPs, advertisers, regulators,

and civil society organizations contribute to a strategic menu of technical and policy options that highlight mutually beneficial paths to second best solutions. Particular solutions are developed through a process of issue (re)framing focused on identifying common metaphors that highlight shared concerns, reduce overall information asymmetries, and surface the requirements for governance and privacy tools that address emerging risks. To illustrate this reframing process, common deficiencies identified in the set of privacy policies are presented along with strategic options and examples of potential reframings.

Thesis Supervisor: Dr. David D. Clark
Title: Senior Research Scientist, Computer Science and Artificial Intelligence Laboratory

Proposal

Innovative secondary uses of consumer information, most commonly for customization of services, is a key competitive differentiator for online service providers (OSPs) such as Google, Amazon, and Facebook. Unless harms can be clearly demonstrated, reactive privacy regulatory environments such as the US accommodate secondary uses as a necessary component of technological innovation. In contrast, the EU's more stringent privacy regulations proactively espouse socially protective policies that prioritize individual privacy as a fundamental human right. Among other transnational conflicts, the Internet milieu brings these privacy paradigms formally insulated by terrestrial boundaries into increasing conflict. Most recently[1], the EU Data Protection Directive's attempt to protect EU citizens' privacy created economic externalities that threatened to halt trans-Atlantic dataflows to countries deemed to have "inadequate" privacy regimes (in particular, the US). Despite international agreements on FIPs as guiding principles, substantially different policy instruments arise when interpreted and implemented under different governance paradigms [17, 18][2]. Initial reactions from the legal and political science literature indicate the EU Data Protection Directive is a failed attempt at policy convergence [19, 77] and the compromise that resolved the conflict, the US Safe Harbor Agreement, is an empty formalism that actually exacerbates and prolongs the conflict [103][3].

Given this context and motivation, the thesis will evaluate whether the US Safe Harbor Agreement is in fact an empty formalism and evaluate its efficacy as a hybrid privacy regime. Empirical evidence will be drawn from a substantive content analysis of ten major OSP privacy policies[4]subject to the US Safe Harbor Agreement. A key trade-off is the tension between the economic efficiency of OSPs and privacy controls required by regulation [101]. To understand the dynamics of this trade-off, the thesis

[1]Westin [131] provides a nice, succinct history of international information privacy conflicts that have occurred since the mid-1960's.

[2]Although Bennett's survey of privacy policy instruments in the US, UK, West Germany, and Sweden was performed before the EU Data Protection Directive came into effect, the categorization of policy convergence strategies and the source of policy instrument differences is both valid and useful for describing the critical path through a governance structure from abstract FIPs to a concrete policy instrument. Bennett's original work [17] is cited extensively as a seminal work on differences in governance structures [] and is followed by [18].

[3]Reidenberg is the only author to flat out indicate the US Safe Harbor agreement is an empty formalism. Others [19, 71, 77, 101] perform fairly impartial analyses that imply it is an empty formalism but do not say so explicitly.

[4]As of 30 July 2009 these ten policies covered 18 of the top 40 web sites ranked by Alexa. Four of the top five are covered by the privacy policies of Google, Yahoo!, Facebook, and Microsoft.

will explore how OSPs, based on their publicly posted privacy policies, conform to the fair information practices (FIPs) set out by US Safe Harbor Agreement, will evaluate these privacy policies as decision making tools for consumers, and elaborate the attendant conflicts that arise from differences in interpretation and operationalization of privacy concepts and assumptions under conventional and hybrid privacy regimes. A preliminary analysis of OSP privacy policies indicates compliance with the *black letter* of the US Safe Harbor Agreement FIPs, but only in the weakest (broadest) interpretation. In other words, OSPs' privacy policies, especially those of OSPs vested in the less stringent US privacy paradigm, deviate from stronger interpretations preferred by EU states. It is expected that a deeper analysis will provide empirical evidence that the US Safe Harbor Agreement is an empty formalism.

The tension between OSPs' economic efficiency and privacy controls is exacerbated when a FIPs operationalization favors one group of stakeholders over another. OSPs' claim that privacy advocates' strong interpretations of privacy regulations create "onerous" technical and management burdens is an example of an argument against an operationalization favoring strong consumer privacy. More recent arguments by OSPs highlight advertising, most recently customization via behavioral advertising, as the essential factor in the economics of "free" online content. This latter argument is still one of economic efficiency versus individual privacy, but narrows the discussion to the information asymmetries identified in the preliminary analysis of OSPs' privacy policies. The content analysis takes these arguments into account and will illustrate (1) the nominal data categories and uses of consumer information and (2) how information asymmetries hobble the self-regulatory enforcement espoused by the US Safe Harbor Agreement.

Finally, the criticisms of the US Safe Harbor Agreement as a hybrid privacy regime are not intended to condemn hybrid privacy instruments as a whole. Rather, these criticisms are intended to highlight the implications and effects of the US Safe Harbor Agreement as the negative results of a privacy policy experiment. Lessons learned are expected to inform how broadly-scoped privacy concepts are interpreted and the detrimental implications of the empty formalism for innovations that can better surface actual privacy preferences. The thesis intends to proffer attendant recommendations and design implications that inform stakeholders of the importance of precise privacy standards regarding choice and subsequent implications. Perhaps more importantly to this adversarial regulatory space, the thesis also intends to recommend characteristics of a framework for incentivizing the consistent, sustainable development of standards through collaborative processes.

Acknowledgments

This thesis concludes two years in the Technology and Policy Program (TPP) focusing on online privacy policy. I am indebted to the Advanced Network Architecture (ANA) group, in particular my advisor, Dr. David Clark, for advice and direction. Dr. Clark provided excellent advice, insights, and references as this work developed from a simple survey of privacy policies into what is hopefully a contribution to the literature on privacy regulation. Dr. Clark patiently helped me unravel many of the interesting issues in this work. I would also like to thank Dr. Frank Field for support as the TA for ESD.10 and for allowing me to tack thesis topic feedback and discussion onto the end of TA meetings. In addition to the support of the ANA, I would also like to thank Daniel Weitzner and the Decentralized Information Group for their early support and feedback on this project. I would also like to thank Jim Rice and Mahender Singh of the Center for Transportation and Logistics for supporting my first year in TPP.

Contents

List of Tables

Chapter 1

Introduction

> [T]he primary civil-liberties issues raised by the uses of computers are not matters of information security to be resolved by technical specialists but are instead policy choices as to privacy, confidentiality, and due process.
>
> *Alan Westin and Michael Baker, Databanks in a Free Society [132]*

As online services such as search engines, social networking platforms, and online marketplaces increasingly mediate individuals' daily lives, extant and emerging privacy risks become more noteworthy. Some features of online services do not behave as conventional experience goods[1], requiring technical knowledge to unravel the full scope and implications of attendant privacy hazards. For the casual, non-technical user, the result is an information asymmetry that limits users' ability to identify and reason about certain categories of privacy hazards. Ideally, privacy hazards and their implications should be clearly and prominently explicated in online service provider's privacy policies. The dearth of privacy standards attendant with tacit data collection performed by online service providers is an illustrative instance of a privacy policy information asymmetry. The motivated, technically proficient user can evaluate and understand tacit data collection mechanisms to derive a more comprehensive set of privacy implications[2], but the signals and experience necessary to recognize these implications are not a prominent part of the casual user's experience.

This thesis identifies factors contributing to deficiencies in privacy policies that are proximately rooted in the content of online service provider privacy policies, but that can also be traced to the systemic implications of institutional design decisions.

[1]Nelson [88] contrasts search goods with experience goods. For *search* goods, inspection is sufficient to determine the utility of a particular good. If inspection before purchase (here engagement) is not sufficient, if it pays to evaluate by purchasing the good and the price of the good is sufficiently low, the process of searching for the good is less value. The process of evaluating through purchase is what Nelson calls "experience" and goods that are more effectively evaluated by experiencing one or more purchases is a conventional experience good. The assumption of conventional experience goods is that the consumer (here the user) can, based on their possibly limited technical capabilities, trace the effects of the good back to a particular experience or set of experience with the good and understands the implications of purchasing the good.

[2]Following the description of an experience good, this is an instance of a costly search process that is much greater than simply purchasing (engaging) the service.

Empty formalism is evident both proximately in online service provider privacy policies and in the application of the US Safe Harbor to online service provider privacy policies. Walker's articulation and reflection on empty formalism highlights the interaction between policy and institutional failures:

> Do rules degenerate into an empty ritual which respects the letter of the law but not the spirit of justice? Since the answer may not be a simple yes or no, a more subtle way of posing the question might be: How empty is empty formalism? [128]

This work hopes to answer the subtler question by understanding the tactics evident in online service provider's privacy policies, the immediate implications of these tactics, and how the failures of institutions charged with regulating online privacy perpetuate empty formalisms.

A qualitative analysis of a ten online service provider privacy policy sample[3] provides evidence of proximate empty formalism. The deficiencies identified are evident in the technical components of the policies, but are arguably technology agnostic and need be addressed by policies at the institutional level. Nominally, privacy policies are compliant with the black letter of the imprecise norms of the US Safe Harbor. Despite this compliance, proximate information asymmetries in privacy policies, such as those related to tacit data collection, obscure privacy hazards from the casual user. The general failure of the regulatory institution, here the US Safe Harbor, is that it provides only vague normative guidance. In particular, the US Safe Harbor does not establish processes for developing and assessing evolving privacy standards necessary to proactively identify new privacy hazards. Both the failed policy outcomes (policy failures) and the abstract privacy norms set out in the US Safe Harbor can be (and are argued in this thesis to be) empty formalisms[4].

The institutional design decisions contributing to the US Safe Harbor are products of a recent compromise in the ongoing conflict over privacy regulatory regimes between the US and the EU. In October of 1998 the EU Data Protection Directive came into effect. The nominal objective is to establish privacy as a fundamental human right for EU citizens by harmonizing EU states' privacy regulations based on the Fair Information Practices (FIPs)[5]. The EU ruled that the US privacy regulatory regime lacked both the comprehensive laws and centralized enforcement regime necessary to adequately protect EU citizens' data handled by US-based online service providers. The resulting compliance externality threatened the suspension of trans-Atlantic data flows worth approximately $120 bn [64].

The US developed the US Safe Harbor as a compromise. The US Safe Harbor is a hybrid privacy regulatory regime that articulates a comprehensive privacy regime

[3]The detailed analysis supporting black letter compliance is presented in Appendix C.

[4]For clarity, the individual instances and categories of empty formalisms in online service provider's privacy policies will be referred to as *policy failures*; empty formalism at the institutional level will be referred to as *institutional failures.*

[5]The FIPs are the product of the OECD, COE, and the Younger Committee, and the US Department of Commerce attempting to identify a common, comprehensive set of privacy principles. See Chapter 2 for histories of the FIPs in the US and EU; Section 2.1 articulates the six fundamental principles based on Bennett's summary in [17]

based on the FIPs, but that adopts a market-based (decentralized), self-regulatory enforcement mechanism. Initial reactions from the legal and political science literature indicate the EU Data Protection Directive is a failed attempt at policy convergence [19, 77] and that the US Safe Harbor Agreement is an empty formalism that actually exacerbates and prolongs the underlying normative conflict [103][6]. With respect to providing constructive guidance for the development of privacy standards for online service providers, the deficiencies identified here highlight a lack of standards guidance in the US Safe Harbor.

The privacy deficiencies presented here in online service provider privacy policies as evidence of the US Safe Harbor as an empty formalism with respect to providing constructive guidance for the development online service provider privacy policies and regulatory enforcement. More precisely, the policy analysis demonstrates how portions of online service provider privacy policies are negative outcomes that satisfice to the overly abstract privacy norms articulated by the US Safe Harbor Agreement and obscure relevant privacy implications. Black letter compliance with vague data purposes also blurs the threshold of policy compliance. Coupled with the absence of regulatory guidance on the precision of data purposes, the institutional failure is further aggravated by undermining self-regulatory mechanisms based on providing evidence of harms based on clear-cut privacy violations.

Tracing policy deficiencies to the contributing institutional factors highlights institutional design failures, but also highlights areas of improvement at the institutional and technical level. Although the deficiencies identified here can be resolved with a reactive regulatory decisions[7] and complementary technical changes. Such incrementalist approaches are quick fixes that, like the US Safe Harbor, provide a false sense of resolution (empty formalism) that obscures solutions to emerging problems. Albeit critical of the US Safe Harbor, this thesis is not a wholesale condemnation of hybrid self-regulatory frameworks. Rather, Chapter 5 proposes approaches supporting institutional and technical design features[8] that can facilitate the continuous feedback necessary for the concurrent design of privacy policies and online service features that support these policies.

The proposed methods are based on a synthesis of Schön's notion of generative metaphor [106] and Antón's Goal Based Requirements Analysis Methodology (GBRAM) to privacy policies [6, 9, 10]. Schön's generative metaphor is applied to (re)frame privacy problems. As currently framed, largely technical descriptions satisfice to abstract norms and create information asymmetries. The reframing process is expected to give insights into metaphors meaningful to both designers and lay users. The result is a meaningful policy with fewer knowledge barriers. The reframing process also surfaces insights previously occluded by one-sided metaphors such as these

[6]Reidenberg [103] is the only author to directly indicate the US Safe Harbor agreement is an empty formalism. Others [19, 71, 77, 101] imply it is an empty formalism but do not say so explicitly.

[7]One-off, reactive, and chaotic are all adjectives used to describe the US privacy regulatory regime; see Section 2.3 for a historical overview.

[8]Institutional design decisions, as implied, apply to the design of institutions. Technical design decisions apply to the design of privacy tools available to users by online service provider or tools provided by third parties that supplement engaging online services.

focusing on "service-and-utility" or "privacy versus surveillance." These insights also contribute to the development of supporting tools.

Antòn's GBRAM is a requirements analysis methodology based on grounded theory that is used to surface requirements from organizational artifacts (such as privacy policies). The product of GBRAM is structured natural language representations that facilitate comparison and contrast of system goals. Structured natural language representation avoids overly technical requirements specifications by preserving the vernacular of organizational artifacts, allowing non-technical actors to participate in effective identification of redundancies and conflicts in system goals. This thesis proposes a process for using GBRAM to surface hidden, common goals amongst conflicting actors based on the negotiation and identification of conflicting metaphors. Schön's generative metaphor is applied to reframe these into metaphors that highlight commonalities, overcome information asymmetries, and provides insights into the design of privacy tools.

The proposed institutional design decisions intend to incentivize technical design processes to use feedback about context-based differences in privacy preferences. Such feedback can provide valuable feedback for surfacing emergent privacy risks. Coupled with third party monitoring of privacy compliance, it is argued that such sociotechnical relationships can more efficaciously surface new privacy risks rather than relying on policies and tools that satisfice to abstract static norms. This proposal does not choose winners or losers by suggesting a particular configuration of actors and specific technical approaches. Rather, this work identifies and frames the design choices that affect the mutual advantages possible among regulators, online service providers, and ultimately, end users.

1.1 Privacy Policy Failures

Ideally, users rely on an understanding of the online service provider's privacy policy to decide whether the online service provider's data purposes are acceptable relative to the user's personal privacy preferences[9]. Information asymmetries arise from vague and abstract articulations of practices that do not provide precise information on how all information collected is used. As a decision making tool, the policy fragments[10] addressing tacit data collection exacerbate information asymmetry be-

[9]Economic privacy policy analysis couches policy analysis in a cost-benefit analysis, viewing policies as tools for deciding whether an online service's privacy practices are compatible with a user's privacy preferences and utility gained by sharing information [1, 2, 3, 16, 20, 29, 62, 67]. This thesis follows these analyses by conceiving policies as decision making tools, but focuses on overcoming knowledge barriers (formally information asymmetries) to efficaciously convey implications rather than focusing on the cost-benefit trade-offs of commoditized information.

[10]A policy fragment is the set of statements that address the same policy concern. In this case, the policy fragment is the set of all statements salient to describing tacit data collection methods. A policy fragment may be contiguous or non-contiguous. As may be implied, contiguous fragments comprise a sequence of adjacent statements that address a particular concern. A non-contiguous fragment indicates the statements describing a concern are not adjacent and may be scattered throughout the policy document or across multiple policy documents.

cause they only provide nominal technical information about methods and technical data. The average user does not have the technical expertise to translate technical policy fragments into meaningful implications or to map the processes described to meaningful experiences that can be used to make decisions relative to individual privacy preferences. Following the framing as an experience good, the process necessary for a technically proficient reader to surface privacy implications and persistent privacy deficiencies is an extensive search process. In terms of metaphor, this is especially evident when comparing the intuitions embodied in descriptions of e-mail correspondence, instant message-style chat tools, and bulletin-board posting policies with unintuitive technical discussions covering tacit data collection such as the use of web beacons, client-side scripting, and cookies. Existing analyses of privacy policies indicate they are all generally at the same reading level [8]. The analysis in Chapter 4 confirms, modulo online service provider-specific domain, privacy policies convey essentially the same normative content. More interestingly, the analysis presented here identifies factors that contribute to two pervasive problems: mixed context and the aggregate image. Efficacious resolution of these problems will require both proximate policy redesign and institutional redesign to avoid one-off, incrementalist solutions that only address technical issues. In general, as noted above, although the two problems identified here are surfaced via an analysis of current technical infrastructure, they are arguable technology agnostic, sociotechnical privacy problems whose social and market components facilitate persistence across evolving online technologies.

Mixed context is the first problem. Mixed context occurs when users are confronted with a superficially seamless environment that is actually comprised of objects from multiple sources and is similarly governed by a dynamic set of distinct, possibly conflicting privacy policies. This can be ascertained from privacy policies, but requires close inspection and an understanding of the underlying technical components[11]. The technical knowledge necessary for this analysis contributes to the information asymmetry. The privacy deficiency is the tractability of any user making privacy decisions in the face of mixed context. Pragmatic resolutions requires coordination amongst online service providers, advertisers, advertising networks, civil society organizations, and online application developers[12].

The aggregate image problem is the second pervasive privacy deficiency. The notion of aggregate image relates to whether a collection of information is considered personally identifying. Privacy policies comply with black letter requirements in that they acknowledge data about visitors is collected automatically, but do not meaningfully articulate the implications of how that information is (or may be) (re)combined. Although individual items such as age and interests may be innocuous when taken individually, when combined they may constitute a very accurate image of the individual user. Again, the information asymmetry is the unraveling of technological methods to surface the problem. The deficiency is the lack of meaningful articulation of the hazards of tacit data sharing and the implications of recombination. Although

[11]Once again, following the experience good framing, this is an extensive search process.

[12]Section 5.4.1 describes the application of the proposed analysis framework (based on Schön and Antòn, Section 5.2) to the mixed context problem.

the debate over whether this image is personally identifying is not novel[13], it is independent of the underlying technology and arguably a persistent issue[14] that must be addressed by policy rather than by incremental, technology specific interventions.

Given the failings of the privacy policies surveyed here, this thesis recommends the concurrent design of privacy policies and online environments that facilitate explicit, bi-directional feedback from user to online service provider and from online service provider to the user. The former recommends online service providers incorporating tools into user workflows that allow feedback regarding the types of data collection applied in these environments. The latter is a signal from the online service provider to the user, based in efficacious metaphor, that conveys to the user the implications of engaging a particular environment. Together, these provide the continual feedback necessary to concurrently design efficacious environments whose policies[15] are based on empirically derived articulations of users' privacy preferences. From a design perspective, this is a functionally complete solution; pragmatically, the analysis, design, and allocation of implementation "burden" requires coordination between more than just users and online service providers.

1.2 Institutional Failures

The policy failures described above are partially born of institutional failures. The discussion of policy failures has implicitly implicated institutional failures, but has thus far only considered users and online service providers. One failed approach is to place the burden of resolving privacy dilemmas on a single category of actor—in this case online service providers[16]. One-sided burden gives rise to incrementalist, adversarial[17] solutions that, rather than surfacing constructive reframings, occlude possible compromises and common metaphors. The complete set of actors includes users, online service providers, advertisers, advertising networks, regulators, and civil society organizations. Historically, regulators architect the institutions charged with developing and policing privacy laws and regulations. Here, the EU Data Protection Directive and the US Safe Harbor are presented as institutional arrangements that

[13]Recently, the FTC has addressed issues related to tacit data collection and identity [53, 54]. Earlier incarnations of this debate have addressed the issue in terms of whether there is sufficient information for an individual to be identified, such as Kobsa's discussion in [72].

[14]For instance, the same problems with conventional web browser based cookies have been identified in Flash based cookies [21]. The information collection problem will migrate to mobiles as advertisers pursue mobile-based technologies that synthesize accurate geo-location with explicit and inferred user preferences. As another instance, IPv6 will allow every device to have a permanent, unique identifier (IP). This will allow more effective mapping of behavior to individual, strengthening the argument that IP may be personally identifying.

[15]In particular, the empirical data on the functional categories of information that is and is not private in a given context is critical; see Section 3.2 for a discussion of contextual integrity.

[16]This is common to the larger problem of Internet governance conflicts, such as the dispute between Yahoo! and France. See [139, 73] for extensive discussion.

[17]The notion of adversarial relations amongst actors is in the sense of Jasanoff's deconstruction and reconstruction of technical knowledge [66]. This is revisited in Section 2.4 when reviewing arguments related to the "onerous" burden associated with comprehensive privacy paradigms

highlight the role of regulatory bodies, such as the European Data Authorities and US Department of Commerce, that intend to enforce privacy rights and (nominally) audit online service providers to ensure compliance with applicable law. In both institutional arrangements, failures can be traced to particularistic regulatory design by one category of actors with little input from others.

Conceptually, these problems are rooted in the institutional design process. As per Heisenberg [64], the actors involved in the development of the EU Data Protection Directive were the EU representatives and state enforcement agencies: businesses were noticeably absent. In the US Safe Harbor development, the opposite was the case: the design of the US Safe Harbor was almost exclusively influenced by the Clinton administration's Department of Commerce, relying on US industry for expert guidance. Although both institutions base their normative regulatory tenets on the FIPs, the differences in implementation are characterized in terms of membership, flexibility, and enforcement mechanisms[18]. Succinctly, the EU Data Protection Directive membership is exclusive, flexibility is present but entails lengthy bureaucratic processes, and enforcement is, at least on paper, centralized in state authorities. In contrast, US Safe Harbor membership is inclusive, flexibility is at the discretion of online service provider participants, and enforcement is distributed (decentralized) and "self-regulated."

The EU Data Protection Directive is presented as the impetus of the US Safe Harbor and as a contrast to the US privacy regulatory regime. While this work highlights criticisms and failures in the EU Data Protection Directive, it focuses on tracing policy failures to institutional failures in the US Safe Harbor Agreement. With regards to the US Safe Harbor, the combination of online service provider centered flexibility and enforcement places too much discretion in the hands of online service providers. Under the EU Data Protection Directive, this is resolved in part by monitoring and in part through consultation with state-based data authorities; although good for enforcement, it is argued that this further exacerbates the two-actor problem. As a compromise, the US Safe Harbor attempts to preserve the normatively liberal privacy regulatory regime by implementing enforcement using reactive, self-regulatory mechanisms that handle privacy violations as they occur. These trade-offs in the institutional design space foreshadow their implications for the design of online service provider privacy tools.

As implied earlier, it has been argued that the US Safe Harbor agreement may be abused as an empty formalism. Both Reidenberg [103] and Heisenberg [64] indicate that the US Safe Harbor agreement is an empty formalism, but do not elaborate the mechanisms that give rise to the empty formalism. For completeness, a comparison of online service provider privacy policies with the FIPs articulated by the US Safe Harbor agreement was performed (Appendix C). The analysis in Chapter 4 focuses on the implications for continuously evolving privacy regulations and the commensurate privacy features necessary to implement efficacious privacy policy architectures and instruments. Online service providers' privacy policies are placed in the context of

[18]These parameters are based on Koremenos [74] overview of rational institutional design and are discussed further in Section 5.1.

the letter, intent, and history of the EU Data Protection Directive and US Safe Harbor agreement to better understand the implications of the US Safe Harbor as an instance of a hybrid privacy regime. This thesis will show how information hiding and all-or-nothing tactics identified in the sample contribute to information asymmetries, privacy deficiencies, and ultimately undermine market-based self-regulation necessary to keep pace with changing technologies. These deficiencies provide evidence that the US Safe Harbor agreement, as applied to online service provider privacy policies, is a failed application of unmodified self-regulatory enforcement to a comprehensive privacy regime.

Although the US Safe Harbor agreement is flawed, this analysis is not intended to stigmatize all hybrid combinations of self-regulatory enforcement and comprehensive privacy regulation. Rather, the application of the US Safe Harbor agreement discussed here is treated as a negative result that lends insight into how to design privacy regimes in which market-based reputation can bind as an enforcement mechanism. The policy failures (deficiencies) discussed in Chapter 4 can be couched as outcomes of these institutional failures. This work proposes a framework that will foster continuous, empirically informed analysis and design feedback amongst the full set of privacy actors. Chapter 5 will argue this will foster more efficient allocation of resources and more efficaciously identify emerging privacy conflicts and latent deficiencies. These recommendations suggest governance structures that facilitate third party mediation that can facilitate flexible and sustainable collaborations. Although domain-specific collaborations are certainly necessary, the framework presented will illustrate how GBRAM and generative metaphor can be incorporated into a cooperative governance regime to identify domain-specific operationalizations that align regulatory, industry, and consumer priorities based on shared concerns.

1.3 Roadmap

This analysis is grounded in the texts of online service provider privacy policies that are the product of organizations that must cope with the complexities balancing data sharing, advertising-based revenue, and reconciling these with national and international privacy regulatory regimes. Although the data analysis is bottom-up, the institutional context and attendant privacy theories will be presented top down. Chapter 2 presents an overview of the development and evolution of privacy regulations. This covers privacy practices as they have evolved in the US and the evolution of the Fair Information Practices and the EU Data Protection Directive in the EU.

Chapter 4 presents the privacy policies evaluated, a qualitative analysis of their efficacy as privacy decision making tools, and highlights deficiencies. The major asymmetries this thesis focuses on are rooted in tacit data collection. Other asymmetries include overly vague data categories and purposes. These asymmetries are evidence of satisficing to the black letter of equally vague privacy principles, contribute to imprecise articulations of exactly what constitutes personally identifiable information, and limit clear distinctions between advertiser and online service provider privacy policies and obligations. To highlight the role of policy framing, the implications of

privacy policy deficiencies are described and different framings of the sample policies, given different actors' interests, is presented. These include the distinctions and implications of data sharing, data collection, and surveillance.

Chapter 5 builds on the evidence and implications of Chapter 4 to propose concomitant institutional and online service provider design requirements, methods, and functional objectives that can resolve the kinds of privacy deficiencies identified in the policy sample. Concrete instances based on mixed context and aggregate image deficiencies will describe solutions to the immediate problem. The learnings developed in these instances will be used to propose research directions for developing a feedback framework between regulatory institutions and online service provider's. A number of the theoretical models described in Chapter 3 highlight Nissenbaum's context-based notions of privacy [89] and Solove's family of related concepts to help better develop privacy policies [110]. The feedback framework proposed here could potentially overcome the problems of data collection indicated by Nissenbaum [89, pp. 153–155] and better understand how to translate Solove's family of privacy concepts into requirements that can inform online service provider's design of more effective privacy tools.

Chapter 2

Regulatory Background

Almost without notice and in the name of efficiency our technological progress has moved us toward the 'big brother' supervision predicted in George Orwell's book *1984*.

US Representative Ralph Regula, quoted in Legislating Privacy [102, p. 81]

Conflicts over privacy policy span regulatory paradigms, differences over implementation, and differences over theoretical grounding. Differences in regulatory paradigms[1] set the stage for conflicts amongst regulatory institutions—the EU-US privacy paradigm conflict shaped the EU Data Protection Directive and the US Safe Harbor agreement as institutions. Although both institutions are nominally based on the FIPs as privacy *norms*, neither provide sufficient guidance in implementing privacy standards. Section 2.4 compares and contrasts critiques of these norm-based institutions. Many of the critiques focus on the level of abstraction and the absence of concrete guidance or standards.

Reidenberg's discussion of governance paradigms narrows the distinction amongst western democracies into two groups: normatively liberal and socially protective. One of the key distinctions amongst ostensibly similar western democracies[2] is how much they trust government institutions and the culture and social roots of these beliefs. Reidenberg argues each of these governance paradigms "reflect the country's choices regarding the roles of the state, market, and individual in the democratic structure" [103]. US governance is normatively liberal: private information is

[1]Differences in privacy paradigms and the resultant implementations is most extensively covered by Bennett [17, 18, 19].

[2]For completeness, one should note that Westin also linked privacy to governance. Westin identified privacy as a key element distinguishing totalitarian regimes from democratic regimes [130]. Westin argues that privacy is key to the ability of individuals, organizations, and even some aspects of government to make autonomous decisions without intermediate processes being interrupted or derailed. In many ways, privacy protects each of these decision processes from the heckler's veto, ideally allowing an honest exchange of ideas absent the burden of political ramifications. Along those lines, Westin argues that totalitarian governments impose control by creating an environment devoid of many of the aspects of privacy that allow individuals and groups to make their own decisions without fear of government or political retribution. In contrast, democracies provide various forms of privacy protections that allow individuals and organizations to make decisions.

an alienable commodity and policies dictating how this commodity is handled are market-dominated in an ad hoc, often sectoral, manner. In contrast, most European nations' governance is socially protective: individual privacy is an inalienable right, protected by comprehensive policies that are uniformly applied to all uses of personal information in the public and private sphere. Reidenberg's analysis presents nations' privacy policies as natural products of the governance paradigm. Moreover, as products of their respective governance paradigm, each policy reinforces the fundamentals of that paradigm. Viewing privacy through these governance "lenses" results in incompatible and conflicting interpretations of the FIPs.

Under the traditional Westphalian notion of sovereignty, terrestrially defined borders insulate these two governance paradigms from one another. As argued by Kobrin [71, 70] and others [77, 101], a ubiquitous cyberspace facilitates near frictionless international data flows that wear at terrestrial barriers. In turn, efforts at information policy convergence by powerful actors (both nations and multinational organizations) wear at traditional notions of sovereignty. Reidenberg argues that the EU Data Protection Directive and US Safe Harbor agreements bring these two governance paradigms in direct conflict. The effect is the imposition of socially protective privacy norms onto US institutions that have been inculcated in a market-driven, liberally normative regulatory environment. From the US perspective, this information is a commodity that has been freely exchanged for services; from the EU perspective, it is the lawfully protected property of it's citizens and comes with strict limitations on use.

2.1 Fair Information Practices

Despite differences in regulatory paradigms, the FIPs evolved in the emerging privacy climate of the 1960's and 1970's as a response to the increasing use of mainframe computers by governments [17, 130, 131]. The general notion of FIPs has been referred to as "principles for privacy safeguards," "principles for handling personal information," "principles of fair information practice," and "data protection principles" [17]. Bennett[17] and Reidenberg [103] summarize the FIPs; Bennett's six general principles are presented, noting finer-grained distinctions from Reidenberg where appropriate.

Principle of Openness The principle of openness requires the existence of a personal data repository be either publicly known or known to those about whom the data is stored. Stated as a negative rule, an organization should not hide the existence of a data repository. The openness principle ensures individuals are cognizant of the existence of repositories and can exercise the rights imbued by subsequent principles.

Principle of Individual Access and Correction Individuals should be able to access and correct personal files. This typically means that they have the ability to correct "any portion of the data that is not accurate, timely, or complete" [17]. This right is only available to the individual to whom the data corresponds.

Principle of Collection Limitation The collection of data is limited to "lawful and fair means" [92] and should be collected only with the knowledge and consent of the data subject. In Bennett's terms, this is intended to avoid "fishing expeditions." Data should only be collected for a specific function or purpose. Bennett notes that collection cannot be separated from use and disclosure.

Principle of Use Limitation Once collected, there are limits to the uses to which personal information may be put. Bennett highlights the notion of *relevance*: information may only be used for the purposes specified when it was originally collected. Restated as a negative rule, data collected for one purpose may not be used for another without the consent of the data subject. Implicit in Bennett's articulation, Reidenberg makes explicit that use limitation may also indicate that information may only be *retained* for as long as necessary.

Principle of Disclosure Limitation Also tied to the notion of relevance is disclosure limitation: data may not be shared with another organization without the permission of the data subject or without express legal authority. Reidenberg refers to this as the *finality principle.*

Security Principle Security is intended to introduce safeguards that reinforce earlier principles. The security principle requires safeguards against inadvertent or intentional destruction of data. The security principle also safeguards against unauthorized use, modification, or access of personal data. It is important to note that in this role security supports privacy principles; security and privacy are not interchangeable.

Influential sources of the FIPs were fair labor practices in the US [17], Britain's Younger Committee (early 1970's), Westin and Baker's recommendations to the National Academies in [132] (1972), and nascent articulations in the US 1970 Fair Credit Reporting Act [117] and the 1974 Privacy Act [118]. The two most influential international articulations of the early FIPs are the Council of Europe's (COE's) Convention for the Protection of Individuals with Regard to the Automatic Processing of Personal Data (COE Convention) in 1981 [36] and the Organization of Economic Co-operation and Development's (OECD's) Guidelines Governing the Protection of Privacy and Transborder Data Flows of Personal Data (OECD Guidelines) in [91][3]. In the late 1960's the COE began discussions to determine if the introduction of advanced technologies, in particular the computer, would pose a threat to the privacy protections in the European Human Rights Convention [37]. The COE Convention was the culmination of studies comprised of evaluating electronic data in the private sector (1971-1973) and the public sector (1973-1974). During that period it was also argued that a stronger normative instrument, based in international law, was necessary to protect and disseminate European notions of privacy as a fundamental human right. The full Convention was adopted by the COE Parliamentary Assembly in February

[3]These have been cited frequently in the literature ([77, 17]); the texts of these are drawn from [105]

of 1980 and by the Committee of Ministers in September of 1980. The convention ultimately received 18 nation's commitment to enact privacy laws

The OECD Guidelines were a response to an increasing volume of transnational data flows that were occurring both within and outside of the European reach of the COE Convention. The OECD comprises not only the EEC, but also Japan, Southeast Asia, Australia, and North American countries, giving it greater international reach than the COE Convention. The OECD Guidelines is a non-binding international agreement addressing threats to privacy and individual liberties independent of processing mechanism. Although the FIPs objectives are the same, this is different than the COE Convention's focus on automatic processing and binding mechanisms for cooperation amongst those that ratify the COE Convention [92]. The OECD Guidelines generally follow the FIPs above, recommending principles regarding collection limitation, data quality, purpose specification, use limitation, security safeguards, openness, individual participation, and accountability.

The evolution and further articulation of the FIPs by the COE and OECD are evidence of parallel efforts to harmonize privacy regulation across borders. This history of harmonization is further realized in the EU Data Protection Directive, where the FIPs are leveraged as the foundations for comprehensive, socially protective regulation that applies uniformly to public and private organizations. These general principles are widely accepted and are the nominal basis of both the EU Data Protection Directive and the US Safe Harbor Agreement. As broad statements, the DPD-FIPs and SH-FIPs are conceptually similar in substance and can be mapped back to Bennett's FIPs above. Political contention arises over interpretation of the FIPs and the scope of enforcement mechanisms applied by a particular regulatory paradigm.

2.2 Privacy in the EU

In October 1998, the EU Data Protection Directive came into effect. As a socially normative regulation, the objective is to establish privacy as a fundamental human right for European Union citizens. A second objective is to improve economic efficiency among member states by harmonizing data protection regulations. Both objectives come with enforcement mechanisms that require non-EU states meet a standard of adequacy, often requiring the development of a comprehensive privacy regime comparable to those being developed by EU member states. The US did not meet this adequacy requirement—the result was an externality that threatened to suspend EU-US data flows worth $120 billion [64].

The EU privacy regulatory environment is characteristically socially protective. The EU views privacy as an unalienable human right that requires government intervention to achieve efficacious protections. Although there are some distinctions amongst EU states, the EU Data Protection Directive imposes a set of comprehensive privacy regulations uniformly on both public and private data controllers. The socially protective privacy regime may be described as proactive, providing broad, blanket protections and relying on derogations of these regulations to accommodate pragmatic exceptions. The socially protective paradigm is embraced by EU states

and privacy advocates as a means to ensure efficacious enforcement of privacy, but opponents argue it places an "onerous" burden on businesses and can potentially stifle innovation.

2.2.1 Development of the EU Data Protection Directive

Both the framers of the EU Data Protection Directive and the US Safe Harbor Agreement avoided conflicts among substantially different interests by inviting only one set (category) of interests to the drafting table. In the case of the EU Data Protection Directive, the first (and critical) draft of the EU Data Protection Directive was written largely by the data protection authorities of member states. Heisenberg notes that "80% of a piece of EU legislation is contained in the Commission's first draft" [64][4]. Business interests did not become aware of the full extent of the EU Data Protection Directive until after they were given an opportunity to review the full draft. Despite arguments that obligations to notify the supervisory authorities of changes in data usage policies and issues of prior checking[5] were onerous and costly, very few business interests were reflected in subsequent changes. The majority of second draft changes were minor clarifications that did not change the essential substance of the EU Data Protection Directive.

At the time of drafting, the Chair of the Council of Europe's Data Protection Experts Committee and Chairman of the Commission's drafting group was Spiros Simitis, the German state of Hesse's data protection commissioner. Simitis is crediting with ensuring the focus of the document on privacy as a fundamental human right:

> Contrary to most other documents and nearly for the first time in the history of the Community, the Commission in its draft said that the need for the Directive is based on the need to protect human rights within the Community. This is why, when we speak of data protection within the European Union, we speak of the necessity to respect the fundamental rights of the citizens. Therefore, data protection may be a subject onwhich you can have different answers to the various problems, but it is not a subject you can bargain about. [27, p. 42]

Simits' statement highlights the socially protective privacy paradigm in Europe. The last point is especially salient to this thesis: the means of protecting privacy may vary across states and cultures, but privacy may not be bargained away completely[6]. Two objectives were achieved in preserving the status of data protection as a human

[4]Heisenberg attributes this to Hull [65].

[5]See Appendix A for a more detailed mapping of notification and prior checking. Regan also discusses this in terms of control of information and the granularity of opt-in and opt-out [101].

[6]The notion of bargaining about privacy is rooted in treating privacy as a good that can be exchanged. Law and economics works such as Posner [100] espouse this approach to privacy, among other legal issues. The notion of an exchange of privacy for services is a fundamental difference between the US and European perspective, in particular Europe's view that privacy is an inalienable, fundamental human right. The contrast will be revisited in the comparison of the EU Data Protection Directive and the US Safe Harbor Agreement in Section 2.4.

right: (1) privacy automatically trumps (no bargaining) other rights except similarly fundamental rights such as freedom of press and (2) it ensured the EU Data Protection Directive would raise the overall privacy protections across member states, not impose a lowest common denominator.

The result was a privacy regulatory structure that was ostensibly strong on enforcement, but was very unpopular among businesses and other information stewards [64]. In the EU, businesses and organizations are accustomed to more invasive, controlling EU regulations. Businesses in other regulatory environments (in particular the US) are not accustomed to substantive government intervention in business practices. An effect of this one-sided regulation building is that it missed the opportunity to leverage strong support of privacy as a fundamental right while also finding a compromise with business interests. The strong enforcement and auditing features of the EU Data Protection Directive that resulted from the influence of data authorities is lauded by privacy-proponents, but creates conflicts when proposed as a model for normatively liberal privacy regimes.

2.2.2 The EU Data Protection Directive

The EU Data Protection Directive intends to buttress privacy as a fundamental human right and further normalize the privacy regulations as a means to improve the efficiency of the EU Internal Market. Privacy as an inalienable human right is not new to the EU; it originally derives from the COE Convention for the Protection of Human Rights and Fundamental Freedoms [35] and is reflected in the Charter of the Fundamental Rights of the European Union [45]. Constructing privacy as a fundamental human right also suggests that privacy requires government intervention to be efficacious, hence a socially protective paradigm. Further, the EU population is, historically more trusting of socially protective government institutions [130]. As the guardian of individuals' privacy rights, EU regulations apply uniformly to the public *and* private sector institutions.

The EU Data Protection Directive is characteristic of general EU regulations. The EU Data Protection Directive is structured as a set of guidelines for designing new national laws or redesigning and/or refactoring member-states's existing laws. As guidelines from a supranational governance organization, the EU Data Protection Directive also intends to harmonize member states' legal structures as a means to lowering regulatory barriers and further integration of the EU Internal Market. The role of the Internal Market is to facilitate fluid, ongoing economic relations amongst the member states. The harmonization effort was intended to simplify data transfers amongst economically dependent member states by providing a common privacy platform.

The EU Data Protection Directive is characteristically proactive. As a fundamental right, individuals must have the ability to control how their personal information is used. Under this assumption, the rights of the individual are given primacy. The EU Data Protection Directive reifies this by requiring national laws that, in turn, require organizations respect individuals' privacy by giving open notice of data collection, use, disclosure, and by ensuring proper tools necessary for access and correction are

available. The enforcement effort introduces supervisory authorities that, on paper, have sufficient capabilities to monitor and enforce the EU Data Protection Directive's objectives. In effect, the EU Data Protection Directive may be viewed as high-level guidelines for implementing the FIPs described above. The following section describes the EU Data Protection Directive in terms of substantive guidelines for implementing Bennett's FIPs and as guidelines for reifying enforcement mechanisms that ensure proactive application.

2.2.3 Highlights of the EU Data Protection Directive

A generalized summary of the substantive privacy objectives of the EU Data Protection Directive are to ensure that information stewards:

- Collect only the information necessary for the stated purpose
- Only use information for the stated purpose
- Provide mechanisms for reviewing, updating, and removing personally identifying information from a system
- Entitle users to know how information was collected and from whom

Bennett [19] effectively maps the substantive Articles and subpoints describing the EU Data Protection Directive's privacy concepts to the FIPs. Appendix A provides a finer-grained mapping of the EU Data Protection Directive to the FIPs that is compatible with Bennett's mapping; a basic mapping, highlights, and deviations are presented here. In particular, the EU Data Protection Directive addresses *openness* (Articles 10, 11, 18, 19, and 21), *access and correction* (Article 12, 14, and implicit accuracy requirements in Article 6[d]), *collection limitation* (Article 6, 7, and 8; exceptions in Articles 8, 9, and 13), *use limitation* (Articles 6[b,c,e] and 7[b-f]; exceptions in Article 8, 9, and 13), *disclosure limitation* (Articles 6[b], 7[a], 10[c], 11, and 14[b]), and *security* (Articles 16 and 17). The *finality principle* is not applied absolutely. Exceptions are made for processing that is in the interest of the data subject, required by law (Article 7), or is part of a government function related to safety, security, national defense, or the prosecution of a crime (Article 13). Notions of accountability are loosely covered under Article 16, which addresses confidentiality of processes. This indicates that information may only be processed as per the instruction of the controller; this is interpreted as applying to both employees and third parties employed by the controller. The EU Data Protection Directive also addresses knowledge and consent (Article 7), but does not differentiate between opt-in and opt-out policy mechanisms [101].

2.2.3.1 Privacy Regulation and Enforcement

Ensuring a uniform enforcement of the espoused privacy principles is a key political element of the EU Data Protection Directive. The EU Data Protection Directive requires member states each have national laws compatible with and equivalent to

the data protection principles set out in the EU Data Protection Directive. The enforcement mechanism itself is described in Article 28; Articles 29 and 30 describe the role of the working group in maintaining the EU Data Protection Directive and coordinating with authorities

Under Article 28, each member state must have a supervisory authority responsible for monitoring data controllers (organizations acting as information stewards) and enforcing the EU Data Protection Directive. The supervisory authority is expected to be an independent policing organization capable of performing investigations, initiating interventions before processing starts, invoking legal proceedings, and hearing complaints against data controllers [19]. Bennett notes that this substantially extends the powers of pre-existing supervisory authorities. The Data Protection Directive's Working Group (Articles 29 and 30) advises these authorities, but does not have the power to change the EU Data Protection Directive. Rather, recommendations go to what Bennett refers to as a "mysterious body" referred to as "The Committee." The Committee comprises member state representatives. The Committee has the power to make decisions and regulations within the scope of the directive. The Committee also determines which countries meet the "adequacy of protection" requirements laid out in Article 25 and derogated in Article 26.

2.2.3.2 Transfer and Derogations

Articles 25 and 26 are largely responsible for the externality that the EU Data Protection Directive imposes on non-European Economic Area countries. Article 25 establish that private data may only be transferred to third countries deemed adequate by The Committee. Article 26 establishes derogations from the adequacy requirements intended to accommodate transfers acceptable to the data subject.

Article 25 requires that a third country ensure the adequacy of their privacy standards. This is distinct from assessing adequacy itself, which may be performed by an individual country, or as per Article 31, by The Committee. Bennett [19] makes an important note that "these rules and measures must be *complied with*, obviously meaning that symbolic value statements are insufficient." It may be that self-regulation (viz. the US) may read very similar to "symbolic value statements" under some interpretations.

Under Article 25, if a member state recognizes a third country as not having adequate privacy protections, it notifies other member states and the Commission. Further, member states are thus required to prevent data transfer to the inadequate country. Finally, Article 25[5] does indicate the Commission should, "at the appropriate time," negotiate some form of agreement that would remedy the adequacy problem and allow data transfers to resume. In this case, that negotiation produced the US Safe Harbor agreement.

Article 26 makes it possible for EU data subjects to waive the right to the EU Data Protection Directive protections and consent to data transfers to countries that have not met the EU Data Protection Directive's adequacy standards. Article 26 takes the form of a list of exceptions for which derogation is valid: among these are unambiguous consent, fulfilling a contract, legal obligation on the part of the data

controller, or acting in the vital interest of the data subject. While EU member states may make these arrangements of their own volition, they must do so on a case by case basis with data controllers and must report these authorizations to the Commission for review.

The European Commission (EC) originally found the sectoral, self-regulatory privacy paradigm of the US inadequate. One factor was that the fragmented sectoral privacy regulations did not meet the more comprehensive expectations of the EU adequacy standards. In terms of the enforcement objectives of the EU Data Protection Directive, a self-regulatory scheme is substantially different from the supervisory authority's powers of monitoring and enforcement. Article 25[5] and the desire to preserve essential trans-Atlantic data flows gave rise to the negotiations that ultimately resulted in the US Safe Harbor agreement.

2.3 Privacy in the US

Privacy is not a first-class notion in the US Constitution or the Bill of Rights; the word privacy does not appear in either document. In the US's normatively liberal regulatory environment, privacy is assumed to behave like an experience good. Privacy is expected to emerge as a natural product of extant norms expressed through self-regulated market activities. Under the self-regulatory approach, it is only when market forces fail and individuals experience substantive, identifiable harms that the government steps in with sector-specific regulations. In the case of privacy, HIPAA, the Bork Bill, and the Privacy Act of 1974 are examples of sectoral privacy regulation. In contrast to the EU's perspective on privacy as a fundamental human right, the US views privacy as a good that may be exchanged[7] to increase utility and that privacy "protections" are available to citizens through their choices regarding when to engage in these market exchanges.

The evolution of privacy thinking in the US is played out in the evolution of 4^{th} Amendment precedent regarding the bounds of privacy as it relates to police searches (Section 2.3.1). 4^{th} amendment precedent illustrates the fundamental mistrust of government institutions characteristic of the American model of democracy[8]. The following discussion is not to establish rigid boundaries regarding what is and is not private based on the scope of the cases, precedent, and regulations presented. Rather, this discussion highlights the dynamics of the regulatory environment in which privacy has evolved as a legal concept in US jurisprudence and in subsequent legislation. These dynamics give insights into US legal and political responses to new technologies in their legal and social context.

[7]Exchange is the less negatively-laden form of Simitis' "bargain[ing] away" of privacy rights discussed earlier.

[8]Westin provides an interesting distinction amongst the governance patterns of ostensibly similar western democracies in Chapter 2, *Privacy in the Modern Democratic State*, of [130]. Bennett contends that states tend to reuse regulatory instruments whose requirements, implications, and outcomes are familiar from repeated use [17]. It is arguable that the EU Data Protection Directive and US Safe Harbor are both instances of this tendency.

2.3.1 4^{th} Amendment Privacy Precedent

Privacy is considered a tacit component of the 1^{st}, 2^{nd}, 4^{th}, and 14^{th} Amendments; most of the seminal work on protecting privacy is found in 4^{th} Amendment adjudications and precedent. The basis of American privacy "theory"[9] is Warren and Brandeis notion of "the right to be left alone" [129]. Warren and Brandeis argue that as societies, cultures, knowledge, and values change, so must the common law in its "eternal youth."[10] Until 1967, police search was dictated by the trespass doctrine, which defined 4^{th} Amendment privacy violations in terms a physical search of an individual's residence or property without probable cause[11].

In *Olmstead v. United States*[12], the Supreme Court ruled that telephone conversations *were not* private and thus not subject to protection under the 4^{th} amendment. Although Brandeis articulated The Right to Privacy in 1890, in 1928 this notion remained the minority dissent in *Olmstead*. It is important to note the social context of the telephone in 1928: telephones were still fairly novel, they were not as integral a mediating technology of everyday activity as they were 30 years later (and even less so than now), and party-lines were still the norm in many areas. In this sense, society had not yet imbued telephone conversations with the privacy characteristics of an in-person conversation behind closed doors. It was not until *Katz v. United States*[13] in 1967 that *Olmstead* and the trespass doctrine were overturned.

Katz also established the Katz test for privacy relative to 4^{th} Amendment searches. The Katz test is satisfied if (1) an individual has an expectation of privacy in a given context and (2) society recognizes that expectation as reasonable. The Katz test serves as a legal instrument for identifying an articulable harm, identifying the context in which that harm occurred, and determining the reasonable expectations individuals have regarding privacy in that context. The Katz test, coupled with the concomitant ruling that telephone conversations are private, may be perceived as an exemplar of the ad hoc privacy regime in the US. Although the Katz test provides an established process, it is important to note that it is not applied on a continuous or proactive basis; rather, it is only applied reactively when substantial harm calls attention to a particular context and regulation (or lack thereof). The evolution of privacy precedent with regard to the telephone was repeated as e-mail became a

[9]Solove [110] argues that Warren and Brandeis [129] were not articulating a theory, but instead were identifying the problems characteristic of treating privacy within the American legal system and some nascent solutions. Their fundamental conceptualization, "the right to be left alone" was appropriated as a theoretical construct after the fact.

[10]This type of change can be considered a form of adaptation and self-correction. In *Olmstead* Brandeis explicitly notes the need to incorporate flexibility that can accommodate new technologies. A more general view of rules from legal theory implies flexibility, and adaptation can be identified in Hart's discussion of primary and secondary rules (Chapter 5 of [63]). Primary rules are those that enjoin behaviors; secondary rules confer powers to create or modify primary rules. In particular, an entire class of secondary rules, rules of change, specify how primary rules may be added, removed, or modified *at the pace of change of group behavior.*

[11]The common law roots of the trespass doctrine go all the way back to English Common Law and Semayne's Case in 1604. This is also considered the root of "A man's home is his castle."

[12]277 U.S. 438 (1928)

[13]389 U.S. 347 (1967)

common, well-understood means of communication. Recent debates focus on privacy in the cloud, behavioral advertising, and notions of what is and is not personally identifiable (or identifying) information [51].

In the spirit of Brandeis' youthful common law, the increasingly mutable character of Internet technologies creates almost limitless contexts in which "reasonable expectations" may be formed. Reasonable expectation is informed by the ubiquity of the technology's application, how well the implications of that application are understood by the general populace, and how these latter affect expectations of privacy. Linking this to the notion of privacy as an experience good, the expectation of privacy may be viewed as an outcome of sufficient exposure to a situation that leads individuals to assume or imbue a particular context with some degree of privacy. The evolution of the expectation of privacy for telephone conversations seems to have taken this right as the telephone became ubiquitous and took on the characteristics of person-to-person conversations rather than public (party-line) conversations.

A final case in 4^{th} Amendment privacy precedent, *Kyllo v. United States*[14], highlights the impact of new technologies. *Kyllo* reaffirmed that, as per Justice Scalia's opinion, "in the home, all details are intimate details"[15]. Justice Scalia's discussion of inference is especially compelling:

> We think that obtaining by sense-enhancing technology any information regarding the interior of the home that could not otherwise have been obtained without physical "intrusion into a constitutionally protected area,"[16] constitutes a search—at least *where (as here) the technology in question is not in general public use* [emphasis added].

The above establishes that inference is in fact a search when it figuratively (as opposed to physically) reaches into the home. The emphasis above reiterates the notion of reasonable expectations, especially the aspect of only addressing the privacy concerns related to a technology after it has proven to be a common, well-understood tool. In this case, a sense-enhancing technology was used to infer information about the activities within the home, a violation of reasonable expectations. The types of inferences used by online service providers to customize their products and service offerings are common amongst professionals in the area, but may not be well-understood by the general populace in terms of what they are capable of or how extensively they deployed. It is on the basis of common understanding, or what is typically referred to as "informed consent," that later arguments regarding the collection of non-PII and its implications will be built[17].

[14]533 U.S. 27 (2001)

[15]Id. 14.

[16]Silverman, 365 U.S., at 512

[17]It may be argued that automated data collection mechanisms employed by online service providers are a form of sense-enhancing technology that is not well-understood by the general populace. This also draws on the notion that privacy is not an experience good. Absent this understanding of the mechanisms, the implications of allowing automatic data collection and processing is not clearly articulated (Chapter 4). Taking these arguments as given, it is difficult to argue the average user can give "*informed* consent" regarding the automatic collection of non-PII and its implications for his/her privacy.

2.3.2 Highlights of the US Safe Harbor

After substantial negotiation, the US Safe Harbor agreement emerged as a US-specific derogation of the EU Data Protection Directive and its adequacy standards. The US Safe Harbor agreement is literally an agreement between the US and the EU; it is neither a treaty nor a modification of the EU Data Protection Directive. Rather, the US Safe Harbor agreement is the culmination of negotiations between the EU Commission and the US Department of Commerce to ensure the continued flow of data from Europe to the US. The nature of the agreement establishes adequacy standards that Safe Harbor participants are expected to implement. The prescriptive articulation of the US Safe Harbor is laid out as a set of basic privacy principles and FAQs that further explain implementation details and requirements.

2.3.2.1 Development of the US Safe Harbor

Shortly after the EU Data Protection Directive came into effect, the EU Commission requested the US create a regulatory body to monitory privacy violations and respond to consumer complaints [64]. The Clinton administration rejected this suggestion. Instead, the Clinton administration began its own investigations into Internet regulation, relying largely on industry experts to contribute to policies regarding privacy and the information economy in general. The Clinton administration's priorities focused on economic development. In particular, it focused on the role of the Internet in e-commerce, how regulation may stifle innovation, and, subsequently, how these may affect the growth of the then nascent information economy.

An early articulation of the Clinton administration's position was presented by Ira Magaziner in [114], *A Framework for Global Electronic Commerce*[18]. Magaziner (on behalf of the White House) presented five principles critical to electronic commerce. The leading principle is that "the private sector should lead," followed by supporting principles indicating "governments should avoid undue restrictions on electronic commerce" and that "governments should recognize the unique qualities of the Internet" [114]. As a reaction to the EU Data Protection Directive, the notion that the private sector should lead is almost the opposite of the regulatory function of the EU Data Protection Directive's notice to supervisory authorities and requirements for prior checking[19]. The reasoning behind these arguments is that the key to the online environment is unfettered innovation; in the normatively liberal US, unfettered translates to self-regulation with the government stepping in only in cases of gross violation of individual rights.

During the two years following the EU Data Protection Directive coming into effect (one year before member states were to begin acting on adequacy requirements), there was little formal progress toward resolving the conflict between the US and Europe. The slow adoption by EU member states was one contributor, interpreted as the possibility that enforcing the adequacy requirements may not be as strict

[18]Magaziner elaborates his position and the influences on this report in an interview in [79].

[19]See the discussion in Section 2.2.3 and the details of the EU Data Protection Directive in Appendix A.

as implied by European advocates and the black letter of the EU Data Protection Directive itself. US opponents of the EU Data Protection Directive believed the threat to trans-Atlantic trade would also soften the EU position. Although this was not sufficient leverage to completely ignore the EU Data Protection Directive, the trade volume coupled with arguments for self-regulation was sufficient to craft a US-solution rather than capitulating to the EU. The architect of the hybrid US Safe Harbor Agreement was Ambassador David Aaron, appointed Undersecretary of Commerce for International Trade in November of 1997.

Until Aaron's arrival, the US position was that the information economy required the flexibility of self-regulatory privacy. The difficulty with this argument was that the US did not have a self-regulatory solution. Magaziner articulated the position in a statement before the Subcommittee on Telecommunications:

> ...the self regulatory solutions that we are looking towards, we think, ought to be acceptable to [the EU]. And that we will not permit a breakdown of dataflows. We think that we can prevail in that discussion. But only if we have the self regulatory regimes up and going. Because the latest comments they have been making to us is [sic], "Magaziner, you've been talking about self regulation for a year. Fine. Where is it?" And at some point you can't fight something with nothing. [80]

Aaron's background in tax law and the notion of "safe harbors" was the inspiration for the US Safe Harbor Agreement, which for the time being has resolved the conflict and provided the an early test of hybrid self-regulatory mechanisms [48, 47].

The first formal meetings were between Ambassador Aaron and European Commission negotiators John Mogg (DG of the Internal Market and Financial Services Directorate General) and Susan Binns (Director in charge of data privacy). Heisenberg indicates that, unlike previous junior negotiators, Mogg and Binns did not apply the black letter, but took a more interpretive approach and wanted a pragmatic solution [64]. Aaron recollected the first articulation of the idea behind the Safe Harbor during the negotiations:

> Nobody knew how we were going to do this, and we were just sitting in John Mogg's office one day, and I had always been struck by the idea of "Safe Harbor." ... [A]s we were discussing this issue, I thought ... well if we couldn't get the country to be considered "adequate," maybe what we could get considered adequate are the companies. And that if we could set up some kind of a regime that could have adequacy finding for a system, not for a whole country's law and regimes, and so the word just popped into my head, as describing Safe Harbor. [48]

Farrell reports that this piqued the interest of the European Commission negotiators and that the primary contribution of the Safe Harbor was that it opened the door to substantive negotiations that moved beyond deadlock over normative differences [48]. Coupled with a visit by member state representatives to hear from representatives of self regulation organizations and US enforcement officials, the idea of self regulatory, hybrid system gained support from the remaining skeptics.

Once past the negotiations with the EU, drafts of the US Safe Harbor Agreement were distributed to industry groups. Comments from privacy advocates responded that the solicitation of comments was one-sided, most clearly indicated by the salutation "Dear Industry Representative." Heisenberg notes that the comments of privacy consumer groups were near identical to those of European businesses in response to being left out of EU Data Protection Directive drafting process. Although not quite regulatory capture, industry clearly had a substantive influence on the direction of negotiations and the drafts of the US Safe Harbor Agreement.

Upon reviewing the draft of the US Safe Harbor Agreement, the Article 29 Working Party had reservations [15]. The European Parliament rejected the US Safe Harbor Agreement on grounds that the European Commission had not taken the recommendations of the Article 29 Working Party, as the technical experts, into account and that the Commission may have overstepped its bounds in the negotiation of the agreement. Modulo the accusation of overstepping legal bounds in negotiation, the European Parliament does not have the power to determine adequacy[20]. Despite the admonitions of both the Article 29 Working Party and the European Parliament, the Commission announced the Safe Harbor met the EU's adequacy requirement and that it would go into effect on November 1, 2000.

2.3.2.2 Safe Harbor Principles

The key elements of the US Safe Harbor agreement are the Safe Harbor privacy principles and the means of enforcement. The term enforcement is overloaded. Enforcement as a US Safe Harbor privacy principle will be referred to as an enforcement principle. The means by which the seven US Safe Harbor principles will be enforced will be referred to as regulatory enforcement.

The Safe Harbor agreement, as described by the Department of Commerce, is based on seven principles [125]. These principles reflect many of the EU Data Protection Directive principles and their FIPs counterparts. The first Safe Harbor principle is *notice*; organizations are required to give notice to users with regard to what information they collect, the purposes of that data collection, contacts for complaints, and information regarding how information is shared with third parties. The second principle is *choice*; organizations must (1) give individuals the opportunity to opt out of disclosure of their information to third parties, (2) obtain consent for secondary information uses outside of the original purpose, and (3) obtain consent (opt in) for the use of sensitive information along the lines of criteria (1) and (2). The third principle is *onward transfer*; organizations may only share information with other Safe Harbor participants or those with whom they have a written agreement that the third party will provide an equivalent level of protection. The fourth principle is *access*; this is akin to the access and correction principles in the EU Data Protection Directive and in the FIPs. Access ensures that users can correct, update, or delete inaccurate information. The exception is when the effort necessary to provide access is disproportionate to the supposed risk to the data subject's privacy. The

[20]Recall modus vivendi in Section A.2 and Footnote 2.

fifth principle is *security*; this requires an organization take reasonable information security precautions to avoid unauthorized access and to ensure the integrity of the data (in terms of content). The sixth principle is *data integrity*; this ensures the data subject's information is relevant and reliable for the intended use. The seventh principle is the *enforcement* principle. The enforcement principle has three components. First, the organization must provide mechanisms for recourse by which disputes can be reconciled and applicable damages awarded. Second, there must be mechanisms to ensure companies are implementing the espoused privacy policies. Third, there must be sufficient, proportionate sanctions that can be applied if organizations do not comply with the Safe Harbor Principles.

The third part of the enforcement principle alludes to overall regulatory enforcement. In the US, there is no equivalent to an EU supervisory authority that monitors and enforces privacy policy. Rather, sector-specific privacy functions are distributed amongst the bodies that regulate those sectors. The Safe Harbor is self-regulatory. To join, an organization must either develop its own self-regulatory privacy policy that adheres to the Safe Harbor principles or join a self-regulatory program such as TRUSTe [115].

2.4 Comparison and Contrast

These two implementations are similar in that they provide an articulation of the FIPs, but differ substantially in how they propose to impose the FIPs. The following general critiques summarize the problems with the EU Data Protection Directive and US Safe Harbor in their role of providing guidance to the development of online privacy policy standards.

The difficulty with the overly broad scope of rules is largely a problem of enforcement. Kuner uses the instance of 'equipment' used by data controllers that are located outside the EU. Another instance is the broadly scoped definition of personal information. The problem is that there is a dearth of guidance regarding how to create standards defining what constitutes personal information. The problem is further confounded when abstract definitions are placed in context. Development of standards is a solution, but leads to political and economic questions regarding who should create these standards and, in turn, how these standards should be enforced.

Another criticism of the EU Data Protection Directive is that the requirements are burdensome or "onerous." Criticisms of the notification mechanisms are one instance of burdensome requirement. The immediate implication is the additional effort to give notice of each operation an organization is using or intends to use. The notion of burdensome requirements is also invoked in arguments regarding innovation. Businesses [25, 58] and the US government (the FTC under the Clinton administration) argued that comprehensive policies create "onerous" burdens that threatened to stifle innovation [23]. For example, combining notice (Articles 18 and 19) with prior checking (Article 20) before an organization uses a particular operation, a data controller may be required to wait for government approval to deploy a new service.

The failure to keep pace with technology is not unique to privacy. Historically,

regulations have needed revision as either the technology changes or novel uses of the technology emerge. The rate of development of Internet technologies in particular have put the pace of change on "fast forward." A contributing factor is that Internet technologies are generative [138, 140]—these technologies act as a platform that gives rise to emergent, previously unforeseen uses, is designed for novel (re)combination, and supports multiple, potentially competing evolutionary paths to an objective. For instance, behavioral advertising is a recent instance of innovation that has defied conventional categorization under the current conception of personally identifiable information. Kuner indicates that the EU Data Protection Directive was designed in the era of mainframes; this environment is substantially different from today's generative Internet.

Regarding keeping pace with technology, industry has a legitimate claim to expertise. Industry leverages claims of expertise and the pace of technological development to argue for more dynamic regulatory regimes, such as those based on self-regulation. The general claim is that such regimes are necessary to keep pace with a dynamic Internet landscape; a particular instance is the ongoing development of behavioral advertising. As an explicit instance of an online service provider's articulated support for self-regulation, Google has indicated that legislative processes are too slow and that the outcomes are often outdated when they do arise [38].

The result of the US Safe Harbor is a policy instrument that has been characterized as a hybrid privacy arrangement [47, 48, 49] that synthesizes comprehensive character of the FIPs with a self-regulatory enforcement mechanism. The expectation, as articulated in Section 2.3.2 and Appendices B and C, is that private actors can self-regulate, backed by government sanctions when privacy violations are demonstrated (reactively). As evidenced by the policy sample and discussion above, the user has access to conventional PII, but there is little guidance regarding how to *emergent* privacy violations should be handled. The interest-based construction of the US Safe Harbor has placed a great deal of discretion in the hands of private actors regarding emerging privacy threats that lie outside the conventional bounds of PII-based protections. In particular, the US Safe Harbor does not provide guidance for how to ensure users have sufficient (meaningful) information to evaluate novel applications of previously "innocuous" information, such as the construction of an aggregate image.

The argument here is not that the architects of governance instruments should anticipate all possible contingencies. The inability to foresee all possible scenarios is a well know difficulty in the construction of governance systems [63]. The failing is that the US Safe Harbor captures the privacy problems associated with PII-based privacy violations, but relegates unforeseen deficiencies such as the aggregate image and mixed context problems to self-regulatory mechanisms. As discussed here, these are not experience goods, with clear implications to the lay user, thus self-regulatory enforcement mechanisms, based in bringing evidence of these harms to the FTC, does not bind. The privacy policies provide sufficient evidence of the aggregate image in their appeal to customization under the "service-and-improvement" framing. Absent more precise guidelines for data categories and data purposes (again, this is relegated to self-regulation), it is not possible to determine precisely what is in an aggregate image is in order to argue whether it is harmful as a means of engaging

in the self-help processes assumed by the self-regulatory mechanism. Thus, the institutional deficiency is the unmodified coupling of self-regulatory mechanisms with comprehensive policies in a technological context that, absent well-framed metaphors, is not necessarily well-understood by the lay user. The result is that the perception of comprehensive policy and the difficult to demonstrate privacy harms created by the SH-FIPs contribute to obscuring the privacy deficiencies of FIPs "compliant" online service provider privacy policies.

Chapter 3

Privacy Theory

When a conception of privacy is too narrow, it ignores important privacy problems, often resulting in the law's failure to address them. On the other hand, privacy conceptions that are too broad fail to provide much guidance; they are often empty of meaning and have little to contribute to the resolution of concrete problems.

Daniel Solove, Understanding Privacy [110, p. 39]

Despite substantial efforts across social science, law, and computer science, a single theory of privacy has not obtained. Solove indicates that "[t]he most prevalent problem with the conceptions [of privacy] is that they are either too narrow because they fail to include the aspects of life we typically view as private or too broad because they fail to exclude matters that we do not deem private" [110, p. 13]. Abstract definitions may be covering, but are not specific enough to give insights into concrete applications. To apply abstract definitions, substantive understanding of the context and subsequent efforts in framing are necessary for efficacious application. Moving top-down, a common problem with privacy theory is the tractability of operationalization. On the other end of the spectrum, bottom-up privacy conceptions are narrowly tailored to the issue (or narrow operationalizations of) and give little insight into conceptually adjacent privacy problems.

Privacy rules, legislation, and policy are similarly susceptible to the problems of overly broad or narrow prescriptions. Bennet's articulation of the FIPs, the DPD-FIPs, and the SH-FIPs are broad, normative statements that leave a great deal of interpretation to those implementing the FIPs. In both the US and Europe, despite each having a common set of FIPs, implementation and enforcement differs within each, leading to differing mechanisms for developing standards, guarantees, and protections. To provide insights into balancing these problems and to highlight the technical and policy strategies that may provide robust, sustainable solutions, this work will draw from two relatively new frameworks for reasoning about privacy in context: Nissenbaum's notion of contextual integrity [89] and Solove's family of privacy concepts [110]. In contrast to a grand theory that attempts to capture as many divergent instances of privacy as possible, these frameworks provide a conceptual vo-

cabulary for characterizing the dimensions of problem frames that can be applied to understand normatively divergent instances of privacy. In other words, these theories provide a robust, pragmatically rooted conceptual framework for articulating the bounds of a private context, what contributes to a privacy violation, what constitutes the violation itself, and the types of outcomes. This vocabulary reinforces the distinction between the knowledge barriers (information asymmetries) that contribute to privacy deficiencies.

To reinforce the contrast between conventional theories, Westin's "modern" definition of privacy is briefly presented. Nissenbaum's contextual integrity is presented as a framework for bounding privacy contexts. Solove's typology[1] is then presented as a family of concepts for reasoning about context specific privacy. Together these provide a distinct terminology for discussing the context specific roots of knowledge barriers, privacy objectives, privacy deficiencies, and implications. Following these theoretical frameworks, Schön's notion of generative metaphor is presented as a tool for reasoning about first whether a policy provides meaningful conceptions in its role as a decision tool and second as a means to better surface (or elicit) the kinds of requirements necessary to develop context-specific privacy tools.

3.1 Westin

Westin provides the modern articulation of privacy:

> Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others [130].

Westin's definition provides a very broad categorization of privacy that covers individuals, groups, and institutions. Of particular interest is that Westin identifies group and institutional actors as having a claim to privacy—in the context of online service providers, an organization's privacy translates to trade secrets that protect the details of innovative secondary uses of private information and competitive advantage.

In [130] Westin addresses both individual aspects and categorizations of privacy and the implications of privacy for a society[2]. Of particular interest to this discussion is the idea that groups and institutions have a right to privacy. Privacy on the part of an institution facilitates efficient discussion of ideas and concepts without the fear of outside interference in these decision processes. Here, the objective is to allow a group to explore a range of popular and unpopular ideas that may ultimately lead to an efficient, efficacious, and socially acceptable outcome.

[1] Solove's characterizes what is referred to here as a typology as a taxonomy. The family of concepts presented by Solove is more appropriately a typology because the categories are not discretely delineated as in a true taxonomy.

[2] Although not directly applicable to this thesis, Westin identifies privacy as a distinct element of democracy [130]. Privacy facilitates the free discussion of topics that may not be popular amongst the ruling party. It is this freedom to form private groups that is a key distinction between democratic governments and totalitarian regimes that seek to control the ideas and information shared amongst individuals.

The key distinction is that the decision process, intermediary steps, alliances, and good-faith compromises amongst nominal adversaries are allowed to evolve unhindered. In this scenario, the determination of "when, how, and to what extent" described above protects negotiations and ensures that strategic exchanges, decisions, and compromises remain in their proper institutional context. This is different from the traditional application of the sunlight principle[3], which, in its strongest form, expects complete transparency in all decision processes. Rather, in this work, the sunlight principle is to be applied to the disclosure of *outcomes and implications*, not to the technical development processes or as a precautionary principle. Framed in this way, the trade-offs between the economic efficiency of online service providers and individual privacy become a trade-off between the (private) processes that allows organizations to develop competitive services and individual privacy. Stated this way, the process highlights the quantity and kind of information shared amongst privacy-related actors, not the typical dichotomy of whether or not information should be shared or not.

3.2 Contextual Integrity

Nissenbaum claims that the conventional framework for reasoning about privacy "fail[s] to clarify the sources of their controversial nature" [89, p. 119]. Rather, contextual integrity is presented as a justificatory framework for privacy policy and law that draws on moral, political, and social values to highlight the root of the problem. Nissenbaum's criticisms of conventional privacy are based in failures of court cases to set applicable precedent[4] for future cases and that many disputes are characterized more by adversarial encounters between specific interests than by addressing the root of the problem or the social value of privacy[5]. Contextual integrity is not presented as a privacy theory of the same breadth as Westin's, but rather as a framework for pragmatically reasoning about the root of privacy problems in a way that establishes applicable precedent. Nissenbaum also criticizes the typical dichotomies used to describe privacy policy issues: sensitive and non-sensitive, private and public, government and private. Following the focus on identifying the root of privacy conflicts, Nissenbaum's discussion focuses on a more textured depiction of individuals' activities as they move "into, and out of, a plurality of distinct realms" [89, p. 137].

Context has thus far been used rather loosely and often synonymously with the

[3]'Sunlight is the best disinfectant; electric light the best policeman,' U.S. Supreme Court Justice Louis Brandeis [24].

[4]This is also a problem in the larger class of Internet Jurisdiction cases. For instance, the case of Yahoo! and France in 1998 was a conflict over the content of online auctions [73, 139]. Recently, similar conflicts over what is an is not to be considered a counterfeit luxury good have created disparate rulings on each side of the Atlantic [26, 28, 55, 61, 81, 82, 83, 95, 96, 97, 111, 113].

[5]Both Nissenbaum and this thesis draw on Regan's analysis of interest-driven legislation of privacy [102]. Nissenbaum cites Regan regarding the devolution of debates over privacy into adversarial confrontations between interest groups intent on promoting their interests over their opponents' [89, p. 122].

notion of an environment[6]. Privacy norms are socially constructed, and, as such, are embedded in socially constructed contexts that may persist across multiple built environments. Privacy norms are rooted in the details of these contexts (or situations). Contexts may thus be defined in terms of distinct sets of norms that comprise notions of "roles, expectations, actions, and practices" [89, p. 137]. The objective of contextual integrity is to provide a cohesive perspective on how these contribute to meeting individuals' privacy expectations within a given social context.

The notion of contextual integrity gives insight into how to accurately and consistently surface individual privacy preferences. In one sense, this is the dual of behavioral profiling leveraged for targeted advertising. In both cases, preferences are inferred based on contextual indicators. It will be later argued that one approach to improving context-based privacy will be the applications of technologies and methods used to infer interests in advertisements to provide feedback tools for eliciting actual (as opposed to imposed or espoused) privacy preferences. Contextual integrity "ties adequate protection for privacy to norms of specific contexts, demanding that information gathering and dissemination be appropriate to that context and obey the governing norms of distribution within it" [89, p. 119]. This definition builds on two type of norms: norms of appropriateness and norms of flow or distribution.

Norms of appropriateness dictate what information is fitting to reveal in a particular environment. Contexts differ along the dimensions of explicitness and completeness. In terms of explicitness, a context may have very low barriers, finding individuals sharing detailed, intimate information. An example of one-way sharing of such information is the patient-psychiatrist relationship. An example of a two-way sharing is between extremely close friends or between long-time romantic partners.

Norms of distribution (or flow) dictate what information can be transferred to others while respecting the contextual norms under which it was shared. Following the earlier examples, either a spouse or a psychiatrist sharing intimate details with others outside the original context would violate contextual integrity and may be perceived as a breach of privacy. In the case of the spouse, the binding norm violated would be that of personal trust. In the case of the psychiatrist, the binding norm violated would be that of a professional patient-physician relationship, a formal proxy for trust.

Although this is an elegant and insightful way of describing privacy, Nissenbaum admits it is very difficult to operationalize. The difficulty lies in the conceptual and empirical research necessary to understand a context sufficiently well to concretely define norms of appropriateness and norms of distribution. Such an effort would, ideally, yield well-formed rules describing appropriateness and distribution. Although this effort is difficult using conventional data collection methods, the (built) online environments that create a particular online context may be instrumented to help collect this information in real time, as it occurs.

[6]In this discussion, the term environment is technically laden with the connotation of "built environment" or an "engineered environment" in contrast to the more purely socially constructed environments implied by Nissenbaum's context.

3.3 Solove's Privacy Framework

In [110], Solove critiques the existing approaches to privacy theory and their application in law and policy. As per this chapter's epigraph, too narrow a theory does not capture all the issues that fall under an intuitive domain of privacy and too broad a theory (or principle in the case of the FIPs) provides too little guidance, often giving insights into the abstract but little substantive direction for concrete policy making. Solove draws on Wittgenstein's notion of family resemblances[110, pp. 42–46][7] and pragmatist philosophy[110, pp. 46–51][8] to approach a typology of privacy concepts from the bottom-up. The argument is that this approach to theory building can help identify the middle ground between overly-narrow, under-inclusive privacy theory and the overly-broad theories that are empty of substantive meaning.

Solove's typology provides a terminology for describing the implications of the privacy policies described in Chapter 4. As an empirical typology of privacy concepts built around concrete situations, Solove has identified characteristic activities that contribute to defining instances of privacy contexts as a configuration of concepts from the consistency of this typology. The model identifies the lifecycle of a data subject's information, identifying four categories of potentially harmful activities: information collection, information processing, information dissemination, and invasion. The four categories are further decomposed into subcategories that refine these activities and identify implications of these activities. Details of this typology can be found in [110, pp. 101–170] with descriptions of the concrete situations that exemplify and contribute to this typology. The following summarizes and defines the terminology used in the remainder of this thesis.

3.3.1 Information Collection

Solove decomposes information collection into interrogation and surveillance. The categories are similar to collection limitation principles in the FIPs. Applied as guides to contextually informed policies, they contribute to understanding precisely what information is collected in a given situation or scenario (context). *Surveillance* is defined as "the watching, listening to, or recording of individual's activities" [110, p. 104]. *Interrogation* "consists of various forms of questioning or probing for information" [110, p. 104].

3.3.2 Information Processing

The subcategories of information processing describe the activities and outcomes of how information is manipulated and used by a data controller. The first subcategory is *aggregation*, which deals with how different pieces of information about a person are combined, potentially resulting in additional information. This is different from the statistical aggregation commonly described in online service provider privacy policies.

[7] Solove directly references Wittgenstein's *Philosophical Investigations* [134], but also refers the reader to [60, 98] for a better understanding Wittgenstein's family resemblances.

[8] Solove draws primarily from the work of John Dewey and William James.

Based on the sample, aggregation denotes a statistic describing a particular category of information, such as how many (or what proportion of) individuals within a certain age range viewed a given category of products. Solove's aggregation describes the combination of heterogeneous data that may provide more complete information, and thus a more complete picture of an individual. This has implications for other categories of processing, such as identification as well as distortion (under information dissemination) and invasions of privacy[9].

Identification is broadly defined as linking information to particular individuals. The definition of what is personally identifying is another critical element of the privacy debate. The notion of what is personal identifying information (PII) and what is non-PII has been contended from a number of perspectives. An individual's identity may be derived from their community (online and/or terrestrial), ethnicity, religion, political beliefs, nationality, physical attributes, interests, or some combination of these and/or other categories of information. Operationalizations, or more accurately the lack of precise operationalizations, of identity are a critical component of characterizing information as PII and the subsequent restrictions on how information is used and how it must be protected.

One problem with information processing and identity is the threshold at which individually innocuous categories of data, in aggregate, identify an individual. Name, social security number, telephone, etc. are individual categories that can be directly linked to individuals and are considered personally identifying. The collection of categorical information such as age range, neighborhood, and specific interests, do not, individually, identify a person but collectively may describe an individual quite accurately. Whether the processing that aggregates this data creates an "identifying" image of the individual and whether the further processing to link this image to a unique identifier constitutes the collection and maintenance of PII is currently being debated [51].

Insecurity addresses how well information is protected from illegitimate access through intentional or unintentional means. Information leaks that occur when information is poorly protected often lead to other violations related to identity (theft), aggregation, and distortion. (In)security is often considered a gateway to many categories of privacy violation and it has received substantial attention in the form of security principles in the FIPs. Security violations are the focus of many of the FTC court actions related to privacy; almost all of these are actions against organizations for either misrepresenting their information security practices or neglecting to correct well-known vulnerabilities[10]. The focus on the security of information also

[9]It is this kind of overlap that makes this classification scheme a typology rather than a taxonomy.

[10]The majority of FTC cases cited by [51] deal with the misrepresentation of technical security protections and negligence regarding well-known vulnerabilities. The only FTC case published as of this thesis dealing with tacit data collection is [52]. This case deals with Sears deploying a stand alone application that runs on a users' computer and that collects information about users online behavior. It is arguable that this was provable only because the complete client executable was available and analyzable, allowing regulators to determine precisely how and what information was collected. Moreover, because it resided on the user's computer, it is more easily argued as an invasion.

contributes to the conflation of security and privacy, leading to the frequent misconception that strong security implies strong privacy protections. This conflation mistakenly portrays technical security (via constructions of privacy that focus on the control of information and technical security as the guardian of information control mechanisms) as a necessary and sufficient condition for privacy. Rather, focusing on security ignores other mechanisms by which organizations may legitimately share information (or engage in information dissemination).

Secondary use addresses the situation when information is collected for one purpose and subsequently used for other purposes without the consent of the data subject. Secondary use is evident in the FIPs under use limitations and was also stated in the HEW Report, which indicates that "[t]here must be a way for an individual to prevent information about him obtained for one purpose from being used or made available for other purposes without his consent."[11] Secondary use is not necessarily a violation of privacy, may have beneficial outcomes, and may require balancing societal outcomes against individual rights. For instance, secondary uses that enhance the service provided to an individual may in fact provide greater (albeit perhaps unexpected) utility to the user. The dual is that secondary use may be undesirable. As such, secondary use is often confounded with opt-in and opt-out issues regarding which is appropriate to a given situation. This coupling of secondary uses with the spectrum of opt-in and opt-out highlights the source of business claims that strong opt-in policies are onerous and stifle innovative secondary uses.

Another factor regarding secondary uses is that many individuals simply do not know the range of *potential* secondary uses to which seemingly innocuous data may be put. Schwartz highlights the possible information asymmetries that result from users' lack of understanding of potential secondary uses:

> [I]ndividuals are likely to know little or nothing about the circumstances under which their personal data are captured, sold, or processed. This widespread individual ignorance hinders development through the privacy marketplace of appropriate norms about personal data use. The result of this asymmetrical knowledge will be one-sided bargains that benefit data processors. [108]

Although the privacy landscape has changed substantially since Schwartz made this observation in 1999, the outcome remains the same even though both the EU Data Protection Directive and the US Safe Harbor agreement both make nominal claims requiring disclosure of data purposes via collection and use limitations. This is also a form of exclusion, described below.

Exclusion is relates to the interplay between openness, access, and control principles found in various articulations of the FIPs. Exclusion occurs when individuals are denied information about how their information is being used. As such, exclusion is not often considered a harm unto itself. Exclusion does facilitate other harms when the user does not know they need additional information to make informed decisions. The distinction between exclusion as a contributor to privacy violations,

[11] Referenced in [110, p. 130], originally in [126].

but not a violation itself, is conceptually homomorphic with the distinction between knowledge barriers (information asymmetries) and privacy deficiencies discussed in the Chapter 1.

Exclusion is also a function of notice. When an individual is not given notice regarding how her information is used, they are effectively excluded from decisions about information processing. Subsequently, they have less influence over the kinds of secondary uses that may occur.

Exclusion also has a fiduciary element related to the understanding of secondary uses described above. Fiduciary elements intend to introduce accountability, especially when one party to a relationship has a particular power over another party (or parties). For instance, in Nissenbaum's doctor-patient relationship, because of the implications of information uses and the decisions made based on both the information and secondary uses, physicians are required to acquire informed consent. In the strong form of informed consent (opt-in), this may require disclosing future or potential uses of the data. The notion of opt-out, which assumes consent unless an individual explicitly expresses otherwise, does not support a fiduciary relationship. Moreover, as will be seen, because opt-out is constructed by the online service provider, it is often constructed in an all-or-nothing fashion that discourages opting out because one loses substantial utility from the beneficial (innocent) secondary uses.

3.3.3 Information Dissemination

Information dissemination is concerned with the legitimate and illegitimate means of spreading or revealing information about an individual. Solove identifies a number of outcomes of information dissemination that recur in various bodies of privacy law. Each is outlined below.

The distinction between breach of confidentiality and disclosure is subtle. In both cases, information that is harmful is disclosed outside of the intended context. In the former, a *breach of confidentiality* is a form of betrayal by a trusted actor, regardless of the information. In the latter, *disclosure* of information is the dissemination of information relating to a private matter that might be "highly offensive to a reasonable person" [110, p. 137]. In the case of disclosure, the focus is less on the perpetrator of the illegitimate dissemination, but rather the nature of the information.

Disclosure relates to the dissemination of true information that, if exposed to the public, may harm the data subject. Some may argue that this is contrary to free speech and thus two fundamental human rights conflict. An efficiency argument is that limiting disclosure inhibits the ability to determine an individual's credibility and thus the ability to imbue the individual with trust. Richard Posner argues that disclosure limitations protect the "power to conceal information about themselves that others may use to their disadvantage."[12] This articulation of disclosure limitations illustrates the overlap with the related notion of exposure.

The notion of *exposure* captures the act of revealing physical or emotional attributes of an individual that may be embarrassing or humiliating. Although related

[12]Referenced by Solove [110, p. 142]; originally in [100].

to disclosure in that it involves the dissemination of true information, the information involved in exposure is a subset of that covered by disclosure and related directly to an individual's body, health, or emotions. Issues relating to aggregation and exposure may also be a result of distortion, where the implications of true facts result in inaccurate or imprecise information linked to an individual that may elicit the same feelings related to exposure, even though the data is not the "true information" associated with breach of confidentiality, disclosure, or exposure. One contributor to aggregation and potential distortion is increased accessibility.

The trite response to issues related to increased accessibility is that information is already publicly available. In other words, the nominal accessibility of the information has not changed. The issue of *increased accessibility* speaks to the implications of simple automatic access to a wide variety of information sources that may facilitate aggregation. The issue of increased accessibility is also related to use limitations. Information may be public and was collected for a specific purpose—secondary uses emerge when multiple sources are combined to reveal additional information via aggregation. As such, the issue of increased accessibility may be considered along with security as a factor in privacy violations, but whose presence or absence does not constitute a privacy violation in itself[13].

Appropriation was the first privacy tort recognized after Warren and Brandeis' seminal paper. The tort of appropriation occurs when "the defendant must have appropriated to his own use or benefit the reputation, prestige, social or commercial standing, public interest, or other values of the plaintiff's name or likeness."[14] A clear instance of appropriation is identity theft, where the appropriation of another's identity facilitates theft and may create subsequent harms through damage to the victim's credit. Less obvious are Solove's implications regarding the role of appropriation in safeguarding what society believes to be appropriate when constructing an individual's identity. In the case of online service providers, as noted earlier in the discussion of identity, an online service provider's built environment dictates how the user may access or change the information about herself held by the online service provider.

The concept of *distortion* has been referenced several times earlier in this discussion as a negative outcome related to other forms of information dissemination. In terms of these concepts, distortion is an outcome that results from information dissemination of true or false information that makes the individual "appear before the public in an objectionable false light or false position, or in other words, otherwise than as he is."[15] Distortion captures defamation, false light, and reputational torts.

[13]Although Solove does not make it explicit in his discussion of these factors, security and increased accessibility are neither necessary or sufficient for pragmatic enforcement of privacy or for privacy violations. In contrast, they are clearly related and contribute to factors that do result in privacy violations, such as disclosure and the negative outcomes of aggregation. It is this distinction that makes clear that although there are no clear ("bright white") boundaries, the conceptual theory is bounded by the coherent combination of these related concepts.

[14]Drawn from Solove [110, p. 155]; originally from the Restatement of Torts §652C & cmt. c (1977).

[15]Referenced by Solove [110, p. 159]; originally in the Restatement (Second) of Torts §652E cmt. b.

The key distinction between distortion and other forms of information dissemination is the element of integrity intrinsic in its application. In the EU Data Protection Directive, the notion of data integrity is a protection against distortion. In this case, FIPs regarding openness, access, and change (paralleled by concepts protecting against exclusion) ensure that information about a person is not distorted and thus does not give rise to unexpected harms resulting from misinformation or false light.

3.3.4 Invasion

One way to view invasion is to envision how the effects of previously discussed concepts encroach on the activities of the individuals affected. In this sense, unconsented-to effects create disruptive, meddlesome, and disturbing incursions into the lives of the data subject. Solove identifies two subcategories of invasion: intrusion and decisional interference.

The protection against *intrusion* is most likened to Warren and Brandeis' "right to be left alone" in that it protects individuals from uninvited interactions. One of the torts is that of seclusion, which helps clarify the state that is being violated when one intrudes "upon the solitude or seclusion of another or his private affairs or concerns" [129]. Intrusion is related to a number of the concepts above in that they facilitate the intrusion: surveillance, interrogation, disclosure, aggregation, and increasing access are a few facilitators and violations that combine to create an intrusion. Solove also points out that this is not purely a spatial (terrestrial) phenomena. Rather, spam and other unsolicited communications intrude upon individuals' daily activities, potentially violating forms of seclusion no matter where the individual is terrestrially. The notion of a realm of *exclusion* is also introduced, moving away from the dichotomy of either being secluded or not. Instead, it is more akin to the contexts identified by Nissenbaum. For instance, individuals in a restaurant may not be secluded (per se) but may have reasonable expectations that they will not be stared at (surveiled). The case of *Sanders v. American Broadcasting Companies* gave rise to this interpretation of exclusion:

> [T]he concept of 'seclusion' is relative. The mere fact that a person can be seen by someone does not automatically mean that he or she can legally be forced to be subject to being seen by everyone.[16]

These two cases highlight that seclusion and exclusion are more textured than the simple dichotomy of whether one is alone or not.

Solove defines decisional interference as "interference with people's decisions regarding certain matters in their lives" [110, p. 166]. An initial evaluation indicates this is a question of autonomy, not privacy. Westin argues that autonomy and privacy are linked: the ability for individuals or groups to reason about an issue without fear of surveillance (and subsequent intervention) by the government or blackmail by others is fundamental to preserving freedom. The sources of interference are rooted in a number of the earlier concepts: risk of disclosure, avoiding engagement for fear of harms of intrusion, and an indirect link with blackmail as a side effect.

[16]Drawn from Solove [110, p. 165]; originally from *Sanders v. American Broadcasting Companies.*

3.4 Generative Metaphor

The notion of generative metaphor is a useful tool for understanding how a particular problem is framed, how that framing influences the design and implementations of solutions, and the insights that can be gained from comparing different framings [106]. Schön describes metaphor as a means to recognize that a problem has been conceived from a particular perspective and that other perspectives may shed light on previously unacknowledged characteristics or generate new insights into the problem. Schön argues that the problems developing (social) policy are rooted in how the problem is framed (by the metaphors used) rather than that means of the problem itself. Rather, the framing of the problem shapes how users and designers perceive a situation and subsequently the solutions that are applied. One framing of a problem may characterize a service as suffering from "fragmentation" and prescribe "coordination" as the remedy [106, p. 138]. Alternately, the service may be described as "autonomous," which does not imply a problem as opposed to the connotation of fragmented services having once been elements of a more integrated whole. As such, the framing of the problem shapes how it is perceived and the set of tools brought to bear in solving it.

A key difficulty Schön addresses is the difference between problem setting, which is often characterized by how the problem is framed, versus the process of problem solving. Simon's problem solver explores the problem-space to optimally satisfy some objective function [109]. Schön argues that part of this process assumes that the problem (and its framing) is given—the framing can be seen to shape the problem-space and thus may introduce unintended and potentially unrecognized constraints. In Schön's words

> Each story constructs its view of social reality through a complementary process of *naming* and *framing*. Things are selected for attention and named in such a way as to fit the frame constructed for the situation.

This process highlights the "salient" features and relations between objects, simplifying a potentially complex situation[17].

A *generative* metaphor is one that creates "new perceptions, explanations, and inventions;" in effect, a generative metaphor provides the designer with new insights that were absent and/or conceptually occluded by alternate metaphors used in other framings of the problem. In this thesis, the policy analysis provides evidence of the metaphors used by online service providers to shape users' perception of their privacy practices. To some degree, the discussion of Solove's privacy concepts has already highlighted alternate metaphors for some of the privacy practices surveyed here. The analysis surfaces implications of privacy policies that are occluded by either the absence of meaningful metaphor and information asymmetries rooted in the need for technical interpretation. Given these implications, policy implications are reframed using metaphors drawn from Solove's privacy concepts, moving away from the language of utility, services, and data *sharing* to a framing that focuses on

[17] As an aside, Schön also relates the problem setting process to a construct by Dewey referred to as the "problematic situation," which he references at [106, p. 146] and refers the reader to [39].

the exchange of privacy information for services and implicit data *collection* (or in less flattering terms, surveillance).

The process of problem setting then becomes what Schön calls "a kind of policy-analytic literary criticism" [106, p. 149] that helps understand the framing and the generative metaphors of which they are comprised. Starting with a new situation, the frame setting process suggests cognizance of existing, conflicting framings of the problem (frame conflicts) and the implications of each. Schön argues that frame conflicts are often dilemmas because the ends are couched in frames that give rise to incompatible meanings ascribed to the situation. A possible solution is frame restructuring, the process of constructing a new problem-setting story by drawing from the conflicting relations while preserving the integrity (coherence) of the new story. Schön argues that this process "gives us access to many different combinations of features and relations, countering our Procrustean tendency to notice only what fits our ready-made category schemes" [106, p. 152].

Chapter 4

Privacy Policies and Deficiencies

In a society in which the typical citizen cannot figure out how to program a VCR, how can we legitimately expect the American public to understand the privacy implications of dynamic HTML, Web bugs, cookies, and log files? The commercial models, however, are predicated on "personalization" and "customization" using these technologies.

Joel Reidenberg in E-Commerce and Trans-Atlantic Privacy [104]

Despite differences in structure and organization, privacy policies provide largely the same precision of information regarding data collection, data purposes, and implications. A content analysis shows these policies give little more information than the guidelines provided in the US Safe Harbor agreement itself. A detailed comparison of policies to the US Safe Harbor Agreement is provided in Appendix B and confirms black letter compliance. Beyond black letter compliance, this chapter provides evidence that the privacy policy sample does not provide sufficient information to either substantively differentiate online service providers in terms of privacy guarantees or to determine the implications of the data collected. Data categorizations and purposes articulated by the sample build a "service-and-utility" framing of privacy policies. In some cases, meaningful metaphors are used to convey implications, fulfilling the spirit of a privacy policy as a decision tool. In other cases, technical language unfamiliar to the lay reader, coupled with a dearth of metaphors, create information asymmetries that occlude making informed decisions. Have identified these asymmetries, deficiencies can be more clearly articulated. As per the previous chapter, information asymmetries and privacy deficiencies will be presented in terms of contextual integrity and Solove's privacy typology.

One compelling aspect of this analysis is the *lack* of precision with which online service providers describe data purposes and how this is reinforced by the "service-and-utility" framing. Following the notion of a generative metaphor contributing to design, lack of precision is also reflected in the limited range of opt-out mechanisms and the framing of the implications of these opt-out mechanisms. In both cases, online service providers satisfice to the black letter of broadly scoped FIPs but, in application, the range of options is limited in both the precision of tools and explications of

the implications of the decisions that are available.

In terms of conventional personally identifying information (PII), the categories of data are similar—notice that this data is collected and reassurances that this information is not shared, sold, or rented is common. Less clear are the data purposes and precise implications of aggregate images inferred from non-PII. Despite different structural formats, a common set of data categories were identified (Section 4.2.3.2). Data collection practices are reported in terms of these common categories, the mechanisms used to collect data, and the visibility of these mechanisms to the average user (Table 4.4 and Section 4.2.4).

4.1 Online Service Provider Categories

The privacy policy sample focuses on online service providers that collect users' personal information and have a substantive user base. The sample includes the privacy policies for Amazon.com [5], Facebook [46], Yahoo! [136], Twitter [116], MySpace [86], LinkedIn [76], Google [56], Overstock.com [93], eBay [40], and Microsoft [84]. Online service providers are categorized as social networking, information services, and sales. An online service provider may have elements of more than one of these categories, but the primary function of the online service provider general falls into one of these broad categories. For instance, eBay primarily falls into the sales category, but has distinct social networking elements that contribute to its focus on creating an online community within the markets it provides. Each online service provider was also categorized as either a *single* service provider or *platform* service provider based on the number of services and whether these services fell into the same basic functional category. A service is considered single if the differentiated features serve and support the same primary function. For instance, Amazon has social networking features, but its primary objective is to sell products. A service is considered a platform if the online service provider provides a variety of differentiated services supported by a common technical platform. Facebook is a platform because, although its primary function is social networking the platform provides services and an API for developing embedded applications that are as varied as the types of information that can be published on Facebook. In addition, the platforms surveyed here all allow developers to create online applications that have different privacy guarantees than the hosting online service provider. Table 4.1 summarizes these categorizations.

4.2 Content Analysis

This analysis highlights the information that contributes to understanding an online service providers' data purposes, what data purposes it allows, and attendant implications. Despite structural and organizational differences (Section 4.2.1), privacy policies convey data collection methods (Section 4.2.2), types of data collected (Section 4.2.3.2), data purposes and sharing practices (Section 4.2.4), and opt-out mechanisms (Section 4.2.6).

Online Service Provider	Category	Services
Amazon	Sales, social networking	Single
Facebook	Social networking	Platform
Yahoo	Information services, sales	Platform
Twitter	Social networking	Single
MySpace	Social networking	Platform
LinkedIn	Social networking	Platform
Google	Information services, social networking, sales	Platform
Overstock	Sales	Single
eBay	Sales, social networking	Single
Microsoft	Information services, social networking	Platform

Table 4.1: Functional Categorization of Online Service Providers Each category comprises the primary categorization and a subcategorization. The subcategorization indicates that elements of this category are used by the online service provider in support of the functions comprising the primary category.

4.2.1 Structure

Information service providers, especially those that host a number of different services and provide a developer platform, typically provide an umbrella privacy policy and supplementary policies for each service. The umbrella policy provides general guidelines for how information is used across the family of services provided by an organization, typically capturing data purposes such as database management, billing, logistics, data analysis, etc. If a specific service functionality deviates from the assertions made by the umbrella policy, these are described in a supplementary policy that describes these differences and that serves as notice. Yahoo! is an instance of an information service provider that publishes a primary privacy policy [136] and a list of supplementary policies [137] that describe the differences in what data is collected, the purposes to which this data is put, and with whom the data is shared.

4.2.2 Data Collection Methods

Data collection methods described in privacy policies can be partitioned into explicit, user-generated information sharing and tacit (automatic) data collection. Any information explicitly shared or generated by a user via a web form, through either free-form text entry or by selecting a discrete data element from a list, is considered explicit data sharing. The presence of some data entry method (typically a web form) makes this an active form of interrogation. Information shared explicitly includes name, e-mail address, gender, date of birth, etc. Explicit information may be required or voluntarily shared. Required data such as user name, e-mail, password, and, if purchases are made, payment information and physical address comprise a

Online Service Provider	Supplementary Policies?
Amazon	No
Facebook	No
Yahoo	Yes
Twitter	No
MySpace	No
LinkedIn	No
Google	Yes
Overstock	No
eBay	No
Microsoft	Yes

Table 4.2: Yahoo!, Google, and Microsoft all provide supplementary policies describing the various services they provide. In each case, the umbrella policy indicates supplementary information can be found by visiting the service. Each of the umbrella privacy policies links to a list of services.

subset of explicit data that is typically used by the online service provider to create a user profile. Voluntarily shared information is typically information that supplements a user's profile, such as personal description, descriptions of interests, or other information that is considered to enhance the user's experience. Sharing of voluntary explicit information is often encouraged. The data purposes associated are often covered by blanket statements in other portions of the policy[1].

Tacit data collection methods collect information from the user's browser; HTTP session data automatically sent with each request, data encoded in cookies, data transmitted via web beacons, and embedded executable scripts are all automatic data collection tools. These data collection methods are passive relative to the user and may be categorized as a form of surveillance. This information includes basic browser information and comprises data elements used to construct clickstream information that contributes to inferring behavioral trends. A simple instance of this information referenced by nearly all privacy policies in the sample is IP address, browser type, operating system, and other information that accompanies HTTP requests. Clickstream information is collected by instrumenting the browser using programmable scripting mechanisms such as Javascript or VBScript. Clickstream information is used to track how users navigate the online service provider's web site, what portions of the page a user's mouse hovers over (indicating interest), and the elements within an interactive page the user viewed. Tacit information also includes search parameters included in the HTTP request as well as information about the web site the user is coming from and the web site to which the user is going, even if these are not pages within

[1]Section 4.2.3.2 describes blanket category statements relative to more precise data categorizations.

the organization's web site. Automatic information has a clear legitimate use for debugging a web site, reproducing errors, and improving the efficiency and design of a site. Automatic information also contains information that can be used to monitor and record behavioral information as a means to construct an aggregate image of an individual.

The choice to differentiate explicit information *sharing* and tacit data *collection* is intentional. The implication is that explicit action is taken by the user to share this data with the online service provider and potentially other members of a community. As an explicit action, an individual can choose to withhold information. In contrast, tacit data collection occurs automatically and users do not typically have a choice regarding which individual datum are collected, stored, and processed by the online service provider. Privacy preferences and opt-in/opt-out choices related to interests (Section 4.2.6.1) provide a certain granularity of choice, but it is not at the granularity of individual choices made when the user chooses to share (generate) information through an explicit process.

Another key distinction between explicit information sharing and tacit collection methods is how privacy policies categorize this information. In the case of explicit information, the categories are framed in terms of familiar metaphors, such as what data is personally identifiable, contact information, registration information, favorites, etc. In the case of tacit information, the categories[2] map to the mechanism: HTTP session data, cookies, and web beacons are the most commonly referenced. Although the user is a participant in the processes that activates these mechanisms, they are not as clearly linked to meaningful actions such as e-mail correspondence or group chats. Rather, reporting the techniques used to collect abstract categories of technical data satisfices to black letter notice that these data are collected. The absence of explanatory metaphors occludes recognizing and reasoning about the implications and consequences that are neither elaborated nor obvious to the casual (non-technical) user.

Some privacy policies do make the distinction between explicit (visible) and tacit (automatic) data collection. Others simply describe the mechanisms without making any distinction between explicit and tacit clear. Table 4.3 indicates whether an online service provider makes this distinction explicit. For instance, Facebook distinguishes between "personal information you knowingly choose to disclose that is collected by us and Web Site use information collected by us as you interact with our Web Site" [46]. eBay and Yahoo!, on the other hand, simply list largely operational categories of information, such as registration information, transactional information, information necessary for logistics, community discussions, interaction information, authentication information, and information collected from other companies. In contrast, Amazon explicitly categorizes data as "Information You Give Us" and "Automatic Information" [5]. This level of explicit distinction is preferred for overt clarity but does not address the fundamental information asymmetry.

For the most part, web sites describe the information processing involved with

[2]Here categories map to the naming process in the naming and framing of Dewey's problematic situations discussed in Section 3.4.

Online Service Provider	Explicit Visibility Distinction
Amazon	Explicit
Facebook	Explicit
Yahoo	Methods
Twitter	Methods
MySpace	Methods
LinkedIn	Categories
Google	Categories
Overstock	Categories
eBay	Methods
Microsoft	Methods

Table 4.3: Distinction Between Explicit Sharing and Tacit Collection
Explicit indicates the privacy policy, in one or two adjacent sentences, clearly distinguishes between categories of data roughly equivalent to explicit and automatic. *Categories* indicates this distinction can be determined by inspection of the categories of data the online service provider discusses when giving notice of the data it collects. *Methods* indicates that either the categories did not distinguish between explicit and automatic and the distinction must be determined from an understanding of the methods described, such as web beacons, client-side scripting, and cookies.

transferring information from the user to the online service provider in terms of data collection. Overstock's privacy policy takes a different tone, describing information as being provided by the user, using language referring to "details you have supplied us with" and "you may supply us with ..." [93]. LinkedIn takes a similar approach, making it clear that users are sharing information and that users bear the implications of this data sharing. LinkedIn is especially interesting because it not only takes a different tone, but stresses that the social network is geared towards business networking.

> The purpose of the LinkedIn website is to permit Users to voluntarily provide information about themselves for the purposes of developing, maintaining and enhancing a network of professional contacts. You willingly provide us certain personal information, which we collect in order to allow you to benefit from the LinkedIn website. *If you have any hesitation about providing such information to us and/or having such information displayed on the LinkedIn website or otherwise used in any manner permitted in this Privacy Policy and the User Agreement, you should not become a member of the LinkedIn community.* [76, Emphasis added.]

This is arguably different than the descriptions of the immediate impacts of information shared on Facebook, which has a substantially different culture geared towards largely personal pursuits. Here, the differences in social network purposes are manifest in the tone and implications used to frame the privacy policy.

4.2.3 Data Categories

The primary focus of most regulatory criteria for a privacy violation is the disclosure of PII. As documents framed to provide legal protection[3], sample policies stress PII, its use, handling, and storage. Non-PII is often described with the qualifier that it cannot be used to personally identify the user, often presenting it as an "innocuous" category of data used largely to improve service. This framing of non-PII, especially when coupled with the implications of aggregate (in the sense of Solove) non-PII, obscures the implications of detailed descriptions of individuals these aggregates create and the commensurate level of security that should be afforded to these descriptions.

4.2.3.1 PII and non-PII

The category of personally identifying information is rooted in forms of identification that, individually, can be used to establish contact with an individual. Most of the instances of PII given as exemplars also fit the characterization of either contact information (name, telephone number, address) or unique information that is linked to validating an individual's identity (driver's license, social security number, credit card number). Both of these classes of PII provide a gateway to accessing additional details about an individual by either directly contacting the individual in person, via some communications medium, or by accessing information associated with that unique information. In terms of Warren and Brandeis' notion of privacy, the "right to be left alone," protecting PII is an efficacious way to ensure actors outside of the context in which PII is shared do not have undue access to the individual.

Of the policies in the sample, none give a formal, non-circular definition of PII; the closest is a nominal effort by Google with an example-based entry in the glossary linked to from its privacy policy. Google's definition is fairly characteristic of those in the sample:

> "Personal information" is information that you provide to us which personally identifies you, such as your name, email address or billing information, or other data which can be reasonably linked to such information by Google. [57]

While the examples are useful, they are not nearly covering (for instance, social security number and driver license number are not included). This type of definition is circular; personal information is defined as information that "personally identifies" you, which does not give any more precision than the actual term. Following the dearth of metaphors conveying meaningful implications thus far, the definition does not describe why one category of information is "personally identifying" and why others are not.

The notion of an aggregate image of a user has been used loosely based on Solove's definition of aggregation. Solove's aggregation describes the combination of heterogeneous data that incorporates multiple sources of information, creating a more com-

[3]Framing for legal protection is exemplified by differentiated privacy policies discussed in Section 4.2.4.3.

plete picture of an individual [110]. The aggregate image of an individual is the collection of individually "innocuous" instances of data (attributes) that together describe characteristics and behaviors that distinguish an individual from others characterized by the same attribute categories. In contrast to conventional PII, the aggregate image by itself may not directly link the owner of that image to the individual, but may provide a rather detailed description of their characteristics, interests, and activities. Based on the sample, an aggregate image may be inferred from the synthesis of explicitly shared information and tacitly shared data used to infer behavioral trends.

A problem with maintaining an aggregate image, even if it does not contain PII that links directly to the individual, is re-identification. Sweeney reports that 87% of the US population can be uniquely identified by an attribute triple comprised of 5-digit zip code, gender, and date of birth [112]. This data was based on 1990 US Census data. The process of mapping "anonymous" data that does not include conventional PII to a unique individual is re-identification. Given a second data set that contains a subset of the attributes in the aggregate image, it is possible to match corresponding attributes with statistical significance, such as with the Census data above. The threat of re-identification could be categorized as a form of insecurity, especially if the level of security is commensurate with the policies' claims that this data is innocuous.

The aggregate image also suffers from exclusion, or what under the FIPs would be classified as a violation of rights to access and correction. It is important to note that the aggregate image constructed by either an online service provider or an advertiser is based on the collection of tacit data and the categories these organizations have assigned to particular events and gestures. The primary means for making any change to an aggregate image is to opt-out of tacit data collection (if possible) or to completely opt-out by disabling cookies and executable scripts. The only form of access and correction (of tacit data) available in the sample was the ability to remove recent items from the Amazon shopping history (described in Section 4.2.6 and shown in Table 4.8). If categorization and inferencing mechanisms do not work as efficaciously as expected or the data set is tainted with unrelated data, the image produced may be inaccurate and may target ads embarrassing or even offensive to the user. Revealing embarrassing true information is the threat of exposure; the possibility of exposing (embarrassing) false information is a distortion of the user's image.

In both the exposure and distortion cases of information dissemination, the user is forced to trust the information steward to ensure the accuracy of the aggregate image. Vague categorizations limit validating the types of data that contribute to the aggregate image. Moreover, some of these categories may be out of the control of the user. For instance, Facebook indicates:

> We may use information about you that we collect from other sources, including but not limited to newspapers and Internet sources such as blogs, instant messaging services, Facebook Platform developers and other users of Facebook, to supplement your profile. [46]

In this quote, profile refers to the internal information maintained about an indi-

vidual that corresponds to an aggregate image. At least in the case of Facebook, the aggregate image also relies on external data sources, such as newspapers and the information about a user provided by their friends, which may not be the most accurate, desirable, or flattering.

A key distinction between the aggregate image and PII is that an aggregate image is not useful for targeting advertisements if the user with which it is associated cannot be recognized. The aggregate image may be linked to some unique value, such as stored in a cookie, in order to recognize the user's browser (and by proxy the user) and apply information from the aggregate image to customization and advertising processes when they return to the site. Most of the privacy policies in the sample indicate cookies are used to identify the user, or the user's browser, when they return to the site. As per the "service-and-utility" framing, this is nominally so the online service provider can deliver the customized services that are of value to the user. The privacy policies also indicate that cookies are not "personally identifying." This appeals to the connotation of non-PII as a single datum that does not create a direct link to the user. Thus, if "personally identifying" is framed as a description of an individual (which in common usage is also used to identify someone), then cookies linked to an aggregate image are in fact identifiers. The caveat is that the unique value (stored in a cookie or otherwise associated with the aggregate image) *is not* linked to PII. Rather, these unique values are used to recognize agents of the user (typically browsers) so online service providers and advertisers can make use of prior information, collected in potentially different contexts, to serve advertisements customized based on highly descriptive information of an individuals' browsing habits (and by proxy their interests and preferences).

For the remainder of this thesis, a profile, in contrast to the aggregate image, is considered to comprise only the explicit, conventional PII typically referred to in privacy policies (name, telephone number, address, etc.). The aggregate image is intentionally identified as *inferred* because, as per the earlier distinction between shared and collected data, the user does not have access to the information processing that contributes to the construction of that image.

4.2.3.2 Data Categories from the Sample

Online service providers describe data categories in a number of ways, not necessarily using the same categories. In the easy cases both explicit and tacit data categories were nicely laid out in a section about the data a particular online service provider collects. In more difficult cases, instances and examples of information were scattered throughout policies organized around processes rather than organized around data categories and purposes.

The following describes the categories of data surfaced in the content analysis. The labels in parentheses following these categories indicate the data collection method used. The label *explicit* indicates explicit, explicit data collection. Explicit may be modified by *required*, indicating this category is typically required for service; *voluntary*, indicating sharing this information is at the discretion of the user; *defaulted* indicating this data takes on default values that may be changed by the users. The

label *automatic* indicates information that was automatically collected via client-side instrumentation. Instances of these categories, drawn from the sample, are listed in Table 4.1.

Blanket Includes all information described in a privacy policy. This category is included here for completeness; it is used in the next section to describe the granularity of data purposes

PII / Contact (explicit, required) Information the online service provider categorized as "personally identifiable," characteristically information that, atomically, facilitates direct contact with an individual

Preferences (explicit, defaulted) Preferences and settings to be applied to online service provider features

Communication with Online Service Provider (explicit, voluntary or required) Information contained in any communication between the user and the online service provider, typically qualified over any communication medium

Communication with Users (explicit, voluntary) Information from communications between a single user and a specific, known set[4] of users that is either mediated or facilitated by online service provider features

Financial Transaction (explicit, required) Information about purchases and payment information

User Generated and Community Information (explicit, voluntary) Information generated by users that contributes to an online service provider hosted community; this is distinguished from Communication with Users because information is published to the community as a whole, rather than targeted to one or more specific users, such as instant messages or e-mail.

HTTP (automatic) Information from standard HTTP session data sent by a browser when a request is sent to the server

URL (automatic) Information that can be extracted from variables embedded in a URL

Code Instrumented (automatic) Information collected about a user's behavior that is reconstructed from client-side instrumentation tools and scripts

(Statistical) Aggregate Information (automatic) Statistical information about a particular sub-population of an online service provider's users

Other Sources (automatic) External sources of information about users that is used to either supplement or validate information collected from other categories.

Explicit Categories	Instances
PII / Contact	name, e-mail, physical address, telephone number
Preferences	Opt out settings, user communication settings
Communication with OSP	Customer service communications
Communication with Users	Messages to other users, group messages
Financial Transactions	item purchased, purchase history, payment information
User Generated/Community	blogs, wall postings, reviews
Tacit (Automatic) Categories	**Instances**
HTTP	IP address, browser type, operating system
URL	search terms, locale, search settings
Code Instrumented	cookies, web beacons, click events, mouse over events, hover events, referrer address
Aggregate Information	number of uses in a demographic range, quantity of advertisement views
Other Sources	newspaper articles, profiling services

Table 4.4: Instances of Categories of Data Collected by Online Service Providers

The categories above can be partitioned into two groups: those conceptually familiar to the user (they have meaningful metaphors) and those describing technical data collection methods. These categories are not explicit in the privacy policies. They are largely derived from lists of "examples of data we collect." Familiarity with the types of data is important to conceptualizing what is actually being collected and the implications. The partitioning of Table 4.4 highlights the distinction between meaningful and technically obscure categories. Users are familiar with the data and attendant personal information that constitutes[5] the instances listed alongside the explicit categories. The automatic categories are less intuitive. Some of the instances, such as IP address and search terms, may be useful to advanced users, but other instances, such as browser type, various click events, cookies and the vague descriptions of their applications, web beacons, and the description of variables embedded in URLs are not intuitive to the casual user. In many cases, automatic categories are qualified with blanket purpose statements, such as improving services and web site customization, which may make these data categories seem like innocuous technical data. Framing these individual categories as "innocuous" is a positive alternative to the negative connotation users would associate with categorizing these as mechanisms that facilitate surveillance and the construction of possibly inaccurate aggregate images.

[4]The language is intentionally used to distinguish communication between users and the publication of information to an entire population, such as a blog or a review.

[5]The term constitutes is used precisely here; it does not imply an understanding of how these data may be combined with other visible or automatically collected data.

4.2.4 Data Purposes

The data purposes described by the policies, are, like the explicitly described data categories, imprecise and typically defined in terms of examples of general data purposes. Purposes can be categorized by data category and actors involved in the actions that comprise a particular purpose.

Categorizing along data category, the broadest purposes apply to the blanket data category, which includes all data described in a privacy policy. The blanket category is less useful as an informative, discriminatory category. It is used by online service providers to make broad statements that bind to all data collected by the online service provider and data that is covered by the privacy policy. All of the policies in the sample included some form of blanket statement that indicates that information collected from users is to improve service and provide a more customized experience. This language contributes to what has been referred to as the "service-and-utility" framing of online service provider privacy policies. For instance, one of the blanket statements from the Yahoo! privacy policy states:

> Yahoo! uses information for the following general purposes: to customize the advertising and content you see, fulfill your requests for products and services, improve our services, contact you, conduct research, and provide anonymous reporting for internal and external clients. [136]

Privacy policies typically include this information at the beginning of the policy as a blanket statement and repeat it as references to sharing information are made to reiterate the general themes (metaphors) of service improvement driven customization.

Applying Schön's framing awareness, the implied dichotomies in these privacy policies are "customized/generic" and "improvement/stagnation." It is arguable that these metaphors do not act as generative metaphors: they do not give insights into privacy implications or the tools that attend to these implications. An alternate framing, drawing on Solove's concepts of surveillance and intrusion, does provide insights into the tools that would improve user privacy. For instance, a "surveiled/anonymous" dichotomy might better convey the implications of engaging services that collect and aggregate "innocuous" usage data. This alternate framing may highlight the privacy implications, but obscures the benefits and utility of customized advertisements that some users may prefer.

Privacy policies indicate that the aggregate image created allows them to better target advertisements to an individual's interests and to better customize the services to users' needs. Stated at this level of abstraction, culling the advertisements to those that interest the user and using this information to further customize the service can come across as net positives in terms of users' overall utility. This framing is effective because, when individuals are presented with positive abstract categories, they tend to refine them into the more concrete categories that appeal to them individually. The *precise* interests inferred by online service provider and advertisers' construction of an aggregate image may in fact reflect the information the user viewed or categories of services the user engaged in, but the user may not consider the precise interests inferred appropriate. Thus, vague articulations do not effectively convey the

implications necessary for a privacy policy to act as an efficacious decision tool. In Solove's terms, it may be argued that, without some more precise articulation that conveys meaningful implications, general data purposes result in either blind acceptance without due consideration of the implications or decisional interference due to technical knowledge barriers.

Beyond blanket statements, privacy policies' data purposes focus on the situations under which user information is shared with third parties and advertisers. Across the sample, three categories of third parties are consistently identified: operations support, advertisers, and platform developers. Online service providers reassure users that they do not sell, rent, or trade users' PII to any third parties without users' permission. This does not cover the implications of data collected and shared by third parties themselves. The following sections describe these relationships, as articulated in the privacy policy sample and provide evidence for the mixed context problem described in Chapter 1.

4.2.4.1 Operations Support

Operations support describes general categories of information shared with third parties to provide necessary support functions. Instances of operations support provided by online service providers include order management and fulfillment, order shipping, data analysis, marketing, customer list management, search management, credit card processing, customer service, hosting, processing transactions, statistical analyses, and fraud detection (to name a few listed in the sample). The general trend in the instances given is that online service providers offload non-core services to third parties, mostly along the lines of logistics management and specialized analytics. Like data categories, online service providers do not provide precise criteria regarding what is shared.

Operations support is generally accompanied by a reassurance that PII is shared on a need-to-know basis. Only the information necessary for a third party to perform their function is shared with that third party. For instance, Amazon states:

> [Third party operations support] have access to personal information needed to perform their functions, but may not use it for other purposes. [5]

Amazon does not make mention of contractual obligations to respect users' privacy.

Other online service providers provide stronger "need-to-know" statements. For instance, LinkedIn states:

> These third parties do not retain, share, or store any personally identifiable information except to provide [operations] services and they are bound by confidentiality agreements which limit their use of such information. [76]

A number of the online service providers in the survey further reaffirm need-to-know statements with the reassurance that third party operations support is under contractual obligation that limits their use of the information. Some go even further, indicating that operations support is required to meet the same privacy standards as set out in the policy.

Online Service Provider	Operational Support Contract
Amazon	simple reassurance
Facebook	unspecified contractual
Yahoo	unspecified contractual
Twitter	unspecified contractual
MySpace	simple reassurance
LinkedIn	unspecified contractual
Google	binding
Overstock	simple reassurance
eBay	unspecified contract[a]
Microsoft	simple reassurance[b]

Table 4.5: Categories of Data Collected by Online Service Providers
The category *simple reassurance* indicates that there is a need-to-know statement regarding information shared with and information purposes of third parties. The category *unspecified contractual* indicates the online service provider claims contractual limitations, but does not specify more. The category *binding* indicates that the online service provider indicates contractual limitations guarantee the third party will adhere to the privacy standards set forth in the privacy policy.

[a]This is the loosest invocation of unspecified contract. All of the others are discussed in the context of protecting privacy, but the description in eBay's privacy policy [40] only indicates that operations support is "under contract" with no specific privacy connotations.

[b]This is the strongest of the simple reassurance category. Microsoft indicates operations support is "required to maintain ... confidentiality," [84] implying, but not specifying, a contractual enforcement mechanism.

Table 4.5 summarizes online service providers privacy guarantees relative to third party operations support. Although the difference in guarantees is useful for liability purposes, the specific information shared and the actual third party identities are not fully disclosed. Some online service providers provide partial lists of third parties; all of these qualify any list with the disclaimer that these lists are not complete, may change, and are not authoritative. Moreover, these qualifications usually refer to PII; sharing non-PII and the resultant aggregate images is not typically addressed. This implies non-PII and the associate aggregate image may not receive the same level of security protection as PII, even though it is arguably very descriptive. Coupled with the potential for re-identification, this framing may actually conceal re-identification privacy risks by demoting non-PII to a second-class category of information with respect to information security requirements.

4.2.4.2 Advertisers

Advertisers are the other primary category of third party addressed by online service provider privacy policies. All privacy policies make some mention of advertisers. The first commonality is that all online service providers indicate that their privacy policy binds only to the online service provider. Advertisers (and Platform developers in the next section) may have different privacy policies; the user is advised to review those

Online Service Provider	Specificity
Amazon	explicit
Facebook	none
Yahoo	explicit
Twitter	none
MySpace	implicit[a]
LinkedIn	explicit[b]
Google	implicit
Overstock	none
eBay	implicit
Microsoft	explicit

Table 4.6: Specificity of Third Party Advertiser Segmenting Implications Three levels of specificity of third party advertiser segmenting implications were identified in the sample: explicit, implicit, and no mentions. The category *segment* indicates an explicit statement indicating advertisers may assume marketing segments is included in the online service provider privacy policy. The *implicit* category indicates that examples of segments are given, but a conceptual description of segmenting is not elaborated. The category *none* indicates neither explicit nor implicit mention of segmenting was given by in the privacy policy.

[a]The discussion of non-structured profile information used to serve advertisements could be interpreted as implicit and thus it is categorized as such.

[b]LinkedIn is exceptional in this regard, providing substantial information about the implications of information sharing.

policies[6]. This contributes to the mixed context problem. A second trend is that all of the online service providers in the sample make at least one reassurance that information shared with advertisers, either directly by the online service provider or collected via cookies or other instrumentation, is not personally identifiable.

Although advertisers do not have direct access to PII held by an online service provider, the online service providers admit that advertisers do have direct access to the user through interactions with advertisements served directly from advertiser-owned hosts and/or through web beacons. Typically, the discussion follows the themes established regarding information online service providers do share with advertisers, namely that non-PII is not personally identifiable. Some online service providers (summarized in Table 4.6) do make mention of simple implications of advertisers collecting information about users. With the limited exception of LinkedIn, disclosure of the implications of advertiser data aggregation is a simple statement about segmenting.

An important distinction needs to be made between discussing the implications of segmenting and the practice of segmenting. The objective of many marketing campaigns is to use combinations of user segments to effectively target ads. Online service providers are not required to provide users with any additional information regarding what third party advertisers can or cannot do with non-PII data they can

[6]Section 4.2.6 covers the opt-out choices that sometimes accompany these suggestions

collect[7]. The following is an instance of a segmenting disclosure statement from Yahoo!:

> Yahoo! does not provide any personal information to the advertiser or publisher when you interact with or view a targeted ad. However, by interacting with or viewing an ad you are consenting to the possibility that the advertiser will make the assumption that you meet the targeting criteria used to display the ad. [135]

This particular statement is somewhat abstract. Others give concrete examples of the type of segmenting that may occur

> Although Amazon.com does not provide any personal information to advertisers, advertisers (including ad-serving companies) may assume that users who interact with or click on a personalized advertisement meet their criteria to personalize the ad (for example, users in the northwestern United States who bought or browsed for classical music). [5]

These are both examples of explicit disclosure of simple, uncombined segmenting practices.

4.2.4.3 Platform Developers

Platform developers are another category of third parties that can collect information from online service provider users. As the name implies, platform developers create applications for online service providers categorized as platforms (see Table 4.1). Like the content served by or on behalf of advertisers, online service providers reiterate that the privacy policy only binds to the online service provider and that users should review the privacy policy of the platform developer before using the application or sharing information with or via the application. A range of obligations, relative to the online service provider privacy policy, were observed and are summarized in Table 4.7.

The implications of these differentiated obligations is, like the dilemma with advertisers, that the user is again confronted with another content provider that may impose a different set of privacy rules. For instance, even though the user may have established a relationship with LinkedIn, they must also trust that LinkedIn's partners will behave similarly. In the case of Facebook, users are exposed to applications that may or may not be aligned with the expectations derived from interactions with Facebook itself. As a final example, MySpace provides a third party application platform, but does not support (or encourage) using these applications in its privacy policy:

> MySpace does not control the third party developers, and cannot dictate their actions. When a Member engages with a third party application, that Member is interacting with the third party developer, not with

[7]The FTC, in [51], strongly encourages online service providers to disclose information about how non-PII is used and its implications. As of July 2009, the FTC has not mandated disclosure of practices regarding segmenting or combination of non-PII.

Online Service Provider	Platform Developer Obligations
Facebook	non-contractual
Yahoo	none
MySpace	none
LinkedIn	
Partners	contractual protections, vetted
Standard	contractual protections
Google	none
Microsoft	none

Table 4.7: Platform Developer Obligations with Respect to Online Service Provider Privacy Policy

The content analysis surfaced three categories of platform developer obligation relative to the online service provider's privacy policy: contractual equivalent, contractual protections, non-contractual, and none. The category *contractual equivalent* indicates that the privacy policy indicates platform developers are contractually obligated to protect users privacy at the same level as described in the online service provider privacy policy. The *contractual protections* category contains online service providers that indicate protections are contractually enforced, but does not strictly indicate they are equivalent to those described in the online service provider's privacy policy. The category *non-contractual* indicates that the online service provider asserts platform developers are required to respect users' privacy, but indicates neither contractual obligations or the level of protection relative to the online service provider's privacy policy. Finally, the category *none* indicates no mention of privacy protections is made other than encouraging the user to read the privacy policy of the platform developer. In addition to these categories, an adjacent category, *vetted*, is used to indicate a subset of platform developers have a trusted status with the online service provider and automatically have some level of access to user information.

> MySpace. MySpace encourages Members not to provide PII to the third party's application unless the Member knows the party with whom it is interacting. [86]

The differentiated privacy regimes within a single online service provider architected environment is arguably confusing and forces the user to balance expectations about a relatively known environment (the online service provider) with expectations about a foreign, poorly understood environment (a platform application, advertiser content, and/or combinations of these). Even assuming the user reviews the conflicting policies, there is no guarantee the privacy policies of the platform applications are aligned with users' privacy preferences or the policies of the online service provider.

Confounding this problem is the issue of collateral damage from a users' friends using third party applications. The problem arises when a user Alice is a friend of Bob, who uses an application C developed by Charlie. Alice has chosen to share information set I with her set of friends F, of which Bob is a member. Alice does not wish this information to be shared with others outside of F. Even though Alice does not use platform application C, it may have access to some of the information only intended for (context) F because application C may access any information available to Bob. As a result, Alice's preferences may be violated inadvertently by Bob through his use of application C.

Of the platforms listed in Table 4.7, only LinkedIn explicitly discloses this issue. LinkedIn describes the problem:

> If you, your connections, members of your network, or other Users of LinkedIn use any Platform Application, or if you interact with a Platform Application being used by any of them, such Platform Application may access and share certain information about you with others. Because a Platform Application can make calls on behalf of the User interacting with it to access non-public information, the level of detail accessible by the Platform Application is constrained by the same settings that govern visibility of your data to that User on LinkedIn. [76]

Under this scenario, an immediate recourse to preserve the privacy preferences expected, based on who one has selected as a friend, is to only connect with people who are not running untrusted platform applications. This distorts privacy preferences because the trust relationship between users is no longer based on actions taken directly by the parties involved, but by action taken by their associates, here, a third party application.

The preference distortion can be explained in terms of bounded rationality and rational ignorance [109, 90]. Bounded rationality indicates that humans have a limited capacity for reasoning about a situation in a finite amount of time, thus bounding pure rational choice that assumes perfect information and sufficient time to process all necessary information [109]. A consequence of bounded rationality is rational ignorance [90]. When faced with more information than an individual can process, individuals choose to address the issues they perceive (through their bounded understanding of the problematic situation) to be most salient. Under a model of rational

ignorance, rather than spending the time to investigate individual X's application usage practices to determine whether X poses a threat, the user may artificially limit their social network to only those individuals they already know well enough or, more likely, assume nothing bad is going to happen and not even consider the issue of an application's access to their information. In either case, if the user even considers mixed context, the user's actual preferences may be distorted by a context comprised of elements with different policy guarantees and potentially conflicting privacy implications. This confounds the process of making privacy decisions based on previous experience with the architect of the environment, the online service provider.

4.2.5 Mixed Context

The information that contributes to identifying online service provider environments as mixed contexts is also interleaved with discussions of information available to advertisers, the tacit data collection methods used, and the implications of user segmenting. Providing the information necessary to identify these implications is referred to as *mixed context disclosure.* Mixed context disclosure builds on online service provider's careful articulation that their privacy policy only binds to their content. Their policy does not apply to content outside their domain that they link to or to content provided by and embedded in the environment by third parties. In that sense, when a user visits an online service provider, they are intending to visit a particular context, but because of differentiated privacy policies that are tacitly imposed when advertisements and web beacons are embedded in an environment, they are actually exposed to multiple sets of rules regarding how their behaviors will be recorded and analyzed. Although context is not explicitly discussed by the online service providers in their policies, the existence of differentiated context is surfaced from the implications of online service providers as mixed content environments.

The problem of mixed context is confounded by the fact that the online service provider architected environment comprises content contributed by a dynamic set of actors (observers, in the sense that they can each collect information about users). These actors are referred to as content providers. It is arguable that static configurations of content providers (online service provider, advertiser, developer) could be reconciled into a consistent representation of context that draws on the privacy rules set out by these contributors. In Nissenbaum's terms, each of these particular configurations is a source of privacy norms, would comprise a unique set of actors, and may elicit a unique privacy response from the user. Rather, while the online service provider's privacy policy remains the same, the advertisements are not guaranteed to originate from the same actor upon every engagement. Considering advertisements are presented based on the inferred preferences of the user based on the aggregate image and advertising segments associated with the primary online service provider content being viewed, advertisements may change with each visit, presenting the user with an ephemeral context whose privacy implications have not been considered. The user is faced with two burdens: understand all possible combinations (perfect information) or following up on each new configuration (inviting rational ignorance) as they occur. The first is impossible. The second is made difficult because configura-

tions are ephemeral and unpredictable and do not occur with sufficient frequency to develop expectations even if the user was aware of their existence and implications. A possible solution to this problem is presented in the next chapter, focusing on tools that highlights online service provider architected environments as mixed contexts and the implications for the reputations of contributing actors.

Moving to the substantial implications of mixed context, in terms of norms of appropriateness and norms of dissemination, mixed contexts allow advertisers that serve advertisements across domains to construct detailed images of individuals based on the categories the advertisers have targeted. In term's of Solove's information dissemination, this is a form of increased access; as per earlier discussion, increased access does not constitute a privacy violation itself, but the disclosure, exposure, or distortion resulting from increased access does. Recall online service provider's limited categorizations of online service provider segmenting disclosure in Table 4.6 and the quotes from Yahoo! and Amazon in Section 4.2.4.2. The advertising agent can construct a more detailed aggregate image of the user based on a variety of segment information, collected from a variety of contexts, by combining of a unique, advertiser assigned cookie, knowledge of which user segments advertisements are targeted to, and repeated exposure (across contexts) to a particular advertising agent over the course of multiple sessions. The advertiser may also know the URL and content categorizations of the environment in which the advertisement is being served, providing further information to contribute to an aggregate image of the user. In effect, each of these environments from which information is collected may be considered a context in which the user has "shared" certain information. As a violation of contextual integrity, acting on information collected in one context (violation of dissemination) may be inappropriate in other, unrelated environments.

In terms of Solove's categories, the information dissemination related deficiencies described above constitute an intrusion. In a sense, context creates a boundary and the integrity of that boundary is maintained by the norms of dissemination and appropriateness. An example is an individual reading a forum about Viagra in which an advertiser is legitimately serving advertisements. Using either the content categorization of the forum or the segment information the advertiser is targeting, the advertising network modifies the user's aggregate image to reflect an interest in Viagra. The user later visits a travel web site looking to book a vacation for a few days trip to Bermuda. If the same advertising network serves advertisements for the travel website, recognizes the user, and serves an advertisement such as "Don't forget the Viagra" or "Find local drugstores selling Viagra in Bermuda," this constitutes an intrusion. Based on contextual integrity, information about the user's interest in Viagra may be inappropriate for the travel context, especially if he is discussing travel plans with other individuals that see the advertisement and assume it has been targeted.

4.2.6 Choice and Opt-Out Options

The variety and scope of choice is critical to the efficacy of privacy policies and the tools supporting users' privacy decisions. Like the limited information conveyed by

general categories of data and purposes, the tools supporting opt-out are also limited in scope. For instance, in many cases it is unclear whether opt-out means the collection of tacit data is disabled or whether it continues and the associated advertisements are no longer displayed. General categories of opt-out were identified and apply to various categories of data and data collection methods. Notably absent were online service provider supported capabilities for limiting engagement with specific agents of advertisers and third party applications.

4.2.6.1 Categories of Opt-Out

Three opt-out models were identified in the sample: complete, recent activity, and categorical. The other dimension of opt-out models that received little attention is the ability to opt-out of data collection by some services provided by a platform online service provider, but allowing data collection by others of the same online service provider. This problem also extends to advertisements embedded in online service provider architected environments and to a certain degree platform applications. The problem with managing precisely what agents a user is engaging with will be referred to as the *agent selection problem.*

The first opt-out model is referred to as *complete* opt-out and opts the user out of all tacit data collection by the online service provider. One mechanism for this opt-out model is to disable cookies for either that online service provider or for the browser as a whole. In the policies, this option is often accompanied by an admonition of this practice by the online service provider, indicating the user may not garner the full benefits from the service. For instance, Amazon's privacy policy states:

> The Help portion of the toolbar on most browsers will tell you how to prevent your browser from accepting new cookies, how to have the browser notify you when you receive a new cookie, or how to disable cookies altogether. Additionally, you can disable or delete similar data used by browser add-ons, such as Flash cookies, by changing the add-on's settings or visiting the Web site of its manufacturer. However, because cookies allow you to take advantage of some of Amazon.com's essential features, we recommend that you leave them turned on. *For instance, if you block or otherwise reject our cookies, you will not be able to add items to your Shopping Cart, proceed to Checkout, or use any Amazon.com products and services that require you to Sign in.* [5, Emphasis added]

Another instance discussing disabling cookies is per Overstock:

> We use cookies to provide you with the best level of service by keeping track of what you have in your shopping cart, remembering you when you return to our store and using them to help us show you advertisements which are relevant to your interests. They cannot harm your computer and they do not contain any personal or private information. You must accept cookies to shop through the Site. You can erase or block cookies from your computer if you want to (your help screen or manual should

> tell you how to do this), but the Site may not work correctly or at all if you set your browser not to accept cookies. [93]

The Overstock quote is an excellent example of both the admonishment of disabling cookies and the "service-and-utility" framing. Note that cookies are presented as a necessary tool for maintaining a user's shopping cart. They are also used to help show the user advertisements that are "relevant" to the user, again appealing to the "customized/generic" framing discussed earlier. The last portion of both the Overstock quote and the Amazon quote (and common to all of the privacy policies in the sample), is the "all-or-nothing" character of the admonishment. If one disables cookies for the online service provider, both advertising and online service provider functionality is affected. The policies present the loss of utility as further deterrence to disabling cookies. Some online service provider's allow the user to opt-out of receiving targeted advertisements, but it is questionable whether this simply disables the *display* of advertisements while tacit data is still collected or whether tacit data collection itself is disabled. Table 4.8 lists which online service provider's provide this feature.

Another option to completely opt-out of the display (but not necessarily ensure the complete disabling of tacit data collection) of third party advertising is through the use of browser extensions that limit the execution of scripts and the use of cookies. Instances of Firefox browser extensions are Adblock [94] and TACO (Targeted Advertising Cookie Opt-Out) [133]. These block advertisements, but may not occlude all tacit data collection by the online service provider. For instance, even though scripts and cookies may be disabled, web beacons, data embedded in URLs, and other tacit means of data transfer may be occurring. Although these tools are useful for users that prefer not to share *any* information and that do not garner utility from targeted advertising, this is not a robust solution that acknowledges a heterogeneous space of users' privacy preferences[8]. Strategies to facilitate developing empirically informed, robust solutions are discussed in the next chapter.

The second method of managing tacit data collection is referred to as recent opt-out. This option was only provided by Amazon, but warrants discussion because of both the potential for appropriately granular opt-out options contrasted with the limitations and burdens placed on the user. *Recent* opt-out allows the user to review a list of the most recent items they have viewed that will contribute to the aggregate image of their shopping habits maintained by Amazon. If the user wishes to remove an item so it does not contribute to their shopping image, they can. This list only shows the most recent items viewed, not all items that contribute to the shopping image. To make effective use of Amazon's implementation of recent opt-out, the user must be vigilant to avoid allowing items that do not represent their genuine interests from being committed to their shopping image. Once an item has been committed to

[8]The extreme nature of some of these more represent the ideologically driven privacy preferences of some developers, couching the utility of the tool within their subset of preferences. This ideological privacy framing is the dual of the "service-and-utility" framing in that it is couched in the "surveillance/privacy" framing. Rather, a more balanced reframing would attempt to find generative metaphors that facilitate tools that attend to the most populous segments of the spectrum of privacy preferences.

Online Service Provider	Complete	Recent	Categorical
Amazon	✓	✓	
Facebook	-		✓[a]
Yahoo	-		✓
Twitter	-		
MySpace	✓		✓[b]
LinkedIn	✓		
Google	✓		✓
Overstock	✓		
eBay	✓		
Microsoft	✓		✓

Table 4.8: Opt-out Options by Online Service Provider
The denotations of *complete*, *recent*, and *categorical* are as described in the attendant paragraphs. For the denotation of complete, a "✓" indicates that the online service provider provides an option to opt-out of all ads targeted by the online service provider based on tacit data. This places the customer in the "generic" mode of advertisement targeting, which is also referred to as random. A "-" in the complete column indicates the online service provider provide a discussion of cookie-based complete opt-out. It is important to note that the denotation as complete is presented as an approximation to an ideally "complete" opt-out from all targeted advertising. Revisit the discussion of work-arounds in the accompanying text.

[a]This is a very loose form of categorical because the user is not selecting from a discrete set of categories, but is providing the positive and negative feedback necessary to dynamically infer categories of ads preferred by the user.

[b]The categorical distinction is between structured and non-structured data, not specific categories of interests.

the user's shopping image, the only recourse to "correcting" this image is to remove the entire image. This last resort can be considered a form of "all-or-nothing" opt-out because the user must forgo the utility of the customized service to avoid possible privacy conflicts, in particular the perceived distortion.

The third category of opt-out is *categorical* opt-out. This form of opt-out is also referred to under the name of "relevant" advertising [93], interest-matched (Yahoo!) [136] or what Google refers to as interest-based advertising [56]. Categorical opt-out allows the user to select the categories of advertisements they wish to see based on a list provided by the online service provider. This is akin to the advertisement blockers described earlier, but supported by the online service provider rather than implemented as a third party browser extension. An exemplary instance of the information asymmetry is drawn from Microsoft's privacy policy:

> Even if you choose not to receive personalized advertising, Microsoft will continue to collect the same information as you browse the web and use our online services. However, this information and any information collected from you in the past won't be used for displaying personalized ads. [85]

Microsoft is the only online service provider that made the fact that these preferences only affect the *display* of advertisements, not the underlying collection of data, explicit. A variant of categorical opt-out is the advertisement feedback enabled in Facebook advertisements (this was not discussed in depth in the Facebook privacy policy) and is denoted in Table 4.8.

These mechanisms only ensure that advertisements pertaining to the rough category of interests selected will not be displayed even though that information may still be collected and used. In effect, this linkage illustrates a relationship between the kinds of information conveyed in a privacy policy and the tools developed to support this policy. Absent precise categories and implications, users do not have a baseline of comparison for determining whether a privacy tool does in fact protect them from the privacy implications that may give rise to harms. In effect, these tools are not opt-out options in the sense of the complete opt-out that ensures opting out of the personalized advertisements *because* the user has opted out of tacit data collection. The distinction is whether the primary harm is (a) the nuisance of unwanted advertisements and the potential discomfort attendant with those or if the harm is (b) the maintenance of an aggregate image. The next chapter will link the problem of vague categories and the dearth of implications to absent standards setting mechanisms and tools of engagement at the institutional level.

Chapter 5

Hybrid Solutions

> ... legislative processes can be time consuming and have the potential to be out of date before they can be enforced. That's why self-regulation is a crucial tool for industry to be able to react quickly to immediate policy needs.
>
> *Luc Delany in Google's European Privacy Policy Blog [38]*

This chapter presents approaches that highlight the trade-offs amongst privacy actors. These approaches do not necessarily propose one substantive privacy policy over another, but rather proposes processes that facilitate the design of privacy policies and standards. The essence of this approach creates a feedback process that uses constructive conflict to evaluate the framing and requirements associated with privacy policies and the supporting set of tools available to online service providers, advertisers, civil society organizations, developers, and regulators. Through a process of issue reframing and empirical feedback, shared concerns and strategies may be developed that highlight mutually beneficial strategies rather than pursuing actor-specific first-best solutions. To ground this process in familiar problems, applications to the mixed context and aggregate image problems will be demonstrated.

5.1 Design of International Institutions

The discussion of governance forums follow Koremenos' definition of international institutions "as explicit arrangements, negotiated among international actors, that prescribe, proscribe, and/or authorize behavior" [74, p. 762]. The institutional design elements and characterizations presented by Koremenos are intended to develop institutions that "help resolve problems of decentralized cooperation" [74, p. 766]. Koremenos identifies five key dimensions of institutional design: membership rules, scope of issues, centralization of tasks, rules for control, and flexibility. In terms of these dimensions, the hybrid solution proposed here attempts to correct the failures of the US Safe Harbor compromise.

The EU Data Protection Directive and US Safe Harbor (henceforth the current institutions) were designed by an *exclusive* membership, excluding key actors with

dissonant views. The current institutions also have scope issues, attempting to cover many different privacy contexts with a single, abstract set of privacy principles. The two institutions differ in centralization, but fail in that they go to extremes. In the case of the EU Data Protection Directive, centralization of regulation is a limiting factor, especially considering that it is driven by a single category of actor (regulators) and it is unclear how effective the EU Data Protection Directive has been [75]. The US Safe Harbor is too decentralized an application of the self-regulatory regime, giving rise to the information asymmetries that inhibit self-regulatory processes from binding. In terms of solutions, both institutions place (centralize) the burden of privacy requirements elicitation and privacy tool design and implementation on online service providers and advertisers. Paralleling the problems with centralization, the rules for controlling the institution are overly rigid and overly flexible, respectively. In terms of the EU Data Protection Directive, little has changed since the days of mainframes, despite substantial discussion in Working Papers[1]. With regards to the US Safe Harbor, the information asymmetries and perceptions created by the SH-FIPs obscure emerging privacy deficiencies that could incentivize changes to notions such as what is and is not personally identifying (viz. the discussion of PII and aggregate image). Finally, as products of legislative regulatory processes, both institutions are flexible, as evidenced by the laxity of data categories and data purpose articulations. What is absent is the incentive to leverage this flexibility into the design of more precise standards of privacy, here centered on context-based privacy regimes. These failures are taken as learnings that may be applied to the design of more efficient and effective institutional arrangements for developing privacy standards.

The objective of the process described here is to establish constructive conflict amongst actors whose interests are tied to online privacy regulation and the attendant standards. Drawing on Nissenbaum and Solove's theories of privacy, an effective privacy regime should construct privacy standards based on common framings of empirically-informed contexts. This requires an institution that is sufficiently resilient to maintain a dialog among actors with disparate interests[2]. To this end, the membership of the privacy standards forums (henceforth simply forums) are inclusive, comprising government regulators, online service providers, advertisers, developers, and civil society organizations. Although there is a need for a privacy regime that operates as an umbrella for managing forums, this discussion will focus on developing an empirical understanding of the tools of engagement used within the forums themselves. Concrete instances of this process are presented through application of to aggregate image and mixed context problems (Sections 5.4.1 and 5.4.2, respectively). While forums are centralized venues, the strategies presented structure the discussion of trade-offs and facilitate negotiating a *distribution of tasks* that avoid placing undue burden on individual actors and that lend validity and legitimacy to outcomes.

The basic rules for controlling the institution are posited here as an agreement amongst government regulators. As a backstop against disintegration or empty for-

[1]The reference to the days of mainframes is attributed to Kuner [75], but can also be linked to the control metaphors common to the era when IGOs initially developed the FIPs (see Section 2.1).

[2]See [31] for an interview with the FTC's head of the Bureau of Consumer Protection that includes discussion of this disparity.

malism, government regulators will act as regulatory authority of last resort in the event of deadlock amongst actors. Considering the current response to the threat of FTC regulation [32, 34, 59, 58] and the aversion for static regulation, it is argued this institutionalized backstop will incentivize genuine participation. Other rules will be required to ensure continuous progress, but these may be developed by forum participants based on experience in initial forums, again enforced by the government regulators as a backstop to deadlock or disintegration. Substantive decisions should be a product of non-governmental actors (as those closest to the necessary data and with the motivation to engage in the process), relegating regulators to a position of ensuring self-regulatory mechanisms are in place. In other words, regulators are present to ensure progress is made toward substantive privacy standards, but do not directly contribute to specific outcomes. As may be apparent, these rules are intended to provide sufficient flexibility in allocating substantive tasks such as data collection, analysis, standards evaluation, and validation to relevant actors that best fit the problematic situation and relevant context(s).

5.2 Surfacing Requirements

The proposed approach for "policy driven" design leverages a critical analysis[3] of privacy policies to facilitate a (re)framing process that provides both critical insights into the technologies described and that surfaces requirements that inform more concrete design of the attendant tools. Schön's generative metaphor has been presented in Section 3.4. Earlier discussions have focused on the use of meaningful framings that comprise metaphors that overcome information asymmetries rooted in technical knowledge barriers. Here, the notion of a generative metaphor is leveraged to create "new perceptions, explanations, and inventions" [106] in the domain of privacy policy articulations and the supporting tools. In particular, the (re)framing process is intended to draw a critical eye to the "ready-made category schemes" extant in a particular group's framing of the privacy problem. Through a process of constructive conflict, facilitated by the forums, the reframing process engages actors' ready-made categories to construct framings that provide mutual benefits.

This thesis has presented three possible framings of the privacy problem. The first, evidenced in the privacy policy sample, is the "service-and-utility" framing focused on notions of customization as a means to improve service. A competing framing is centered on the "surveillance/privacy" metaphor, noting the negative connotations of surveillance may dominate any notion of positive benefits from customization. The third framing, comprising notions of the aggregate image (versus innocuous individual categories) and mixed-context (as opposed to homogeneous) as central concepts, is a product of this policy analysis. As generative metaphors, these framings have been used to articulate insights into the privacy implications of engaging online services and the deficiencies in the accompanying tools. Constructed over the course of this thesis, this framing is sufficient to highlight the privacy deficiencies, but would require

[3]The term critical is in the literary analysis sense and serves as a link between Schön's generative metaphor and Antón's goals.

empirical analysis to determine whether they are meaningful to a wider audience. Codifying the insights from the generative metaphors resulting from (re)framing is where Antòn's GBRAM is applied.

A key premise of Antòn's work is the idea of surfacing system requirements from organizational artifacts, effectively grounding requirements in empirical evidence. Antón's thesis [6] develops the GBRAM in an organizational setting, drawing on interviews with employees, reviews of process documentation, and reviews of other organization artifacts that convey perspectives on how the organization achieves its goals. The objective of GBRAM is to extract natural language representations of the goals articulated[4]. Natural language representations facilitate comparisons of goals for consistency, categorizations that can help identify common themes that may have been previously scattered throughout an artifact or across multiple artifacts, and identifying conflicts amongst these goals. In more recent work, Antòn has applied GBRAM to online privacy policies, identifying goals that support privacy and identifying goals that derogate privacy protections with exceptions necessary for efficient operations, among other objectives [9, 14, 11, 8, 10, 7, 12, 13].

A key benefit of this process is that the qualitative, textured narratives elicited from organizations are translated to structured natural language that is sufficient for further refinement into more formal software requirements. In addition to the benefits for comparison, this intermediate form also preserves the vocabulary familiar to organizational actors. The common vocabulary facilitates ongoing discussion and refinement of the accuracy and precision of both policies and requirements before they are translated to more formal representations foreign to the lay user. An important distinction for the application proposed here is the validity of the source of goals; Antòn indicates that:

> [W]e believe that stakeholder disagreements can be incorporated into the goal-based framework simply by admitting multiple sets of goals, indexing each set with the stakeholder that wishes to achieve them. We leave it to the politics of the situation to determine which set of goals (which stakeholder) will prevail. [9, p. 141]

Translating to this work, a manifestation of multiple, conflicting sets of goals are the conflicting framings of privacy policies.

The collaborative approach to policy-driven design will inevitably give rise to conflicts between actors with differentiated vested interests in online service platforms. The forums are intended to give rise to constructive conflicts rather than dysfunctional "compromises." As discussed throughout the analysis, differentiated interests and perceptions (i.e., Schön's ready-made categories) give rise to differentiated framings of the problem. Problem reframing is a process for resolving conflicts over interest-based framings, producing a common framing that is more suitable to an inclusive set of actors. As per Schön, this is not a synthesis of existing framings. Rather, reframing identifies alternate *generative* metaphors that surface novel insights into the system and the design process. Reframing is an iterative process that, given institutional

[4]Or uttered, if you prefer the anthropological and ethnographic terminology.

incentives for ongoing collaborative forums, can contribute to developing context-informed privacy standards derived from meaningful policies that have been vetted by actors dependent on privacy regulation outcomes[5].

Although this process can potentially reduce conflicts over policy framing and provide insights into the design of privacy enhancing[6] tools, it does not resolve the problems of having sufficient empirical understanding of common privacy contexts and expectations. This approach also recommends the instrumentation of online service provider architected environments to include mechanisms that convey (to users) why salient elements of the online environment behave the way they do and to include mechanisms for providing feedback that can be used to better understand users' perception and understanding of a particular context. Under fear of regulation by the FTC, the advertising industry has taken a first step to providing additional information to users regarding why they are seeing a particular advertisement and how behavioral data is used to customize advertisements [32, 33]. The mechanism proposed is a small icon that will appear along with advertisements and that, when selected, will describe why the advertisement is being served. One criteria for this tool should be whether the information provided is meaningful, or if it simply satisfices by providing vague articulations under the "service-and-utility" framing.

5.3 Institutional Arrangements

The members of privacy standard development forums will require input from representatives of government privacy regulators, online service provider legal and development representatives, representatives from advertising networks and agencies, representatives of platform application developers, and civil society organizations (as a proxy for users). In contrast to the interest-biased memberships that designed the EU Data Protection Directive and US Safe Harbor, respectively, these categories of actors represent the (inclusive) variety of interests in privacy standards. For instance, the initial draft of the EU Data Protection Directive excluded business representatives, relying on government authorities focused on enforcement mechanisms. This closed membership resulted in enforcement mechanisms perceived by American businesses as "onerous" threats to innovation in the information economy. Similarly, the US Safe Harbor was designed by the Clinton administration and industry interests. In this case, the closed membership pandered to e-commerce interests and derogated the EU Data Protection Directive FIPs in favor of a self-regulatory compromise embodied by Aaron's 'safe harbor.' In both cases, rather than attempt to balance interests, these institutional configurations simply excluded incompatible interests.

Taking the requirements elicitation process described previously as a set of "meta-

[5]Although these forums seem to be difficult to construct, instances will be described briefly in the context of institutional configurations that can leverage this process to resolve the problems related to aggregate image and mixed-context.

[6]Privacy enhancing does not necessarily imply nominally stronger privacy outcomes in the sense of protecting more information. Rather, it speaks to facilitating more accurate representation and expression of user's actual privacy preferences.

requirements," a number of possible configurations of actors may be possible. For instance, the choice of FIPs as a comprehensive policy-structuring choice and the relative uniformity of the privacy sample provides little evidence for subtler distinctions between different services, such as sales, auctions, search services, social networks, and other categories falling under the umbrella of online services. Depending on the interests and the willingness to collaborate, different strategies of collaboration may be necessary for different domains. For instance, the types of context specific privacy practices and the attendant configuration of actor responsibilities may be different for an online auction site such as eBay than for advertising, photo sharing applications, or even differences between the privacy standards for different domains of social networking (recall the different tone of LinkedIn versus Facebook). The differences in roles may be rooted in vested interests, relative political power, and the real and perceived burdens ("onerous" or not) imposed on actors in one role or another. A common theme in the configurations discussed below will be the role of government regulators, in particular the FTC. The FTC's unique jurisdiction over online service providers may allow it to operate as a mediator that incentives collaborative discourse on policy and technical directions rather than directly participating in the production of substantive policy.

Taking the FTC as the mediator and online service providers' legitimate argument that self-regulation is the more flexible and responsive regulation process[7], the objective is to facilitate collaborative standard setting under the umbrella of self-regulatory paradigm. One outcome is to ensure policies convey meaningful privacy implications and inform users of the implications of engaging online service providers, empowering users to make informed decisions and moving the character of online services closer to experience goods. Another element necessary to understand how to balance the interests involved is considering the current and possible tools of engagement that may be brought to bear by each category of actors. Considering the recent failure of P3P [41] and Bennet's critique of static data purposes and categories [18], one focus is to understand and design tools of engagement. This includes two distinct objectives: investigating the categories of tools of engagement and how to minimize the transaction costs of designing and implementing these tools.

Tools of engagement range from constructive conflicts that illustrate how best to allocate the burden of designing and implementing privacy tools to the collection of empirical data on users' experiences with privacy tools and the surfacing of user requirements. The latter revisits the discussion of understanding context and the data collection necessary to accurately characterize norms of appropriateness and dissemination. One objective of privacy standards setting forums is to develop tools for collecting data about users' privacy choices, the types of contexts they would like to

[7]Self-regulation is a very broad term and captures a wide variety of institutional arrangements. The proposals presented here most resemble notions of private ordering [107] and private authority [127]. These describe a shift from administrative law arrangements, such as substantive rule making performed by an administration such as the FTC, to private authorities comprised of private actors. The efficacy and efficiency of these, in contrast to more conventional arrangement is discussed abstractly by Macey [78]. An instance of a successful, self-sustaining implementation of private authority (a "private" legal system) is regulation of disputes within the cotton industry [22].

construct, and which tools are most efficacious. This in itself has privacy implications and would require explicit opt-in from users that are interested in contributing information about their privacy behaviors and preferences towards improving the related tools. Such a "policy experiment" would require careful design and protection of users' privacy, but could also yield substantial grounded information about how real users make privacy decisions in real environments.

Early laboratory experiments contributed to understanding how users use privacy policies and the divergence between actual and espoused privacy preferences [4, 20, 29, 30, 62, 67, 99]. While valuable, much of this work is based on coarse grained categories of privacy. Here, the objective is to continuously surface common privacy motifs from empirical evidence of privacy choices made *in situ*. Examples of these might comprise how users create groups of friends to share information with and how they categorize different kinds of friends (acquaintances, individuals they interact with terrestrially on a regular basis, categorizations based on terrestrial context, such as work, family, and hobbies, etc.). Understanding these categorizations may facilitate operationalizing these motifs into empirically informed templates of common privacy configurations. These templates may provide an approximation that can be further customized to match unique user preferences.

Disseminating privacy templates contributes to resolving multiple privacy deficiencies. In one dimension, templates reduce the burden on the user to create a configuration of privacy settings "from scratch." Empirically derived templates also have a knowledge distribution aspect that can help users see how other (potentially more privacy savvy) users have chosen to manage their online privacy. This kind of learning from example has the effect of sharing knowledge and experience without requiring substantive technical understanding. The application of templates may have collateral benefits for online service providers as well. Recently, Facebook updated its privacy settings by adding finer-grained controls to content posted, but the "default" settings (template) suggested by Facebook have been criticized as relaxing privacy protections rather than improving them [68][8]. In terms of templates, the effect was that a more public template was imposed alongside finer-grained privacy options—this seems contradictory. An alternative would have been to present users with a menu of context-informed privacy templates that better reflect their usage habits. This strategy would have allowed Facebook to leverage the benefits of legitimately finer-grained privacy settings by folding the more public profile template in with templates that reflect more conservative[9] privacy preferences. This process also returns the discussion to the issue of framing. Regardless of the actual efficacy of

[8]This is also an instance of imposing a particular privacy ideology on a group of users. Mark Zuckerberg, founder of Facebook, has been criticized for claims that this public sharing of information is the new norm [69]. This is in stark contrast with the variety of privacy norms implied by Solove and Nissenbaum. The recommendations here argue that the resolution to these disputes is the empirical construction of templates based on representative subsets of the user population. There are also implications for actors identifying the application of a particular template and making assumptions based on that. The question of what the choice of template says about an individual is an interesting aspect of future research.

[9]In this context, conservative implies a pre-cautionary restriction of information sharing. Zuckerberg's privacy claim may be classified as more liberal.

the templates developed, if the policy framing and presentation of these templates preserves the "service-and-utility" framing characteristic of the policy sample, the benefits will be lost. Identifying meaningful metaphors that emerge from empirical, context-informed templates may contribute to the reframing process described earlier, identifying a more rich vocabulary (and typology) of privacy preferences and implications. In a sense, implementations of the process for constructing templates could be interpreted as collecting usage artifacts from the environment (in the sense of Antòn's grounded process of surfacing requirements in the GBRAM) that preserves a vocabulary familiar to the user.

An immediate criticism of the processes described above is that they are idealistic and require substantive buy-in from the actors involved, in particular online service providers. The collaboration necessary to achieve these objectives will require careful mediation of collaborative forums amongst the relevant actors. The following outlines modes and tools of engagement available to each category of actor, with an eye toward balancing the burdens of implementing privacy protections with the vested interest of the actor at hand. For instance, the data collection necessary for constructing privacy templates may be performed by online service providers and advertisers with requirements to share these templates with other actors, it may be a collaborative effort between civil society organizations that develop browser extensions to collect this information, or it may be a role that can be fulfilled by platform developers. The choice of configuration will depend on negotiations amongst actors, the power relationships between these, vested interests, and how efficaciously mediators and "honest brokers" can help identify imbalances that threaten collateral benefits. Although recent forums have engaged the relevant actors, the interests of these actors are still rather disparate, often debating "basic premises" [31]. The discussions below address existing roles and how a collaborative approach may improve the efficacy and efficiency of regulatory and policy outcomes. This discussion also addresses the sources of information, in particular the capabilities to collect context or usage information, and how the distribution of information affects the distribution of burden in the privacy standard development process.

5.4 Applications

To ground these approaches in familiar problems, these processes will be elaborated in terms of the aggregate image and mixed context problems described earlier. Rather than present a single solution, potential strategic options are presented to highlight combinations of roles, burdens, tools of engagement, and trade-offs that can arise in the policy and tool design space. The typical claim of "onerous" burden is contrasted with scenarios that describe a distribution of burden that requires actors engage in various levels of collaboration to achieve individual and common goals. In many of these cases, the objective is to surface the mutual and collateral benefits of cooperation that may not be obvious when focusing solely on individual, first best solutions. The following framing of applications of this collaboration process is certainly not a complete set of all options, but does represent key interests and trade-offs evidenced

in the current adversarial regulation process. Another key element is the role of a feedback loop that introduces evidence of changing privacy paradigms in the context of emergent technologies and the inevitable deficiencies in the current privacy framework new technologies will highlight. To sustain this feedback loop, the forum must also establish tools for collecting information about the technological environment, a common framing based on this information, and tools that reify the agreed-upon framing. The following illustrates how this process is applied to the aggregate image and mixed context problems.

5.4.1 Aggregate Image Revisited

The fundamental problem with the aggregate image is that, as framed, it is presented as the innocuous collection of tacit data used to customize services. Not only is the framing arguably misleading, but the underlying information asymmetries indicate that lay users may not understand the full privacy implications of tacit data collection. To address this problem, collaborative standards construction processes should develop meaningful policies and attendant tools that inform users of the implications of the aggregate image in a way that preserves the benefits of customization but also highlights the potentially harmful implications. This is in contrast to the one-sided framings such as "service-and-utility" and "privacy/surveillance." One key problem is how to frame the implications of the aggregate image as "personally identifying" or not. Another problem is the ability to access and correct the image to account for distortions. The hurdles are claims of "onerous" burden by the various stewards of aggregate images and the prospect of limitations to innovation.

Depending on the quantity of information, the granularity of detail, and how much of this information can be used in a re-identification process, the aggregate image may or may not be "personally identifying." Adjacent to the fidelity of the aggregate image is the issue of opt-in versus opt-out. Questions regarding how much access to the aggregate image is appropriate and what categories of targeting segments used by advertisers can be made meaningful to users are essential to designing tools that improve users' access to this information. The burden in this case is (1) who should develop a typology of data categories, (2) who should develop the vocabulary for describing these to lay users, (3) who should translate these into goal-based requirements, and (4) who should implement the tools for managing the aggregate image. These steps will require negotiations amongst actors over both the substantive categories and the granularity of the categories. The decisions on categories, in turn, impact the design and implementation of tools for managing categories and the resultant aggregate image.

The information about categories should be based on empirical data regarding what kinds of information users would actually choose to share or keep private. One strategy for collecting the information necessary for operationalization (also applicable in general to mixed context) is to solicit a representative sample of volunteer users to share the tacit data collected about them over a set period of browsing. This is an exercise in designing a social science experiment that collects sufficient information to represent the variance in categorizations and that can capture a representation of

the context in which they occur. This will also require data collection to occur over a sufficient amount of time to be representative of common browsing habits. Taking this as a guideline, the specific experimental design will be a product of the forum. As may be obvious, the collection of data for this sample is intrinsically private and will require explicit opt-in by users. Another factor in the data collection, directly related to burden, is the technical means for collecting this information.

To facilitate constructive conflict, presenting the actors with a strategic menu of options, each representing a potential set of trade-offs, for revising data category standards may facilitate productive negotiations. For example, one strategy for surfacing data categories is to incentivize online service providers to collect this information themselves and share the outcomes with the forum to facilitate development of a continuous data categories standards and revision process. Another strategy that reduces the burden on online service providers is for online service providers to provide an API for accessing (tacit) category information collected during browsing sessions. Platform developers or third party browser extension developers (perhaps supported by civil society organizations) could build tools for users to install in the process of participating in this experiment. Combinations of civil society organizations and online service providers could share the burden of developing analysis tools. Yet another option is the independent development of browser extensions that facilitate annotation and categorization of online service provider practices by civil society organizations. Finally, although not necessarily empirically informed, the backstop option is the possibility of government-backed regulation and development of category standards that may intrinsically deprive interests of the opportunity to contribute as much as the collaborative approach does.

Evaluating this menu of options, the trade-offs are intended to induce constructive compromises based on private negotiations among actors. For example, data collection and distribution to the forum by online service providers and advertisers, although perhaps considered onerous, may be to their advantage because they can ensure their interests in the benefits of customization are presented alongside civil society's interests in privacy implications. The second and third option place the surfacing of categories in the hands of third parties, in particular civil society organizations. As groups focused on privacy as strong information protection, they do not necessarily have any incentive to present a balanced case for the benefits of customization. The third option is a stronger case of framing by privacy-interests if organizations are required to surface data categories on their own, giving rise to outcomes that may not fully represent online service provider and advertiser interests in conveying benefits. The fourth scenario is undesirable for a number of reasons: online service provider and advertisers will argue that it stifles innovation; it is a static, incremental solution that may not weather underlying technological changes; it is not empirically grounded and thus may be just as insufficient as vague categories (and replicate the failures in P3P).

Again, the objective is not to claim that one (or any) of the strategies listed above will solve the problem, but rather to construct a forum for starting from a set of strategies (developed beforehand by participating actors) that will incentivize the kind of bargaining that can identify a more mutually beneficial strategy that preserves the

dynamic, flexible character of self-regulation. For instance, starting from the menu above, some combination of the first and second option may emerge. Such an option may find online service providers and advertisers sharing data categories but also making APIs available to third parties that perform their own analyses, allowing for a form of institutional intercoder validity. Although encouraging two analyses of benefits and implications, one from the perspective of online service providers/advertisers and the other from the perspective of civil society organizations, may conflict, it provides a richer landscape of metaphors that contribute to the reframing process and may give insights to more meaningful metaphors that neither group would have identified on their own. In effect, the collaborative forums are intended to facilitate a form of constructive conflict that will yield compromises that, in contrast to the US Safe Harbor, give rise to a grounded framing of the aggregate image.

Antòn's GBRAM can be applied to the artifacts (documentation) of the common framing to surface system requirements and to identify latent inconsistencies. In some cases, inconsistencies may be technical coordination problems, others may warrant re-evaluation by interested actors. As discussed in the conceptualization of this process, retaining the common vocabulary of the framing accommodates consistency checking and requirements elicitation without creating premature technical barriers. Once requirements are agreed on, allocation of tool development is the next step. Similar to the menu of strategies illustrated above, the development of privacy tools follows the same constructive conflict process of allocating burden based on access to resources and vested interests. These tools would ideally be developed as a framework supporting continuous re-evaluation of privacy framings. Closing the loop, this will help identify new usage patterns related to emerging technologies, attendant deficiencies, and previously unidentified deficiencies.

As an illustrative instance of tool design, one feature would be to capture a representation of the environment in which the tool is deployed. This representation would allow analysts to see the process the individual was engaged in, for example setting the privacy settings on a picture or modifying a particular privacy template. This could be refined into a mechanism for capturing contexts in which privacy tools are used and where users report difficulties. This kind of empirical data collection then allows analysts to better understand contexts based on rich, empirical evidence of usage patterns. Given empirical understandings of contexts, sharing information about these contexts may provide collateral benefits for advertising by helping users best identify what contexts they consider innocuous (and thus fodder for advertising) versus those that are off-limits. This, in turn, can also be incorporated into the design of privacy templates. Thus, the collateral benefits of privacy elicitation processes become a mechanism for eliciting actual preferences regarding privacy *and* advertising, further highlighting the role of designing around shared concerns.

The instrumentation of privacy tools to be context sensitive closes the feedback loop. Like "initial" empirical information regarding data categories and purposes, context information should be made available to all actors and form the basis of the next round of privacy standards discussions. For instance, an emerging advertising domain is the smart phone (device), especially those that provide geo-location services to developers. Although the notion of the aggregate image presented here does not

include implications of terrestrial location, it is not hard to conceive of the aggregate image beginning to incorporate travel patterns. As this technology matures and is deployed more widely, the implications of aggregate image (and mixed context) will spill over from purely online environments into the terrestrial contexts an individual frequents.

5.4.2 Mixed Context Revisited

The process of developing standards for resolving the problems associated with mixed context follows the same iterative feedback loop applied to the aggregate image: collect data, develop a framing that represents the trade-offs amongst common and disparate interests, surface requirements, develop tools, repeat. The following briefly highlights the conflicts and trade-offs amongst interested actors in each step of the feedback process.

Data collection trade-offs are embedded in the articulation of the benefits of providing targeted advertising versus limiting the degree of targeting based on context. For example, one option is, like above, for online service providers and advertisers to collaborate to develop representations of mixed contexts, sharing the these representations with related actors. Another possibility is for some combination of civil society organizations and third party developers to implement browser extensions for use by a representative sample of users. Yet another option that avoids the use of a representative sample is to develop spiders that visit online service providers and record the configurations of mixed context they experience based on searches for popular goods and services. Conflict over the validity, precision, and accuracy of these strategies is to be expected. Again, the backstop might be an option in which government regulators require full opt-in as an "onerous" disincentive to failing to find a useful compromise. As per the earlier discussion, this is the least desired outcome. As evidenced by recent reactions to the FTC's threats of regulation by advertisers, a precisely articulated, higher probability application of this option is expected to incentivize cooperation.

The notion of mixed context frames the problem, but it is not a solution in and of itself. One dimension of the problem is that evidence of mixed context is buried in online service provider privacy policies. A recent solution that places indicators of mixed context directly in the user's workflow is the introduction of an icon associated with advertisements that, when engaged, will inform the user why they have been targeted by that particular advertisement [33]. In addition to targeting information, mixed context information may also be embedded with this icon, providing a meaningful description of how the privacy policy associated with that advertisement differs from the policy of the online service provider. For example, rather than retaining the uniform blue color of the proposed icon, the icon may change colors to yellow or red based on how much that advertisement conflicts with the privacy policy of the online service provider or user's preferences. Coupled with efficacious opt-out tools, the user may examine the sources of the conflict and choose whether to block just that category of advertisements, whether to correct a distortion of their aggregate image, or whether to block that particular advertiser or advertiser network. Following the

theme of instrumenting these tools, data may be collected regarding how frequently opt-outs occur, on what grounds, and what the (inevitable) deficiencies of these tools are. For instance, users that opt out of advertising rather than optimizing their aggregate image may face penalties to incentivize participation necessary to preserve the revenue model while also maintaining an accurate aggregate image. Finding the appropriate balance in such a penalty structure would require empirical analysis of how users react and the impact on online service provider revenues. Ars technica recently performed an experiment where they blocked users that used a certain advertising blocking extension [50]. Formalizing this into an empirical study of the economics of aggregate image accuracy, targeted advertising, and thresholds of utility for users, online service providers, and advertisers is the kind of empirical evidence necessary for efficacious design of privacy and customization tools. An extension for the mobile case may be to choose to opt-out of ads based on location.

This type of strategy places control of how information is used directly in the user's workflow and adds an element of experience that does not require a deep understanding of the technical sources of mixed context. For example, if a user visits an online service provider and all the advertisements have red indicator icons, this will signal the user that the online service provider is not vetting the privacy policies of advertisers and networks effectively. A surfeit of red indicators damages the user's perception of that online service provider. Taking the example a step further, civil society organizations may develop browser extensions that collect metrics on how many conflicts occur as users visit online service providers as a means to give online service providers reputation scores based on whether they frequently accommodate advertisers or networks with poor or conflicting privacy practices. Online service providers may also monitor these values as a "magnitude of preference alignment" that can help them better vet advertisers or remove those whose policies have changed to the detriment of the environment. Taking this a step further still, online service providers may coordinate with civil society to invite re-evaluation of a previously poorly rated page (environment), giving online service providers both a rating in terms of preference alignment *and* responsiveness to ill-behaving advertisers.

Under this scenario, meaningful signals are provided to individuals as a means to make immediate decisions based on information about the privacy implications of a particular environment. In the example of the civil society organization's browser extension, this helps resolve certain aspects of collective action problems and is a strategy that may better incentivize genuine cooperation by online service providers and advertisers. Moreover, this would also have the effect of incentivizing *online service providers* to pressure advertisers and advertising networks to implement better privacy policies or to disallow the collection of personal information in lieu of a targeting protocol that does not give advertisers direct access to user information. In either case, the reputation incentive acts on online service providers, the category of actor that arguably has the closest relationship with advertisers and thus perhaps the most influence.

As discussed earlier, the extent to which these tools are deployed and the substantive form, content, and the allocation of development tasks are products of the compromises and common interests developed in the forum. By employing similar

closure-based mechanisms for collecting and categorizing information about context, future iterations of this standards making process can introduce finer-grained tools for helping users construct representations of context. As with the icon configurations, non-obtrusive incorporation of these tools can facilitate a more experience-based understanding of the online environment and elicit preferences in situ rather than expecting users to express preferences in one visit to a monolithic privacy preferences pane.

5.5 Limitations and Discussion

While this process has potential, it is also limited by the willingness of actors to collaborate. The reactions of the advertising industry to threats of government enforced regulation give some evidence that the backstop strategy described above will work. Although these recommendations focus on fostering constructive conflict amongst categories of actors to resolve the kinds of problems identified in the policy sample, the process also requires more research into the institutional arrangements that will make this constructive conflict process more resilient.

An argument may be made that this is not very different from existing self-regulatory mechanisms. In terms of the feedback loop, it is conceptually similar: identify a harm, collect information, regulate. The difference here is that this process has been institutionalized into a *tighter* feedback loop that concurrently accommodates technological innovation while also encouraging actors to share information so that, as a group, they can proactively identify and reconcile privacy deficiencies early in the design process. In contrast, the current process is dysfunctional in that many compromises are simply empty signals that engage in the bare minimum to resolve conflicts. As noted multiple times, this incrementalist approach may actually obscure solutions to emerging problems. The continuous sharing of information and continuous dialog creates an environment in which innovation can occur, but can be monitored by the actor best suited (or best motivated) to identify privacy deficiencies.

Appendix A

EU Data Protection Directive

Bennett's discussion of the EU Data Protection Directive [19] is one mapping of the Articles describing substantive privacy prescriptions (Chapter 2 of [44]) to Reidenberg's First Principles. Although the secondary analyses provide useful starting descriptions, they were not sufficient for the granularity of analyses applied here. The focus of this work is to highlight the granularity of the FIPs recommendations and the type of enforcement laid out in the black letter of the directive. The following descriptions are a synthesis of a content analysis of the EU Data Protection Directive text [44] and secondary analyses drawn from the literature. The following provides a brief overview of the EU Data Protection Directive's structure, but narrows quickly to a mapping of substantive privacy prescriptions to the FIPS and a description of prescribed enforcement mechanisms. For a fine-grained, point-by-point analysis, see Kuner [75].

A.1 DPD-FIPs

Openness Articles 10, 11, 18, 19, and 21 address issues regarding openness. Article 10 requires the data controller disclose their identity and representatives (Article 10[a]), purposes for processing (10[b]), and information regarding the recipients and categories of recipients of information, whether information queries are voluntary, and the existence of the rights of access and correction (10[c]). This latter element of 10[c] highlights the difference between giving notice of a requirement and fulfilling it. The openness criterion is fundamental because it ensures citizens (users) are aware that information is being collected, with whom it is shared, and what they can do about it.

Article 11 pertains to information indirectly collected about a data subject; in other words, information that is not collected by a controller that has not directly interacted with the data subject. Article 11 requires the same disclosures of controller identity, purposes, recipients, etc. as required for Article 10. The FTC's report on behavioral advertising [51] sheds more light on this distinction.

Article 18 requires data controllers disclose data purposes to the appropriate data authority before carrying out automatic processing. This requirement serves open-

ness because these purposes are subsequently logged in the publicly available register required by Article 21. Article 18 also allows data controllers the option to appoint a personal data protection official that will maintain their own public register of data processing operations. Article 19 provides a more precise set of information necessary when giving notice to the supervisory authority. Notice includes name and address of the controller and the controller's representative, purpose(s) of processing, categories of data and data subjects related to processing, "recipients or categories of recipient to whom the data might be disclosed" [44], transfers to third countries, and a description sufficient for preliminary evaluation of compliance with security requirements.

These openness requirements ensure information is conveyed to the data subject (Articles 10 and 11) and to the supervisory authority (Articles 18, 19, and 21). The public registers required by Article 21 ensure any EU citizen may review the processing practices of a data controller before they choose to engage that controller. Related to review of these practices, these Articles supporting openness do not provide operational guidance regarding how much information should be provided about the categories of data subjects or categories of information collected. In its role as a proactive tool[1] for providing citizens with the information necessary for making privacy decisions, these categories, more precisely, the granularity of these categories, is important for making decisions about whether data purposes are aligned with the citizen's privacy preferences.

The Articles supporting the remaining FIPs are requirements imposed on the data processing operations. From an engineering perspective, these may be considered very high-level requirements imposed by regulation.

Access and Correction Articles 12, 14, and implicitly the accuracy requirements in Article 6[d] support the principle of access and correction. Article 6 addresses data quality issues. In particular, 6[d] requires data be "accurate and, where necessary, kept up to date" [44].

Article 12 articulates the operational elements necessary to support Article 6[d]. In particular, it requires the controller provide access to whether data about a subject is being processed, what categories of data are being processed, and recipients and categories of recipients. This information is the same as required for disclosure by Article 10; henceforth these will be referred to as Article 10 data. Article 12[b] explicitly indicates access and correction measures comply with the "provisions in this Directive" and, as noted above, complies with quality requirements in Article 6. Article 12 also requires the controller notify third parties of updates to changes made to personal data.

Article 14 covers the data subject's right to object to processing on "compelling legitimate grounds" [44], modulo overrides by national legislation or national security interests. Article 14 also gives the data subject the right to object to (and thus halt)

[1]The notion of a proactive tool is akin to the notion of a privacy policy as a decision making tool. It is presumed that this proactive paradigm, especially with regards to openness, is intended to provide the citizen with the information sufficient for making efficacious privacy decisions. The categories described by privacy policies are an artifact of this regulation and are ideally intended to be a source of information in this decision making process.

data purposes related to direct marketing or require the controller notify the subject if data is to be used for direct marketing purposes.

Collection Limitation Notions of collection limitation show up in Articles 6, 7, and 8 with exceptions articulated in Articles 8 and 9 and general exceptions articulated in Article 13. Article 6[a] explicitly indicates data should be "processed fairly and lawfully" [44].

Article 7[a] indicates the data subject must give unambiguous consent. Note that this does not specify how consent is structured; it may be opt-in or opt-out. This is addressed at length in Regan's discussion in [101].

Article 8 indicates certain special (also referred to as "sensitive") categories of data are to be prohibited from processing. These categories include "personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, or the processing of data concerning health or sex life" [44]. Pragmatic derogations of this prohibition are described in Article 8[2], in particular those relating to health data. Article 8 allows these special categories to be processed by explicit consent (opt-in) of the data subject. It is important to note that the option for giving consent (lifting the prohibition) may be expressly disallowed by Member State legislation (Subparagraph (a) of Article 8[2]).

Article 9 makes collection exceptions for journalistic purposes and for purposes of freedom of expression.

Article 13 provides a blanket set of exceptions that apply to Articles 6[1], 10, 11[1], 12, and 21. These exceptions apply to national security, defense, public security, crime investigation and processing, economic and financial interests of the state, operations related to these categories, and for purposes of protecting the rights of others. Henceforth these will be referred to as the Article 13 exceptions.

Use Limitation As indicated in Bennett's articulation of collection limitation, collection, use, and disclosure are often coupled. This is evident in the EU Data Protection Directive. Articulations of use limitation is often interleaved with or adjacent to articulations of collection limitations; disclosure is coupled to these to a lesser degree. Article 6[b,c,e] and Article 7[b-f] prescribe use limitations. Article 8's discussion of special data categories, Article 9's discussion of journalistic and freedom of expression, and Article 13 all apply as exceptions to use limitations.

Article 6 on data quality indicates that data may only be collected for a specific purpose [b] (recall Bennett's "fishing expeditions"), that is must not be excessive [c], and that it may be retained for no longer than necessary (retention limitations) [e]. These give conceptual usage limitations, but do not provide insights into who should decide these limitations or how they should be determined. Kuner's analysis, from the perspective of compliance with EU privacy law, notes that the limitations vary from state-to-state, introducing potentially incompatible and confusing overlaps [75].

Disclosure Limitation As implied above, prescriptions for disclosure limitation are less frequent and often implicit in usage limitations. Disclosure limitation is cov-

ered in Article 6[b], 7[a], 10[c], 11, and 14[b]; each of these applies some form of what Reidenberg referred to as the *finality principle.* Part of Article 6[b]'s requirements of specified, explicit, and legitimate purposes includes limitation regarding who uses data for a particular purpose. Article 7[a] has a similarly implicit disclosure limitation; consent is coupled with information regarding who will have access to information shared by a data subject. The source of these implicit disclosure limitations is Article 10, in particular 10[c], which requires controllers provide information regarding recipients or categories of recipients to the data subject. Similarly, Article 11 requires the same for controllers that do not receive information on a data subject directly from the subject. Finally, Article 14[b]'s limitations on data purposes related to direct marketing ensures controllers cannot use data subjects' information for direct marketing or share it with third parties that intend to use it for direct marketing without consent.

Security The requirements for confidentiality and security are covered by Articles 16 and 17. Article 16 on confidentiality requires controllers requires only employees and agents of the controller directly involved in fulfilling legitimate data purposes and services may access the data subjects' information. Article 17 imposes information security requirements on the controller. Article 17[1] requires "appropriate technical and organizational measures to protect personal data against accidental or unlawful destruction or accidental loss, unauthorized disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing" [44]. This covers what generally falls under the rubric of enterprise information security. The subparagraph of 17[1] provides an important disclaimer: "such measures shall ensure a level of security appropriate to the risks represented by the processing and the nature of the data to be protected." [44]. Paragraphs 17[2] and 17[3] require controllers impose these requirements on third parties that perform operations on data subjects' information on behalf of the controller. This requires a contract or a legally binding act, kept in writing, between the primary controller and the third party.

A.2 Privacy Regulation and Enforcement

Ensuring a uniform enforcement of the espoused privacy principles is a key political element of the EU Data Protection Directive. Long notes that the primary effect of the EU Data Protection Directive is not the substantive privacy prescriptions provided above, but the introduction of powerful national supervisory authorities and their attendant powers to monitor and enforce the EU Data Protection Directive. As a proactive regulatory paradigm, centralized monitoring and enforcement, applied uniformly to public and private data controllers, is one way to ensure FIPs are applied as a comprehensive, socially protective regulatory instrument. The EU Data Protection Directive requires member states enact national laws compatible with and equivalent to the FIPs set out in the EU Data Protection Directive. Supervisory authorities as the monitoring, investigation, and enforcement are described in Article

28; Articles 29 and 30 describe the role of the working group in maintaining the EU Data Protection Directive and coordinating with authorities.

Under Article 28, each member state must have a supervisory authority responsible for monitoring data controllers and enforcing the EU Data Protection Directive. The supervisory authority is expected to be an independent policing organization capable of performing investigations, initiating interventions before processing starts, invoking legal proceedings, and hearing complaints against data controllers. Bennett notes that this substantially extends the powers of pre-existing supervisory authorities [19]. In addition to Article 18 and 19's (discussed with regard to the openness principle) requirements for supervisory authorities to maintain a register of data controllers and attendant data purposes, Article 20 confers powers referred to as "prior checking" [44]. Prior checking requires Member States determine what forms of processing are detrimental and requires the supervisory authorities examine these processes before they are invoked. This depends on Article 18 and 19 requirements that the data controller notify the supervisory authority of such data processing. Although not explicit in the EU Data Protection Directive, it may be assumed that the examination of processes is not immediate; Kuner reports that supervisory authorities are already overwhelmed with other duties [75].

The Data Protection Directive's Article 29 Working Group (created by Articles 29 and 30) is responsible for monitoring and reporting on the efficacy of the EU Data Protection Directive and reporting and advising on issues relevant to the EU Data Protection Directive. The Article 29 Working Group is advisory only; reports are submitted to supervisory authorities, the European Commission, the European Parliament, and the European Council, but are not binding in any way. As per paragraphs (66-67), powers regarding the transfer of data to third countries is conferred on the EC, *modus vivendi*[2]. As will be discussed at more length below, the US Safe Harbor Agreement is an instance of *modus vivendi*: the European Parliament voted against it but the EC used its authority to realize the agreement.

A.3 Transfer and Derogations

Articles 25 and 26 are largely responsible for the externality that the EU Data Protection Directive imposes on non-European Economic Area countries. Article 25 establish that private data may only be transferred to third countries deemed adequate by The Committee. Adequate privacy standards require comprehensive privacy regulation and the attendant enforcement mechanisms[3].

Article 26 establishes derogations from the adequacy requirements intended to accommodate transfers acceptable to the data subject. Adequacy may be assessed

[2]Modus vivendi literally means "way to live" but is more accurately translated to "agree to disagree." This ensures that, despite disagreement between the EC, the Council, and the Parliament, decisions by the EC, once decided, are authoritative and the other bodies will not hamper further progress on other matters in a pursuit of an unrelated disagreement with the EC. This ensures forward progress.

[3]As per Long [77].

two ways: on an individual, case by case, basis or for the entirety of the EU by the Article 31 Committee. On a case by case basis, Article 26 allows individual Member States to authorize data transfers (or sets of transfers) to organizations with states deemed inadequate if the data subject gives consent and it is in the data subject's interest or if the state has determined that, in this case, appropriate safeguards are in place (typically resulting from contractual obligations). Standard contracts exist for this purpose [42, 43]. The Article 31 Committee vets decisions made by the EC. The Article 31 Committee is comprised of representatives from Member States. The Article 31 Committee is not transparent; Bennett refers to it as a "mysterious body" [19] and Kuner indicates that identifying Committee decisions and even the identities of the actual members can prove difficult [75]. A key role of the Article 31 Committee is determining adequacy of non-EEC countries' data laws and the development of model contracts for case-by-case data transfers.

With regard to adequacy criteria, Bennett [19] makes an important note that "these rules and measures must be *complied with*, obviously meaning that symbolic value statements are insufficient." The emphasis on "complied with" further highlights the enforcement role of the national supervisory authorities and the perception that these authorities are a necessary component for upholding privacy. In this vein, self-regulation (viz. the US) may read "symbolic value statements" under some interpretations.

Under Article 25, if a member state recognizes a third country as not having adequate privacy protections, it notifies other member states and the Commission. Further, member states are thus required to prevent data transfer to the inadequate country. Article 25[5] does indicate the Commission should, "at the appropriate time," negotiate some form of agreement that would remedy the adequacy problem and allow data transfers to resume. In this case, that negotiation produced the US Safe Harbor agreement.

Article 26 makes it possible for EU data subjects to waive the right to the EU Data Protection Directive protections and consent to data transfers to countries that have not met the EU Data Protection Directive's adequacy standards. Article 26 takes the form of a list of exceptions for which derogation is valid: among these are unambiguous consent, fulfilling a contract, legal obligation on the part of the data controller, or acting in the vital interest of the data subject. While EU member states may make these arrangements of their own volition, they must do so on a case by case basis with data controllers and must report these authorizations to the Commission for review.

The European Commission (EC) originally found the sectoral, self-regulatory privacy paradigm of the US inadequate. One factor was that the fragmented sectoral privacy regulations did not meet the more comprehensive expectations of the EU adequacy standards. In terms of the enforcement objectives of the EU Data Protection Directive, a self-regulatory scheme is substantially different from the supervisory authority's powers of monitoring and enforcement. Article 25[5] and the desire to preserve essential trans-Atlantic data flows gave rise to the negotiations that ultimately resulted in the US Safe Harbor agreement.

A.4 Role and Efficacy of Data Authorities

According to the letter of the EU Data Protection Directive, the role of the supervisory authorities is to monitor, investigate, and enforce national data protection laws that correspond to the EU Data Protection Directive. Beyond this charge, Kuner reports that violations of the EU Data Protection Directive and attendant penalties do occur, but information is not visible because supervisory authorities do not have the resources to engage in widespread enforcement, the growth of Internet commerce, and that penalties incurred are not publicly available. Kuner provides a short list of enforcement actions across the EU but also indicates that many enforcement actions are settled informally between the controller in question and the DPA. The criteria for enforcement also vary across the EU; some of these operate proactively while others only address "egregious" cases. Penalties available to the supervisory authorities range from orders to take action, fines, and criminal penalties.

The process of bringing claims against controllers also varies across states. In some states, the supervisory authority may bring actions of their own volition, independent of complaints by citizens. The UK DPA is an example of this mode of bringing actions against controllers. Another process requires a data subject complaint first to the data controller and, if no response is received, the complaint may be forwarded to the DPA. An instance of this mode is the Italian DPA. This latter form, in which the process is initiated by a complaint by the data subject, is very similar to the reactive regime espoused by the US (Section 2.3). This latter form also serves as a counterexample to the proactive interpretations of the black letter of the EU Data Protection Directive.

Finally, beyond their black letter role, the DPAs also play a consultative role for data controllers. In many cases, there may be confusion over whether a particular regulation applies and how certain processes may be viewed by the DPAs. Although Kuner reports that routine questions are not asked, the DPAs are considered generally useful for thornier issues, such as transfer of medical records (interpretation of the Article 8[3] exceptions for the transfer of medical data and the related national laws may not be clear). While the DPAs do have substantial experience and can provide very useful advice, they are also reluctant to provide "hard legal advice on an informal basis" [75]

Appendix B

US Safe Harbor FIPs

The US Safe Harbor FIPs (SH-FIPs) are slightly more precise than Bennett's FIPs in Section 2.1, but arguably less so than the DPD-FIPs, as argued by the Article 29 Working Party [15]. Although Aaron's notion of the Safe Harbor was an effective mechanism for finding common ground and moving negotiations forward, the combination of imprecise principles and "off-the-shelf," unmodified self regulatory mechanisms open the door to abuse, satisficing, and ultimately, the categorization of this mechanism as an empty formalism.

The SH-FIPs themselves comprise the principles of notice, choice, onward transfer, security, data integrity, access, and enforcement [124]. Operational clarifications are appended in the form of FAQs [119]. The following maps the SH-FIPs to Bennett's FIPs as a means for common comparison with the DPD-FIPs described in Section A.1.

Openness The Openness principle is implicit in the Safe Harbor requirement for notice. Two elements of the Safe Harbor framework ensure the openness requirement. The first is registration as a participant in the Safe Harbor, described in the Self-Certification FAQ [121]. Self-certification requires submitting contact information, the organization's description, and a description of the organization's privacy policy (including how to access this privacy policy). The list of registered organizations is published on the Department of Commerce web site [123]. Individual organizations' listing include information submitted during certification and whether the registration is current.

The second element of openness is implicit in the SH Principle of Notice. Notice requires an organization inform individuals of the purposes to which their information will be put, how to contact the organization, type of third parties to whom personal information may be disclosed, and choices individuals have regarding use limitation and disclosure. Implicit in notice is the fact that a database of personal information is being maintained by the organization, along with the purposes of this database.

Access and Correction The principle of access and correction is covered by the SH Principles of Access, and Data Integrity. The SH Principle of Access requires individuals be able to access, review, correct, amend, or remove elements of personal information held by an organization as a means to correct inaccuracies. There are

two exceptions to the SH Principle of Access: if the cost of access is disproportionate to the risk to the individual or if access violates the rights of others.

Collection Limitation The Principle of Collection Limitation is not explicitly articulated in the SH Principles. The SH Principle of Notice indicates individuals must be informed of the purposes for which information is collected, which implicitly bounds the types of information to that purpose. Beyond this implicit limitation, the SH Principles focus more on use and disclosure limitations.

Use Limitation The SH Principle of Choice and Data Integrity address use limitation. The SH Principle of Data Integrity ensures that an organization may not use personal data for purposes beyond those for which it was originally collected. Data Integrity also indicates that information should be reliable in the sense that it is accurate, complete and current for its intended use. Choice ensures users may opt out if an organization intends to share their personal information with a third party or if the organization wishes to use the information for purposes other than for those for which it was originally collected. Choice includes an exception: consent is not necessary for disclosure of personal data to third parties working on the behalf (or acting as agents) of the organization. Taken together, this leads to a notify and opt-out structure, where organizations may choose to engage in new data purposes as long as they notify individuals and give a reasonable amount of time for those individuals to opt out.

Disclosure Limitation The Principle of Disclosure Limitation is covered by the SH Principles of Notice, Choice, and Onward Transfer. As per above, the Principles of Notice and Choice are necessary for an organization to share personal information with third parties. Building on Notice and Choice, Onward Transfer focuses more on the liability issues related to disclosure. If the organization ensures that the third party subscribes to the SH Principles, is subject to the EU Data Protection Directive , or is subject to the restriction of a different adequacy finding, the organization is not liable for abuses by the third party (unless the organization makes an agreement otherwise). The organization is liable if it shares information with a third party its knows or should have known would process the information in a manner inconsistent with the restrictions above and did not try to prevent this processing.

Security The Security requirements are quite simple, stated in a single sentence:

> Organizations creating, maintaining, using or disseminating personal information must take reasonable precautions to protect it from loss, misuse and unauthorized access, disclosure, alteration and destruction.

B.1 Self-Regulation and Enforcement

The US Safe Harbor Agreement is considered to fall into the category of self-regulation. The SH Principle of Enforcement describes three requirements for enforcement:

1. independent recourse mechanisms for investigating and resolving disputes and complaints

2. verification procedures for ensuring the espoused privacy practices are true and are implemented as described

3. obligations to correct issues that result from compliance failure along with sanctions and consequences sufficient to ensure compliance

To further clarify the implementation of these, the Dispute Resolution and Enforcement FAQ describes "[h]ow ... the dispute resolution requirements of the Enforcement Principle [should] be implemented, and how an organization's persistent failure to comply with the Principles be handled" [120].

Organizations must first satisfy three additional requirements for points 1 and 3 above. The first is to participate in private sector compliance programs that are compliant with the Safe Harbor Principles and that incorporate appropriate enforcement mechanisms. The second is compliance with the legal or supervisory authorities related to complaints and resolutions. The third requirement is a commitment to cooperate with data protection authorities from either the EU or authorized representatives. The FAQ also encourages the design of other mechanisms for enforcement as long as they meet the requirements set out by the Enforcement principle and related FAQs; this will be revisited in later discussions of more efficacious self-regulatory mechanisms.

To facilitate enforcement, the FAQ indicates that consumers should be encouraged to raise complaints with the organization before pursuing independent recourse. Recourse mechanisms should be clear, easily available and affordable, and should provide individuals the information necessary to effectively navigate the procedure. Following the invocation of recourse mechanisms, if the organization is found to be out of compliance, the FAQ indicates a range of sanctions and remedies. Remedies include reversal and correction of the non-compliant actions by the organization. Sanctions include publicity regarding the violation as a means to affect an organization's reputation, requirements to delete data, removal or suspension of compliance seal if applicable, and compensation.

Authoritative, legal enforcement, in the form of sanction, is through the FTC. Section 5 of the FTC Act[1] prohibits unfair or deceptive acts or practices in commerce. As applied here, if the privacy policies of an organization that self-registers with the Safe Harbor are not upheld, the publication of that policy is considered deceptive and thus actionable under Section 5. This could result in court orders to cease the challenged practices, followed by civil or criminal contempt proceedings if the practices continue. The FAQ also addresses "persistent failure to comply" [120]. In

[1] 15 U.S.C. §45

the case of a repeat offender, Safe Harbor benefits are denied and thus the organization may no longer be able to handle EU citizens' data. In the case an organization fails to comply with dispute resolution mechanisms, it must notify the Department of Commerce or face possible actions under the False Statements Act[2].

B.2 Enforcement History: FTC Actions

In terms of FTC enforcement, a recent report on behavioral advertising [51] indicates that it has brought 23 actions relating to deceptive privacy claims and improper disclosure. A review of the complaints and press releases for these actions indicates that nearly all of these are violations of either Section 5 of the FTC Act or violation of consumer financial information protections under the Graham-Leach-Bliley (GLB) Act. In both of these cases, the organization has been found to have inadequate security measures. In the case of the violations of the FTC Act, these are typically described as misrepresentations of the privacy guarantees made by the organization to consumers. Under the GLB, the case is typically mishandling or improper destruction of consumer information, leading to improper disclosure of this information.

These cases highlight the reactive nature of privacy regulation as enforced by the FTC. These cases are typically brought to the FTC after the face, when an incident endangering user privacy has already occurred. The remedies have been sanctions against the organization, publicity (via the FTC website), and required third party audits of the organizations' security infrastructure for some period of time (typically every two years for 10 to 20 years). This reactive enforcement is an example of sanction-based deterrence intended to encourage organizations to shore up their security practices rather than face the prospect of the burdens imposed by having a third party identify and recommend remedies. As the enforcement mechanism of last resort couched in a largely self-regulatory scheme, the intended effect is to encourage genuine compliance with the FIPs in lieu the imposition of audits that bring with them the onerous burdens organizations are trying to avoid via self-regulation.

An important aspect of these enforcement actions is that they focus almost exclusively on information security practices, not on any of the other behaviors proscribed by the general FIPs, the SH-FIPs, or the DPD-FIPs, such as collection and use limitations.

[2] 18 U.S.C. §1001

Appendix C

Nominal Compliance with US Safe Harbor

In this document, nominal compliance indicates that a given policy complies with the black letter of the US Safe Harbor FIPs. As noted in Section 2.3.2, the US Safe Harbor FIPs comprise the principles of Notice, Choice, Onward Transfer, Security, Access, and Enforcement. In addition to these principles, the US Safe Harbor Agreement also addresses sensitive data, self-certification, verification, and criteria regarding the timing of opt out choices in the US Safe Harbor FAQs. The following describes the compliance of the privacy sample with regards to the six principles and the criteria from the facts, making note of substantive deviations in each category.

Of the six US Safe Harbor principles, the content related to the Principle of Notice receives the most attention and has the most precise information. Although the most precise in comparison to the other principles and criteria, it is still in terms of broad categories of data purposes and categories. The remaining principles and criteria are addressed in varying depth.

C.1 Principle of Notice

The Principle of Notice can be decomposed into *data purposes*, *contact information*, *disclosure to third parties*, *opt out choices*, and *categories of data collected.*

Contact information is the most simple; each policy provides an e-mail address for reporting privacy violations, a corporate contact, the contact information for the department handling privacy issues, and contact information for the dispute resolution provider, if they are participating in a third party program. The remaining categories, data purposes, disclosure to third parties, opt out choices, and categories of data collected are often addressed based on either the category of information collected or the method by which data is collected.

The broad data purposes given by a privacy policy are to provide and improve service, provide the user with a customized experience, and provide relevant advertisements. Under providing service, privacy policies indicate with whom they share information. All privacy policies address sharing information with third parties that

provide operational services, such as contractors. The policies indicate that only the information necessary for performing the contractor or third party's function is shared.

The privacy policies also give notice that clickstream data is collected. As described in the discussion of automatic data collection, the description of clickstream data largely addresses the mechanisms used to collect this data and the most general uses, such as improving efficiency and or customization. Online service providers do not volunteer more precise information regarding the categorization of clickstream data, the inferences made from clickstream data, or the implications of how clickstream data is combined with other data.

An option to opt out of the use of clickstream data is provided by most online service providers. With the exception of Google, the opt out is all or nothing. Google provides a slightly more refined opt out feature, allowing users to opt out completely or select what categories of advertisement they wish to see. If a user opts out, clickstream data is not used to perform any advertisement customization but may still be collected. Some privacy policies provide direct links to opt out of clickstream data collection either through their own site or via the NAI [87]; others simply describe how to opt out with reference to the online service provider's account or profile management tools.

In addition to sharing data with contractors for operations related purposes, the privacy policies also indicate that aggregate data collected about users is shared with other third parties, in particular advertisers. The privacy policies specifically reassure users that PII is not shared. Online service providers indicate that the information shared cannot be used to identify users individually and is often aggregate data, such as how many users viewed a particular advertisement or the number of users browsing a particular topic. These assertions appeal to the fact that, individually, these categories of information are not personally identifiable but do not reassure the user whether the organizations with whom their information is shared can reconstruct their identity based on a larger collection of data. This is particularly the case with regards to advertisers whose data collection spans multiple domains via the use of web beacons.

Online service providers also provide notice that some of the content shown on an online service provider's pages may be provided by and hosted by third party advertisers. Online service providers make it clear that the assertions and guarantees made within a privacy policy bind only to the content provided by the online service provider, not the content and actions of advertisers even though the online service provider allows them to display adds on their site. This disclaimers and qualifications usually come with the reassurance that advertisers are not allowed to collected conventional PII, but may be able to read and write cookies, know what page a user is viewing, and know what demographic and marketing segment the ad was targeted to. The online service provider's privacy policy suggests the users review advertisers' privacy policies, as they may differ from the assertion made in the online service provider's policy. The effect is that the environment presented by a page primarily hosted by an online service provider may contain a dynamic mix of privacy policies comprised of the privacy policy of the online service provider and any subset of the

advertisers that may be serving ads for that page.

Privacy policies also provide information about how long information is retained. Some privacy policies provide explicit information regarding how long information is retained, while others provide vague bounds based on need, legal requirements, and other "reasonable" retention periods. Privacy policies also differentiate between the types of information retained. PII is typically retained for as long as an individual maintains an account with the online service provider. In conjunction with notices regarding access (see below), privacy policies also indicate that older versions of PII may be retained under the auspices of customer service (e.g. undoing changes), following up on fraud and abuse, and for legal purposes. In a similar vein, privacy policies also indicate that information may be retained after an account is closed for similar purposes. Few online service providers indicate how long voluntary information is retained, again referring to "reasonable" retention periods.

Online service provider's also give users notice that non-PII[1] will be combined with categories of information from the same page, the user's session, across user sessions, across services provided by within the domain of the provider, from information gathered by affiliate organizations, or from other public sources or sources the user has shared information with and provided consent to distribute. Each online service provider does not make all of these categories explicit to the user, but these categories can be derived from a careful reading of the policy. The notice of combination is sometimes accompanied by a recommendation to read the privacy policies of other services within the same domain.

C.2 Principle of Choice

Online service provider's privacy policies generally rely on opt out mechanisms to comply with the principle of choice. In most cases, the tone of the policy is that one should share their voluntary information and in exchange one will have access to the value service provided, accompanied by reassurances that their PII is not shared (see discussion in the previous section) and that it is protected, secured, and accessible. In terms of tone (but not substantial information regarding purpose), LinkedIn deviated from this trend. LinkedIn's privacy policy had a "buyer beware" tone that permeated the whole of the privacy policy.

> You willingly provide us certain personal information, which we collect in order to allow you to benefit from the LinkedIn website. If you have any hesitation about providing such information to us and/or having such information displayed on the LinkedIn website or otherwise used in any manner permitted in this Privacy Policy and the User Agreement, you should not become a member of the LinkedIn community. [76]

In contrast to the inviting tone of other privacy policies, LinkedIn explicitly indicates that if a user is in any way uncomfortable with the kinds of information shared within

[1]See Section 4.2.3.1 for the distinction between conventional PII versus non-PII and a discussion of the purposes of online service provider's information combination practices.

the LinkedIn community, they should not become a member.

LinkedIn's blunt expression of expectations and user choice is a refreshingly frank in contrast to tone and language of other online service provider privacy policies that seem to be attempting to elicit a feeling of trust, security, and benefit. The blunt tone of the LinkedIn privacy policy reflects its roots in professional and business networking and belies the attitude projected by the site as a whole: this is a place of business. Despite this distinct tone, the LinkedIn privacy policy, like the other privacy policies in the sample, provide little more than marginal choice regarding how user information, in particular automatic information, is used.

As noted in the discussion of Notice, users are often directed to the opt out feature of the online service provider or the NAI [87], for nearly all of the services provided (except Google), users are given an all-or-nothing choice regarding data collection. The opt out choice is tacitly provided as an alternative to disabling cookies completely or, for the more sophisticated user, disabling cookies for a particular domain. Rather, almost all the opt out mechanisms disable behavioral advertising[2]. The user is generally admonished from doing this, indicating that opting out of the use of cookies and other tracking mechanisms reduce the utility of the site because advertisements are not customized to their habits and preferences. In effect, online service providers tacitly encourage the user to reconsider the tension between the utility of advertisement customization and their privacy preferences[3]. In this case, only a marginal choice is given regarding a subset of the purposes to which automatic data are put. The marginal choice is all-or-nothing: the user may garner the utility of customized advertising by acquiescing to whatever combination automatic information (modulo sensitive data, see Section C.7.1) the online service provider and/or advertisers see fit or garner none of the utility and have access to generic advertising placement. This highlights both the precision of the choices available regarding behavioral advertising and that the user is not given any choice regarding other uses of automatic data.

The alternative to the online service provider supplied opt out mechanisms is to disable all cookies. Disabling cookies is generally admonished by the online service provider privacy policy because it may interfere with the innocent functionality of the online service provider as a web application, depriving users the full benefits of the service. Although disabling cookies will limit the aggregation of data from multiple temporally distributed sessions, IP information is unlikely to change over the course of the average session[4].

[2]Different policies refer to what is essentially behavioral advertising by different names. Google refers to this application of automatic data as interest based advertising [56]. Amazon refers to it as relevant advertising [5]. Nearly all of the online service providers appeal to the benefits and utility of customized advertising in contrast to the nuisance of semi-random advertisements.

[3]This also has implications for understanding whether users' espoused privacy preferences match the the actual preferences expressed by their behaviors when faced with these types of situations.

[4]This is true for desktop machines, still generally so for laptops even though the frequency of mid-session changes to IP is logically higher, but IP address information is probably much less consistent per session for mobile users.

C.3 Principle of Onward Transfer

With regard to privacy policies, online service providers are clear to assure users that they do not share, sell, or rent conventional PII with third parties without the user's consent. The definition of onward transfer from the US Safe Harbor Principles (Section 2.3.2) is defined first as depending on the principles of Notice and Choice to ensure individuals know with whom what categories of information are shared and are given a choice regarding this sharing. The discussions of these dimensions of the privacy policies in Sections C.1 and C.2 cover much of the sample's compliance with onward transfer.

The discussion of Notice in Section C.1 indicates that information may be transferred to agents of the online service provider that provide operations support, affiliated companies in joint ventures, advertisers, or other third parties with whom the online service provider does business. With the exception of some online service providers advertising practices, online service provider's privacy policies do not provide any additional precision to the description of what agents comprise these categories. In the case of advertisers, some privacy policies provide a list of advertisers they allow to serve adds to users, but the onus remains on the user to follow the links, find the advertiser's policy, and reconcile this policy with both the hosting online service provider's policy and the user's own privacy preferences.

C.4 Principle of Security

The principle of security is quite possibly the least precise element of the privacy policies in the sample. Overall, compliance with the security principle gives the impression that it is included purely for purposes of compliance and to avoid liability issues. Generally, security compliance general references "reasonable" application of standard industry grade security practices (as a category). The Security Principle and FAQs criteria regarding access (see below) provide online service providers with an exception: security and access can both be limited in proportion to the probably harm to the user if their request is not fulfilled. This clause is repeated in each of the privacy policies in the sample as limiting the scope, frequency, and type of security and access requests.

Like the other descriptions in the privacy policies, these assertions address fairly imprecise categories of information and processes and do not give detailed guidance on what would be considered onerous to the online service provider. In some cases, examples of onerous requests are provided, such as requesting changes in data stored in archive, removing the contribution of individual data points from aggregate data, or situations where the security or change requested would infringe on the rights of others. In other cases, the categories of protections are only slightly more precise, making general references to the application of industry standard encryption, digital and physical audits of databases, and other business processes. As is the trend with each of the Principles, the privacy policies assert practices that fall into broad categories, but do not provide sufficient information to differentiate them in any significant

way.

C.5 Principle of Access

Compliance with the Principle of Access has been described in part in the discussion of Notice. In the process of describing categories related to scope, type, and temporal limitations on the information collected, online service providers also tell the users that many of these preferences may be modified using the account settings features of the online service provider. In the case of PII, these features usually refer to the forms that allow users to update contact and billing information. In some cases, there are direct links to the account settings features, in other cases simple in text references, leaving it to the user to navigate to these features on their own. References to opt out features are similar: privacy polices references these as they come to the portion of the privacy policy that gives notice regarding automatic information collection as it relates to advertising or behavioral tracking. Like the account settings, some policies provide direct links, others simply describe the feature and expect the user to seek it out.

The lack of information regarding precisely how PII and automatic data are combined (discussed with the Principle of Notice, Section C.1) is also a problem for access to this information. Opt out mechanisms are available to disable behavioral advertising, but this does not indicate whether these opt out mechanisms actually stop the actual collection of the data for other purposes. The description of the Access Principle, access implies the ability to review, correct, and delete personal information collected about an individual. Of the online service providers in the sample, Amazon was the only one that provided instance-by-instance access to a subset of automatic information

Amazon's browsing history provides the user with customized history based on behavioral data. The browsing history account settings are an instance of limited access to automatic information that is used for customization. Upon entering Amazon's site, the user is presented with a customized list of recommended items and the general categories from which these items are drawn in a simple tag cloud. The browsing history allows the user to remove *recent* items viewed and searches, but the ability to modify the categories displayed in the tag cloud is not available.

In many ways, this resembles the onus of checking the privacy policy for changes: the user is responsible for keeping track of searches, within their workflow, and removing these searches as they happen. If the user decides to review the browsing history, it may only contain the most recent searches, not the full history of searches. The history of searches may remain, but the only visible artifact is a corresponding category in the recommendations categories tag cloud. Further, it is not clear how a user would tweak the categories in the tag cloud without using the 'Delete All Items' feature of the browsing history management form. Combining the onus of continually checking the recent browsing history and the loss of all useful customization when relegated to 'Delete All Items' option still preserves the all-or-nothing character of users' exchange of privacy options for the utility of customization. It is also important to

note that although this is the most precise tool for managing automatic information within the sample, it is, as implied above, still rather coarse grained. Further, this tool prioritizes the online service provider's control of the uses of automatic information over configurations that allow the user to better express their privacy preferences, potentially to the detriment of a more complete purview of users' behaviors.

C.6 Principle of Enforcement

Compliance with the Principle of Enforcement competes with the Principle of Security for least coverage in privacy policies. Most policies include reassurances that the privacy policies and practices are reviewed on a regular basis. For those that participate in a third-part compliance program, these assurances are accompanied with a reference to that compliance program. In general, those that self-regulate provide an e-mail address and physical mailing address to which one may address complaints. In the case of those that participate in third party programs, the privacy policy indicates that if the user should first contact the online service provider via the provided contact information and if the resolution is not satisfactory, the user should then contact the third-party compliance program.

C.7 Compliance with US Safe Harbor Criteria

In addition to the US Safe Harbor Principles, the US Safe Harbor agreement also provides additional criteria in the FAQ [122].

C.7.1 Sensitive Data

The sensitive data criteria establish that certain sensitive data (religious and ethnic information, sexual preferences, etc.) should not be recorded without the explicit consent of the user. Only a few sites describe sensitive data precisely what sensitive data is; LinkedIn provides the most useful discussion of sensitive information. In its discussion of consent to processing information about the user, it indicates that information provided in the user generated profile may allow others with access to this information to infer other information such as nationality, ethnic origin, or religion[76].

The primary criteria regarding the sensitive data FAQ is a list of exceptions. These exceptions are generally covered when the privacy policy gives notice regarding what information is collected. The notice regarding legal exceptions and vital interests generally covers both sensitive information uses and sharing of information in the event of a fraud investigation or due to legal requirements.

C.7.2 Self-Certification

The self-certification criteria requires that those online service providers that do self-certify also indicate that they adhere to the US Safe Harbor Principles. This is

generally a single statement either at the beginning of the document when introducing the privacy policy and relevant contact information or included with the enforcement notice near the end of the privacy policy.

C.7.3 Timing of Opt Out

The Timing of Opt Out pertains to ensuring users have reasonable access to opt out mechanisms. With respect to the published privacy policies, this is tied to the Principle of Choice and giving notice of either central (all-or-nothing) opt out mechanisms, links to the opt-out mechanisms, or indicating that any non-administrative e-mail the user may receive will contain instructions or a link to opt out of any further unwanted communications. The objective is that the user has immediate access to opt out mechanisms and that they are obvious and available.

C.7.4 Privacy Rights of Minors

In addition to the privacy requirements imposed by the US Safe Harbor Agreement, privacy policies also contain clauses that indicate compliance with COPPA (Children's Online Privacy Protection Act)[5]. Each of the privacy policies includes clauses indicating age restrictions on users. Although some disallow users under 18, all indicate that users 13 are either strictly prohibited or their account must be created by a parent using the online service provider's parental controls.

[5]Available in [105].

Bibliography

[1] A. Acquisti. Protecting privacy with economics: Economic incentives for preventive technologies in ubiquitous computing environments. Citeseer.

[2] A. Acquisti. Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th ACM conference on Electronic commerce*, page 29, 2004.

[3] A. Acquisti and J. Grossklags. Privacy and Rationality in Individual Decision Making. *IEEE Security & Privacy*, 3(1):26–33, 2005.

[4] A. Acquisti and J. Grossklags. Privacy and rationality in individual decision making. *IEEE Security & Privacy*, pages 26–33, 2005.

[5] Amazon.com. Amazon.com Privacy Notice. Retrieved from `http://www.amazon.com/privacy`, June 2009.

[6] A.I. Anton. Goal-Based Requirements Analysis. In *Proceedings of the 2nd IEEE International Conference on Requirements Engineering*, pages 136–144, Colorado, April 1996. IEEE.

[7] A.I. Anton, E. Bertino, N. Li, and T. Yu. A roadmap for comprehensive online privacy policy management. 2007.

[8] A.I. Anton, J.B. Earp, Q. He, W. Stufflebeam, D. Bolchini, and C. Jensen. Financial Privacy Policies and the Need for Standardization. *IEEE Security & Privacy*, pages 36–45, 2004.

[9] A.I. Antón, J.B. Earp, C. Potts, and T.A. Alspaugh. The Role of Policy and Stakeholder Privacy Values in Requirements Engineering. In *Proceedings of the 2001 International Symposium on Requirements Engineering*, pages 138–145, August 2001.

[10] A.I. Anton, JB Earp, and A. Reese. Analyzing Website Privacy Requirements Using a Privacy Goal Taxonomy. In *Proceedings of the 2002 IEEE Joint International Conference on Requirements Engineering*, pages 23–31, 2002.

[11] A.I. Anton, Q. He, and D. Baumer. The Complexity Underlying JetBlue's Privacy Policy Violations. *IEEE Security & Privacy*, 2(6):12–18, 2004.

[12] A.I. Antón, W.M. McCracken, and C. Potts. Goal decomposition and scenario analysis in business process reengineering. In *Proceedings of the 6th international conference on Advanced information systems engineering table of contents*, pages 94–100. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 1994.

[13] AI Anton and C. Potts. The use of goals to surface requirements for evolving systems. In *Software Engineering, 1998. Proceedings of the 1998 International Conference on*, pages 157–166, 1998.

[14] Annie I. Antón. *Goal Identification and Refinement in Specification of Software-Based Information Systems.* PhD thesis, Georgia Institute of Technology, Atlanta, GA, June 1997.

[15] Article 29 Working Party. Opinion 7/99 on the Level of Data Protection provided by the "Safe Harbor" Principles as published together with the Frequently Asked Questions (FAQs) and other related documents on 15 and 16 November 1999 by the US Department of Commerce.

[16] S. Bellman, E.J. Johnson, S.J. Kobrin, and G.L. Lohse. International differences in information privacy concerns: a global survey of consumers. *The Information Society*, 20(5):313–324, 2004.

[17] C.J. Bennett. *Regulating Privacy: Data Protection and Public Policy in Europe and the United States.* Cornell University Press, 1992.

[18] C.J. Bennett and C.D. Raab. *The Governance of Privacy: Policy Instruments in Global Perspective.* Ashgate Publishing, 2006.

[19] Colin J. Bennett and Charles D. Raab. The adequacy of privacy: The european union data protection directive and the north american response. *The Information Society*, 13(3):245–264, Jul-Sep 1997.

[20] Bettina Berendt, Oliver Günther, and Sarah Spiekermann. Privacy in e-commerce: stated preferences vs. actual behavior. *Commun. ACM*, 48(4):101–106, 2005.

[21] Fernando Bermejo. Audience manufacture in historical perspective: from broadcasting to Google. *New Media Society*, 11(1-2):133–154, February 2009.

[22] Lisa Bernstein. Private Commercial Law in the Cotton Industry: Creating Cooperation Through Rules, Norms, and Institutions. Technical Report 133, University of Chicago, 2001.

[23] R. Bessette and V. Haufler. Against All Odds: Why There Is No International Information Regime. *International Studies Perspectives*, 2(1):69–92, 2001.

[24] L.D. Brandeis. *Other Peope's Money, and How the Bankers Use It.* BiblioLife, 2009.

[25] M.D.J. Buss. Legislative threat to transborder data flow. *Harvard Business Review*, 62(3):111–118, May/June 1984.

[26] Doreen Carvajal. EBay Ordered to Pay $61 Million in Sale of Counterfeit Goods. `http://www.nytimes.com/2008/07/01/technology/01ebay.html`.

[27] Fred H. Cate. *Privacy in the Information Age.* Brookings Institution Press, 1997.

[28] David Chartier. ebay to end all louis vuitton auctions in fake purse dustup. http://arstechnica.com/news.ars/post/20080630-luxury-manufacturer-ebay-cant-sell-our-stuff.html.

[29] R.K. Chellappa and R.G. Sin. Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information Technology and Management*, 6(2):181–202, 2005.

[30] R. Clarke. Internet privacy concerns confirm the case for intervention. *Communications of the ACM*, 42(2):60–67, 1999.

[31] Stephanie Clifford. Groups Far Apart on Online Privacy Oversight. *The New York Times*, December 2009. Retrieved fom `http://www.nytimes.com/2009/12/08/business/media/08adco.html` on 7 February 2010. Published 8 December 2009.

[32] Stephanie Clifford. Industry Tightens Its Standards for Tracking Web Surfers. *The New York Times*, July 2009. Retrieved fom `http://www.nytimes.com/2009/07/02/business/media/02adco.html` on 5 February 2010. Published 1 July 2009.

[33] Stephanie Clifford. A Little 'i' to Teach About Online Privacy. *The New York Times*, 2010. Retrieved fom `http://www.nytimes.com/2010/01/27/business/media/27adco.html` on 5 February 2010. Published 26 January 2010.

[34] Stephanie Clifford. F.T.C.: Has Internet Gone Beyond Privacy Policies? *NYTimes Media Decoder*, 2010. Retrieved fom `http://mediadecoder.blogs.nytimes.com/2010/01/11/ftc-has-internet-gone-beyond-privacy-policies/` on 5 February 2010. Published 11 January 2010.

[35] Council of Europe. Convention for the Protection of Human Rights and Fundamental Feedoms (1950). In Marc Rotenberg, editor, *The Privacy Law Sourcebook 2000: United States Law, International Law, and Recent Developments.* Electronic Privacy Information Center, 2000.

[36] Council of Europe. Convention for the Protection of Individuals with regard to the Automatic Processing of Personal Data. In Marc Rotenberg, editor, *The Privacy Law Sourcebook 2000: United States Law, International Law, and Recent Developments.* Electronic Privacy Information Center, 2000.

[37] Council of Europe, Consultative Assembly. *European Yearbook*, chapter Recommendation 509 (1968) on Human Rights and Modern Scientific and Technological Developments. Martinus Nijhoff Publishers, 1971.

[38] Luc Delany. Safer Social Networking and Self Regulation. *European Public Policy Blog*, 2009. Retrieved fom `http://googlepolicyeurope.blogspot.com/2009/02/safer-social-networking-and-self.html`.

[39] John Dewey. *Logic - The Theory of Inquiry.* Henry Holt and Company, November 1938.

[40] eBay.com. Privacy Central. Retrieved from `http://pages.ebay.com/securitycenter/privacy_central.html`, June 2009.

[41] EPIC. Pretty Poor Privacy: An Assessment of P3P and Internet Privacy, 2000. Retrieved fom `http://epic.org/reports/prettypoorprivacy.html`.

[42] European Commission. Commission decision of 15 june 2001 on standard contractual clauses for the transfer of personal data to third countries, under directive 95/46/ec. *Official Journal of the European Communities*, 44(L 181):19–31, 2001.

[43] European Commission. Commission decision of 27 december 2001 on standard contractual clauses for the transfer of personal data to processors established in third countries, under directive 95/46/ec. *Official Journal of the European Communities*, 45(L 6):52–62, 2002.

[44] European Parliament, Council. Directive 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Communities*, (L 281):31–50, October 1995.

[45] European Parliament, Council. Charter of Fundamental Rights of the European Union. *Official Journal of the European Communities*, (C 364):1–22, 2000.

[46] Facebook.com. Facebook's Privacy Policy. Retrieved from `http://www.facebook.com/policy.php`, June 2009.

[47] H. Farrell. Negotiating Privacy Across Arenas: The EU-US "Safe Harbor" Discussions. In A. Héritier, editor, *Common Goods. Reinventing European and International Governance*, pages 105–126. Rowman & Littlefield Publishers, Lanham, MD, 2002.

[48] H. Farrell. Constructing the international foundations of e-commerce—the eu-us safe harbor arrangement. *International Organization*, 57(02):277–306, 2003.

[49] Henry Farrell. Privacy in the Digital Age: States, Private Actors, and Hybrid Arrangements. In William J. Drake and Ernest J. Wilson, editors, *Governing Global Electronic Networks: International Perspectives on Policy and Power*, The Information Revolution and Global Politics, pages 375–400. The MIT Press, Cambridge, MA, 2008.

[50] Ken Fisher. Why Ad Blocking is devastating to the sites you love. *Arstechnica*, 2010. Retrieved fom `http://arstechnica.com/business/news/2010/03/why-ad-blocking-is-devastating-to-the-sites-you-love.ars` on 20 March 2010. Published 4 March 2010.

[51] FTC. FTC Staff Revises Online Behavioral Advertising Principles.

[52] FTC. In the Matter of Sears Holdings Management Corporation, a corporation. FTC File No. 082 3099.

[53] FTC. Online Behavioral Advertising: Moving the Discussion Forward to Possible Self-Regulatory Principles. # 228; Project No. P859900.

[54] FTC Staff. FTC Staff Report: Self-Regulatory Principles For Online Behavioral Advertising.

[55] Monica Gaza. Hermes wins lawsuit against ebay. http://news.softpedia.com/news/Herm-s-Wins-Lawsuit-Against-eBay-87388.shtml.

[56] Google.com. Google Privacy Center. Retrieved from `http://www.google.com/privacy`, June 2009.

[57] Google.com. Privacy FAQ. Retrieved from `http://www.google.com/intl/en/privacy_faq.html#toc-terms-personal-info`, June 2009.

[58] Andy Greenberg. Congress Mulls Online Privacy Law. *Forbes.com*, 2009. Retrieved fom `http://www.forbes.com/2009/06/18/google-yahoo-privacy-technology-advertising.html`.

[59] Andy Greenberg. Struggling Online Admen May Face Privacy Squeeze. *Forbes.com*, 2009. Retrieved fom `http://www.forbes.com/2009/06/17/online-advertising-privacy-technology-security-congress.html`.

[60] P.M.S. Hacker. *Insight and illusion: Themes in the philosophy of Wittgenstein.* Oxford University Press, USA, 1986.

[61] Katie Hafner. Tiffany and ebay in fight over fakes. Retrieved from `http://www.nytimes.com/2007/11/27/technology/27ebay.html`.

[62] I.H. Hann, K.L. Hui, T.S. Lee, and I.P.L. Png. Online information privacy: Measuring the cost-benefit trade-off. In *23rd International Conference on Information Systems*, 2002.

[63] H.L.A. Hart. *The Concept of Law.* Clarendon Law Series. Oxford University Press, 2nd edition, 1994.

[64] D. Heisenberg. *Negotiating Privacy: The European Union, the United States, and Personal Data Protection.* Lynne Rienner Publishers, Boulder, CO, 2005.

[65] Roger Hull. Lobby Brussels: A View from Within. In Mazey Sonia and Jeremy Richardson, editors, *Lobbying in the European Community.* Oxford University Press, 1993.

[66] S.S. Jasanoff. Contested boundaries in policy-relevant science. *Social Studies of Science*, 17(2):195, 1987.

[67] C. Jensen, C. Potts, and C. Jensen. Privacy practices of internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies*, 63(1-2):203–227, 2005.

[68] Cecilia Kang. Facebook adopts new privacy settings to give users more control over content. *The Washington Post*, 2009. Retrieved from `http://www.washingtonpost.com/wp-dyn/content/article/2009/12/09/AR2009120904200.html?wprss=rss_technology`. Published: December 10, 2009.

[69] Marshall Kirkpatrick. Why Facebook is Wrong: Privacy Is Still Important. *ReadWriteWeb*, 2010. Retrieved fom `http://www.readwriteweb.com/archives/why_facebook_is_wrong_about_privacy.php`.

[70] S. J. Kobrin. Territoriality and the governance of cyberspace. *Journal of International Business Studies*, 32(4):18, 2001.

[71] S.J. Kobrin. Safe harbours are hard to find: the trans-atlantic data privacy dispute, territorial jurisdiction and global governance. *Review of International Studies*, 30(01):111–131, 2003.

[72] A. Kobsa. Personalized hypermedia and international privacy. *Communications of the ACM*, 45(5):64–67, 2002.

[73] Uta Kohl. *Jurisdiction and the Internet.* Cambridge University Press, 2007.

[74] Barbara Koremenos, Charles Lipson, and Duncan Snidal. The Rational Design of International Institutions. *International Organization*, 55(04):761–799, 2001.

[75] Christopher Kuner. *European Data Privacy Law and Online Business.* Oxford University Press, Oxford, UK, 2003.

[76] LinkedIn.com. Privacy Policy. Retrieved from `http://www.linkedin.com/static?key=privacy_policy&trk=hb_ft_priv`, June 2009.

[77] W. J. Long and M. P. Quek. Personal data privacy protection in an age of globalization: the us-eu safe harbor compromise. *Journal of European Public Policy*, 9(3):325–344, 2002.

[78] Jonathan R. Macey. Public and Private Ordering and the Production of Legitimate and Illegitimate Legal Rules. *Cornell Law Review*, 82, July 1997.

[79] I.C. Magaziner. The framework for global electronic commerce: A policy perspective. *Journal of International Affairs*, 51(2):527–528, 1998.

[80] Ira Magaziner. Transcript of a Speech given at IBM Colloquium, "Privacy in a Networked World: Demands, Technologies, and Opportunities". http://researchweb.watson.ibm.com/iac/transcripts/internet-privacy-symp/iramagaziner.html.

[81] Amber Maitland. Ebay slapped with lawsuit by louis vuitton and dior couture. http://www.pocket-lint.co.uk/news/news.phtml/4844/5868/ebay-vuitton-dior-lawsuit-counterfeit.phtml.

[82] Carol Matlack. Hermès Beats eBay in Counterfeit Case. `http://www.businessweek.com/globalbiz/content/jun2008/gb2008066_845380.htm`.

[83] Carol Matlack. LVMH vs. eBay: A Counterfeit Suit. `http://www.businessweek.com/print/globalbiz/content/sep2006/gb20060922\_888836.htm`.

[84] Microsoft.com. Microsoft Online Privacy Statement. Retrieved from `http://privacy.microsoft.com/en-us/fullnotice.mspx`, June 2009.

[85] Microsoft.com. Personalized Advertising from Microsoft. Retrieved from `http://choice.live.com/advertisementchoice/`, June 2009.

[86] MySpace.com. Privacy Policy. Retrieved from `http://www.myspace.com/index.cfm?fuseaction=misc.privacy`, June 2009.

[87] NAI. Network Advertising Initiative. See `http://www.networkadvertising.org/`.

[88] Phillip Nelson. Information and Consumer Behavior. *The Journal of Political Economy*, 78(2):311–329, April 1970. ArticleType: primary_article / Full publication date: Mar. - Apr., 1970 / Copyright © 1970 The University of Chicago Press.

[89] H. Nissenbaum. Privacy as contextual integrity. *Washington Law Review*, 79(1), 2004.

[90] Mancur Olson. *The Rise and Decline of Nations.* Yale University Pres, New Haven, CT, 1982.

[91] Organization for Economic Co-operation and Development. Guidelines Governing Protection of Privacy and Transborder Data Flows of Personal Data. In Marc Rotenberg, editor, *The Privacy Law Sourcebook 2000: United States Law, International Law, and Recent Developments.* Electronic Privacy Information Center, 2000.

[92] Organization for Economic Co-operation and Development. OECD Privacy Guidelines (1980). In Marc Rotenberg, editor, *The Privacy Law Sourcebook 2000: United States Law, International Law, and Recent Developments.* Electronic Privacy Information Center, 2000.

[93] Overstock.com. Privacy and Security. Retrieved from `https://help.overstock.com/cgi-bin/overstock.cfg/php/enduser/std_adp.php?p_faqid=65&`, June 2009.

[94] Wladimir Palant. Adblock Plus. See `https://addons.mozilla.org/en-US/firefox/addon/1865/`.

[95] Roger Parloff. eBay Denied Stay in LVMH Case. `http://legalpad.blogs.fortune.cnn.com/2008/07/11/ebay-denied-stay-in-lvmh-case/`.

[96] Roger Parloff. eBay Scrambles to Reverse Loss in LVMH Case. `http://legalpad.blogs.fortune.cnn.com/2008/07/09/ebay-scrambles-to-reverse-loss-in-lvmh-case/`.

[97] Roger Parloff. eBay Triumphs in Tiffany Counterfeiting Case. `http://legalpad.blogs.fortune.cnn.com/2008/07/14/ebay-triumphs-in-tiffany-counterfeiting-case/`.

[98] H.F. Pitkin. *Wittgenstein and justice.* Univ. of California Press, 1972.

[99] I. Pollach. What's wrong with online privacy policies? *Communications of the ACM*, 50(9):103–108, 2007.

[100] Richard A. Posner. The Economics of Privacy. *The American Economic Review*, 71(2):405–409, May 1981. ArticleType: primary_article / Issue Title: Papers and Proceedings of the Ninety-Third Annual Meeting of the American Economic Association / Full publication date: May, 1981 / Copyright © 1981 American Economic Association.

[101] P.M. Regan. Safe harbors or free frontiers? privacy and transborder data flows. *Journal of Social Issues*, 59(2):263–282, 2003.

[102] Priscilla M. Regan. *Legislating Privacy: Technology, Social Values, and Public Policy.* The University of North Carolina Press, Chapel Hill, NC, 1995.

[103] Joel R. Reidenberg. Resolving conflicting international data privacy rules in cyberspace. *Stanford Law Review*, 52(5):1315–1371, 2000.

[104] J.R. Reidenberg. E-commerce and trans-atlantic privacy. *Houston Law Review*, 38:717–749, 2001.

[105] Marc Rotenberg. *The Privacy Law Sourcebook 2000: United States Law, International Law, and Recent Developments.* Electronic Privacy Information Center, 2000.

[106] Donald A. Schön. Generative Metaphor: A Perspective on Problem-setting in Social Policy. In Andrew Ortony, editor, *Metaphor and Thought*, chapter 9. Cambridge University Press, 1993.

[107] Steven L. Schwarcz. Private Ordering. *Northwestern University Law Review*, 97(1):319–349, 2002.

[108] P.M. Schwartz. Privacy and Democracy in Cyberspace. Available at SSRN: `http://ssrn.com/abstract=205449` or DOI: 10.2139/ssrn.205449.

[109] Herbert A. Simon. *Models of Bounded Rationality.* MIT Press, 1982.

[110] Daniel J. Solove. *Understanding Privacy.* Harvard University Press, Cambridge, MA, 2008.

[111] Richard J. Sullivan. Opinion in ebay v. tiffany. Retrieved from `http://fortunelegalpad.files.wordpress.com/2008/07/opinion-in-tiffany-v-ebay.pdf`.

[112] Latanya Sweeney. K-Anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness, and Knowlege-Based Systems*, 10:557–570, 2002.

[113] Anne Szustek. Louis Vuitton wins $64 Million in eBay Lawsuit. Retrieved from `http://www.groundreport.com/Business/Louis-Vuitton-wins-64-Million-in-eBay`.

[114] The White House. A Framework for Global Electronic Commerce, July 1997. Retrieved fom `http://www.technology.gov/digeconomy/framewrk.htm`.

[115] TRUSTe. TRUSTe Website. See `http://www.truste.com/`.

[116] Twitter.com. Twitter Privacy Policy. Retrieved from `http://twitter.com/privacy`, June 2009.

[117] United States. Fair Credit Reporting Act (1970). In Marc Rotenberg, editor, *The Privacy Law Sourcebook 2000: United States Law, International Law, and Recent Developments.* Electronic Privacy Information Center, 2000.

[118] United States. Privacy Act (1974). In Marc Rotenberg, editor, *The Privacy Law Sourcebook 2000: United States Law, International Law, and Recent Developments.* Electronic Privacy Information Center, 2000.

[119] U.S. Department of Commerce. European Union Safe Harbor Documents. Available at http://www.export.gov/safeharbor/eu/eg_main_018493.asp.

[120] U.S. Department of Commerce. FAQ - Dispute Resolution and Enforcement. Available at http://www.export.gov/safeharbor/eu/eg_main_018383.asp.

[121] U.S. Department of Commerce. FAQ - Self-Certification. Available at http://www.export.gov/safeharbor/eu/eg_main_018388.asp.

[122] U.S. Department of Commerce. Frequently Asked Questions (FAQs). Available at `http://www.export.gov/safeharbor/eu/eg_main_018493.asp`.

[123] U.S. Department of Commerce. Safe Harbor List. Available at http://web.ita.doc.gov/safeharbor/shlist.nsf/webPages/safe+harbor+list.

[124] U.S. Department of Commerce. Safe Harbor Privacy Principles. Available at http://www.export.gov/safeharbor/eu/eg_main_018475.asp.

[125] US Department of Commerce. Us safe harbor agreeement. Retrieved from http://www.export.gov/safeHarbor/.

[126] U.S. Department of Health, Education, and Welfare Secretary's Advisory Committee on Automated Personal Data Systems. Records, Computers, and the Rights of Citizens. Washington, D.C.: Government Printing Office, 1973.

[127] Gus Van Harten. Private Authority and Transnational Governance: The Contours of the International System of Investor Protection. *Review of International Political Economy*, 12(4):600–623, October 2005.

[128] S. Walker. *Taming the system: The control of discretion in criminal justice, 1950-1990.* Oxford University Press, USA, 1993.

[129] Samuel D. Warren and Louis D. Brandeis. The right to privacy. *Harvard Law Review*, 4(5):193–220, 1890.

[130] Alan F. Westin. *Privacy and Freedom.* Atheneum, 1967.

[131] Alan F. Westin. Social and political dimensions of privacy. *Journal of Social Issues*, 59(2), 2003.

[132] Alan F. Westin and Michael A. Baker. *Databanks in a Free Society.* Quadrangle Books, New York, NY, 1972.

[133] Daniel Witte. Targeted Advertising Cookie Opt-Out (TACO). See `http://taco.dubfire.net/`.

[134] L. Wittgenstein. Philosophical investigations, trans. g. anscombe. *Malden: Blackwell Publishers*, 1958.

[135] Yahoo.com. Yahoo! Privacy: Ad Serving. Retrieved from `http://info.yahoo.com/privacy/us/yahoo/adserving/`, June 2009.

[136] Yahoo.com. Yahoo! Privacy Center. Retrieved from `http://info.yahoo.com/privacy/us/yahoo/`, June 2009.

[137] Yahoo.com. Yahoo! Privacy: Products. Retrieved from `http://info.yahoo.com/privacy/us/yahoo/products.html`, June 2009.

[138] Jonathan Zittrain. *The Future of the Internet and How to Stop It.* Yale University Press, New Haven, CN, 2008.

[139] Jonathan L. Zittrain. *Jurisdiction.* Internet Law Series. Foundation Press, 2005.

[140] Jonathan L. Zittrain. The generative internet. *Harvard Law Journal*, 119, 2006.