Explorations in Cyber International Relations

Massachusetts Institute of Technology Harvard University

Mixed Context and Privacy

Jesse H. Sowell

Engineering Systems Division
Massachusetts Institute of Technology

August 15, 2010

This material is based on work supported by the U.S. Office of Naval Research, Grant No. N00014-09-1-0597. Any opinions, findings, conclusions or recommendations therein are those of the author(s) and do not necessarily reflect the views of the Office of Naval Research.

Citation: Sowell, J. H. (2010). Mixed context and privacy. *Proceedings of 2010 TRPC Conference*.

Unique Resource Identifier: https://ssrn.com/abstract=1989157

Publisher/Copyright Owner: © 2010 TRPC.

Version: Author's final manuscript.

Mixed Context and Privacy

Jesse H. Sowell

Abstract—Users engaging online service providers (OSPs) such as Google, Amazon, and Facebook encounter environments architected by a single actor (the OSP), but comprised of content and executable elements potentially provided by multiple actors. For the ten OSPs analyzed, privacy policies only cover content provided directly by the OSP. Content provided by external (third party) content providers, such as advertising networks and third party developers, are governed by a different set of privacy policies. In effect, users face environments comprised of mixed content governed by potentially conflicting privacy policies. Reasonably unraveling these conflicting privacy guarantees confounds the process of determining whether users' privacy preference are satisfied. The notion of a *mixed context* describes scenarios where a user is faced with multiple, potentially conflicting policy guarantees within a seemingly uniform, contiguous environment.

This paper develops mixed context as a metaphor that informs the design of privacy policies and the attendant privacy tools. Mixed context has also led to insights into actor incentives and dependencies that shape the design of policies, online environments, and ultimately the balance between advertising (re)targeting and user privacy. The mixed context metaphor draws evidence from OSP privacy policies and builds on Nissenbaum's notion of contextual integrity [29] as an analytic framework for evaluating privacy implications. This framework describes privacy in terms of participants' context-specific norms that are rooted in an experience-based understanding of the environment and the constraints on the behaviors of other actors in that environment. The instances of mixed context presented here confound this process because, although the environment is architected by a single actor and appears to be a single context, closer inspection reveals it is in fact governed by multiple, potentially conflicting policies. The mixed context metaphor has also helped surface institutional incentive structures that confound the development of meaningful privacy policies and tools. An immediate observation is that many of the actors contributing to the mix are invisible to the casual user. This impedes the development of reasonable expectations about a particular environment based on attributing elements of the experience to particular actors. Second, "invisible" non-OSP actors, in particular advertisers, are not directly accountable to users with regard to how they use information for (re)targeting of advertisements. OSP privacy policies provides *conceptual evidence* of mixed context; recent media investigations [39] have documented (observed) instances of mixed context outcomes "in the wild."

Although superficially a technical coordination problem, resolutions to mixed context problems are rooted in both technical means and the institutional arrangements of actors. The common "service-and-utility" framing identified in the privacy policy focuses on the benefits of targeting while underplaying privacy implications. Mixed context attempts to avoid interest-specific metaphors such as *service-and-utility* and value-laden metaphors such as those focusing on the contrast between privacy and surveillance. As applied here, the focus is to identify shared concerns that contribute to a collaborative understanding of the flow of user information that has collateral benefits for both advertising and privacy objectives. Evidence of deficiencies and mixed context have been identified via a bottom-up analysis of privacy policies. In contrast, design and policy recommendations are couched in a top-down institutional analysis that presents incentives for developing tools that convey the implications of mixed context in situ.

1 Introduction

Recent media coverage provides a number of concrete instances of the mixed context problem and its outcomes [39], [38]. For instance, a teenage girl named Caitlin has performed searches related to weight loss and has now been increasingly targeted by weight loss advertisements. Per Caitlin's statement, these are discomforting because they continuously perpetuate her weight concerns, even when she is not specifically browsing weight loss content. Another user searched for information about a medical condition she was being tested for; after finding out she did not have the condition she continues to receive related advertisements. In another instance, a Marketplace contributor volunteered to use her profile, held by Axoicom, as an illustration of advertising profiling and marketing [38]. She is classified under cluster 26, a 'Savvy Single' and among the attributes is an interest in venues for cocktails and drinks. Although she clarified in the interview she looks for these mostly as a venue to meet friends when on travel and even joked "I'm not that big a drinker, really…" there is the potential harm of this profile being taken out of context. Each of these situations can be characterized as an outcome of mixing individually "innocuous" attributes to create an aggregate image of a user. This work characterizes the scenarios where this happens as mixed contexts.

One source of mixing is the environment architected by online service providers (OSPs). OSPs such as Google, Facebook, and eBay architect environments that provide users with services customized based on user supplied information and (automatic) tacit data collection such as monitoring clickstreams and inferring interest based on content viewed. For the casual user, visiting an OSP's site is akin to visiting a single, branded environment: they are "going to Google" or "going to Yahoo!" Although these environments are architected by a single OSP and nominally behave as uniform context, the underlying structure of the environment is architected as a composition of dynamic elements and content provided by the OSP in conjunction with a variety of third party sources.

A survey of ten OSP privacy policies[1] explicitly note that OSP policies govern only the content and dynamic elements provided directly by the OSP itself. Elements and content embedded in the OSP environment, but developed and distributed by third parties, are governed by separate privacy policies that may not coincide with those of the OSP. The notion of a *mixed context* builds on recent privacy frameworks to describe the architecture and implications of scenarios where a user is faced with multiple, potentially conflicting policy guarantees when visiting a seemingly uniform, contiguous environment.

Ideally, users rely on an understanding of OSP privacy policies to decide whether an OSP's data purposes are acceptable relative to the user's personal privacy preferences. Economic privacy policy analyses couch 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 the utility gained by sharing information [1], [2], [3], [7], [8], [9], [21], [22]. Here, policies are still considered decision making tools, but the focus is on the application of mixed context as policy and design metaphor that provides insights into meaningfully conveying the privacy implications of online behaviors to users *in situ*. Nissenbaum's recent reconceptualization of privacy as contextual integrity [29] provides a useful analytic framework. Nissenbaum argues that privacy is better articulated in terms of a normative context based on rules of appropriate behavior and rules regarding how information may be distributed beyond that context. In contrast to individual violations of contextual integrity, the mixed context problem describes the architectures and actor relationships that perpetuate questionable mixing of context-specific information to create an aggregate image of the user. As will be discussed in Sections 3 and 4, mixing may be beneficial or harmful.

Mixed context highlights a lack of tools for meaningfully conveying privacy implications to users. First, online environments do not necessarily behave as experience goods[2]—it is not obvious through casual engagement the reputation of these providers, which dynamic elements are provided by whom, and whose privacy policies bind to these elements. Although privacy policies indicate that conventional personally identifying information (PII such as name, social security number, address, e-mail, telephone number, etc.) is not shared with third parties, the guarantees and accountability related to tacit data collection (behavioral data collection via clickstreams and cookies) are extremely weak. In particular, under current actor arrangements, advertisers and other third parties have little accountability to users for how profile data is used[3]. Further confounding the situation, the mix of third parties is dynamic. Each encounter with an OSP architected environment may expose the user to a different set third parties. It is argued that it is very difficult, if not impossible, to develop normative expectations for such mixed contexts as they are currently architectedand that new approaches that surface the implications of mixed context are necessary.

Mixed context is presented as a generative metaphor[4] that informs both the design of online privacy policies and the architectures of OSP environments these policies govern. While a centralized policy is necessary for reference, this work recommends introducing explicit elements into the environment that convey meaningful signals to the user. For instance, explicit labels for third party content convey the user is not just interacting with the OSP. Such labels may also provide information about the third party actor's reputation, in particular information about third parties' privacy and profiling practices. Focusing on mixed context, in particular who does the mixing and based on what incentives, has also helped surface information on the collaboration amongst OSPs and third party content providers. Policy and institutional recommendations build on this information to suggests strategies for transitioning to institutional arrangements[5] that foster collaborative analyses and tools for meaningfully conveying privacy implications to casual users. In particular, the discussion will describe strategies that may ultimately incentivize OSPs to better police third party content providers based on their reputation for mixing user data.

To develop the arguments for mixed context and policy framing, overviews of Nissenbaum's contextual integrity and Schön's generative metaphor are presented in Section 2. Section 3 provides evidence of mixed context based on conceptual instances of online privacy deficiencies found in the policy sample and recent, concrete instances reported in the media [39], [38]. Conceptual deficiencies are couched in a brief description of the sample of privacy policies analyzed, categorizations of non-OSP content providers, and relevant privacy policy

1. Henceforth simply the policy sample. The original data analysis was performed by the author for a SM thesis in Technology and Policy [37]; portions of that data are reproduced here.

2. Nelson [28] 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 experiences with the good and understands the implications of purchasing the good.

3. The process of finding tools for managing advertiser targeting, such as those provided by OSPs or the NAI, requires similarly extensive search (in the sense of Nelson) process necessary to surface meaningful privacy implications. Even if these tools are identified, it is unclear whether they truly update the aggregate image of the user or simply satisfice by updating it to indicate the user does not want to see advertisements of a certain variety while continuing to use the attribute in question for other profiling activities

4. Generative metaphor is in the sense of Schön [34]; see section 2.2 for a brief description.

5. Incentive structures are couched in the design and evolution of institutional dynamics; the analysis is in the tradition of institutional economics studies [30] in political science.

excerpts. Section 4 presents an institutional analysis of actor incentives, strategic options for shifting the incentive structure, attendant design decisions, and policy development recommendations. Section 5 concludes with a summary and brief discussion of future work on empirically quantifying mixed context.

2 Background

Mixed context builds on Nissenbaum's framework of privacy as contextual integrity. Mixed context highlights how the architecture of online environments and the attendant policies can potentially violate contextual integrity. Mixed context is presented as a generative metaphor that can help surface the privacy implications of online architectures and the actors that participate in the design of these architectures. Nissenbaum's contextual integrity[6] is introduced below; generative metaphor described in Section 2.2.

2.1 Contextual Integrity

Nissenbaum argues that conventional frameworks for reasoning about privacy "fail[s] to clarify the sources of their controversial nature" [29, p. 119][7]. The instances of profiling in Section 1 illustrate instances of the controversial nature of mixing. The disconcerting feeling created by continuous targeting of weight loss advertisements certainly has implications for individual's choices when they wish to confront personal issues they may be dealing with. In [38], the host repeatedly points out the weighting of certain attributes, notably the interest in cocktails; inappropriate weighting of such attributes can potentially distort the image of an individual. The original context was a search for a local venue to have cocktails with friends while on travel. Although the host made light of this categorization, taken out of context or mixed with other facts out of context, this information could contribute to a negatively distorted image of the subject.

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 approaches are based in failures of court cases to set applicable precedent 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[8]. As will be elaborated Section 3, one root of the mixed context problem is that the process and source of mixing is invisible to the user and, in the case of advertisers, those doing the mixing have little accountability to end users. Contextual integrity is an effective analytic framework because it focuses on reasoning about the root of privacy problems in a way that establishes applicable precedent rather than an abstract theory that, while covering, provides little actionable guidance for policy and design.

Nissenbaum also critiques the typical dichotomies used to describe privacy policy issues: sensitive and non-sensitive, private and public, government and private. It is in the reconstruction of framing and metaphors used for understanding privacy that Nissenbaum's contextual integrity and Schön's generative metaphor (next section) provide complementary analytic frameworks. Identifying the root of privacy conflicts, Nissenbaum's contextual integrity provides a more textured depiction of individuals' activities as they move "into, and out of, a plurality of distinct realms" [29, p. 137] that may not be sufficiently represented by typical dichotomies.[9]

Up to this point, there has not been a precise distinction between context and environment. The term environment provides some notion of place, such as the home, a doctor's office, or a built online environment, such as an eBay auction site or a Yahoo! chatroom[10]. Each of these built environments comprise elements that can contribute to, or detract from, the privacy of behaviors that occur within these environments. Contexts are defined in terms of a distinct sets of norms that comprise notions of "roles, expectations, actions, and practices" [29, p. 137]. These norms evolve socially and, while invocations these norms may play out at a particular time and in a particular environment, the application of contextual norms may legitimately occur in a variety of host environments. Contextual integrity provides a cohesive perspective on how these norms contribute to meeting individuals' privacy expectations within a given social context.

Contextual integrity also gives insight into how to

6. The definition of contextual integrity presented here is from [29]. Barth et. al. have formalized contextual integrity as a linear temporal logic and provide a definition that collapses norms of appropriateness and norms of distribution into a single "transmission norm" [6]. This work builds on the "unpacked" version because it more directly maps to both policy evidence and concrete instances in Sections 1 and 3 as well as the discussion of incentive structures in Section 4.

7. To avoid misattribution, the author chose Westin as the instance of broad, abstract privacy theory; see [29] for Nissenbaum's choice of instances.

8. Nissenbaum cites Reagan's work [32] regarding the devolution of debates over privacy into adversarial confrontations between interest groups intent on promoting their interests over their opponents' [29, p. 122]. Mixed context as a generative metaphor will revisit reconciling actor interests in Section 4.

9. Mixing violates the boundaries of these realms; moreover, the transfer across these boundaries is at the discretion of advertisers. To foreshadow the value-free character of the mixed context metaphor, mixed context may be applied by privacy advocates to argue "how mixed" and advertisers may argue whether mixing is harmful or not. Stated as such, it will be argued that tools that facilitate user input regarding harm is the only empirically valid solution.

10. In this discussion, the term environment is technically laden with the connotation of a "built environment" or an "engineered environment" in contrast to the more purely socially constructed environments implied by Nissenbaum's notion of context. For completeness, environments are not limited to the built environments described by these examples, but this discussion is limited to built environments whose architectures are largely man-made. The impact of the architecture of the environment is also discussed by Solove in [36].

accurately and consistently surface individual privacy preferences. In one sense, contextual integrity is the dual of behavioral profiling applied for (re)targeted advertising purposes. In both cases, user preferences are based on contextual indicators. 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" [29, p. 119]. This definition builds on two types of norms: norms of appropriateness and norms of flow or distribution.

Norms of appropriateness dictate what information is fitting to reveal in a particular context. Contexts differ along 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.

Given these rules, the mixed context problem occurs when the architecture of an environment facilitates repeated contextual integrity violations. In this case, the environment does not provide sufficient notice to the user that data is being collected or who is collecting this data. Norms of dissemination occur when actors tacitly collect behavioral data or inferred interest data for use with data in other contexts. Norms of appropriateness are violated when information from one context is actually used for (re)targeting in a different context.

Although this is an elegant and insightful way of describing privacy, Nissenbaum admits it is 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 the data collection methods available in conventional terrestrial environments, the (built) online environments that host online contexts may be instrumented to help collect this information in real time, as it occurs. Such instrumentation has, in principle, the same privacy connotations as other of tacit data collection such as those used for behavioral profiling. The design of policy experiments necessary to understand the potential for constructing context based on this information and the necessary privacy disclosures is discussed in Section 4.

2.2 Generative Metaphor

The notion of generative metaphor is a useful tool for understanding how a particular problem is framed, how different framings influence the design and implementation of potentially competing solutions, and the insights that can be gained from reframing to surface common interests [34]. 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 developing (social) policy is often rooted in how the problem is framed (by the metaphors used) rather than the means of the problem itself. In effect, the framing of the problem shapes how users and designers perceive a situation and subsequently the types of solutions that are applied.

One framing of a problem may characterize a service as suffering from "fragmentation" and prescribe "coordination" as the remedy [34, p. 138]. Alternately, the service may be described as "autonomous," which does not imply the problems associated with 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. Like Nissenbaum, Schön also eschews oversimplifying dichotomies.

A key deficiency identified in the policy sample is online service providers' one-sided *service-and-utility* framing. The *service-and-utility* framing satisfices to requirements to disclose information collection and how information is used. Although this disclosure complies with the black letter of regulatory norms like the Fair Information Principles (FIPs)[11], these disclosures are embedded in arguments that data collection is to improve *service* and provide the user with novel and valuable *utility*. In effect, as framed by the online service provider , (tacit) data collection is presented as contributing to a process of continually improving the services that provide utility to the customer. In terms of the framing dichotomies above, it may be argued that *service-and-utility* implies a "stagnation/continual improvement" metaphor that implies user benefits are driven by continual improvement and novel data purposes. The potential of some novel purposes to have negative effects is not captured under stagnation or continual improvement.

11. The FIPs are a common normative starting point for developing privacy regulatory frameworks. The FIPs were developed concurrently by a number of international government organizations and regulatory bodies in the late 1970's (The Council of Europe, the Organization of for Economic Co-operation and Development (OECD), the US Department of Health, Education, and Welfare (HEW), and Britain's Younger Committee were contributors). Norms include those enjoining rights to access and change personal data, consent to use data, use limitations, guarantees to the integrity and security of data, and processes for enforcement and redress of violations. Rotenberg's provides an account of a number of the original sources in [33].

Schön's discussion describes the effect of framing on problem solving. Schön highlights that problem setting, in contrast to the process of problem solving, is often characterized by how the problem is framed. Simon's problem solver explores the problem-space to optimally satisfy some objective function [35]. 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[12]. Rather than taking a potentially coloring framing as *the* problem, policy problems should be approached in terms of problem *setting* as a means of comparing different frames. As implied above and developed in Section 3, the *service-and-utility* framing is particularly coloring. Regardless of intent, the *service-and-utility* framing highlights the benefits of customization via tacit data collection but does not convey the dangers of contextual integrity violations enabled by unchecked tacit data collection.

In contrast to the one-sided metaphors described above, a *generative* metaphor is one that creates "new perceptions, explanations, and inventions" [34, p.]. 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. Policy analysis provides evidence of the metaphors used by online service providers to shape users' perception of their privacy practices. The policy sample provides a surfeit of evidence for the *service-and-utility* framing; absent from the sample are metaphors that meaningfully convey the potential dangers of mixing tacit data across contexts. The mixed context metaphor attempts to reframe policy and architectural implications by moving away from the value-laden language of utility, services, and surveillance to a framing that can highlight the positive and negative implications of information flow across contexts. It should be made clear that the potential insights into a problematic situation, here online privacy, are not wrapped up in the generative metaphor waiting to be unpacked. The new perspective (framing) shifts the focus to different elements of the problem, giving these primacy and subsequently drawing the designer's attention to previously unattended dynamics. In this case, the mixed context perspective has helped focus on elements of the privacy policies that have surfaced actor relationships and dependencies that contribute to unchecked mixing of tacit data (evidence in Section 3, discussion of institutional arrangements in Section 4).

The process of problem setting becomes what Schön calls "a kind of policy-analytic literary criticism" [34, p. 149] that helps analysts and designers 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. In this case, what data is mixed, who does the mixing, and the incentives that perpetuate mixing is that 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" [34, p. 152].

3 Evidence and Implications

The original analysis that identified the mixed context problem was based completely on the content of the policy sample. Recent media coverage [39], [38] provides concrete instances and further evidence of mixed context outcomes "in the wild." Evidence from the policy sample will be referred to as policy evidence; evidence of instances of mixed context outcomes from the media will be referred to as observed evidence. Policy evidence comprises textual descriptions of the tacit data collection and descriptions of the OSPs and actors that provide dynamic content. The sample provides qualitative evidence that mixed context problems *exist*. Recent observed evidence is both a confirmation and a lead-in to questions of "How mixed?" addressed in Section 3.4. Policy evidence provides little in the way of the precise data categories collected by OSPs and third parties[13]. Policy evidence does provide sufficient information to identify the mechanisms and architecture that supports mixed contexts. The policy evidence also provides enough information about actor relationships to map out the incentives that contribute to the mixed context problem.

As per earlier discussion, much of this information is embedded in a *service-and-utility* framing that focuses on the benefits of customization. The *service-and-utility* metaphor, when placed alongside a discussion of mixed context, is almost paradoxical in its appeal to trusting the OSP. OSPs elicit trust with reassurances that conventional PII is never shared and claims that non-PII

12. Schön also relates the problem setting process to a construct by Dewey referred to as the "problematic situation," which he references at [34, p. 146] and refers the reader to [13].

13. Recently some third parties have provided mechanisms for selecting which categories of information they would like associated with their identity, but these are monolithic lists of broad interests taken out of context and that require repeated visits to monitor and keep track of. These are discussed and critiqued in Section 4.

is "innocuous" and facilitates better service and utility delivered to the customer. At the same time, and receiving substantially less focus, the policies also provide evidence that OSPs have architected an environment by which third parties can build a comprehensive aggregate images of users. This characterization is not a wholesale condemnation of OSPs for pursuing interests that further their business objectives. This characterization does critique the failure to efficaciously disclose the *implications* of context-violating tacit data collection and critiques the limited choice sets available for the management of tacit data collection, and by proxy, the construction of an aggregate image. The discussion unravels this story by presenting evidence of the *service-and-utility* metaphor (Section 3.1), a refined definition of mixed context (Section 3.2), instances of mixing within and across environments (Section 3.3), and finally an initial discussion of "How mixed?" (Section 3.4) that transitions the discussion from evidence to recommendations espousing a collaborative regime for resolving some of these issues.

3.1 Evidence of the *Service-and-Utility* Metaphor

Across the sample, the *service-and-utility* framing highlights the benefits of customization. The *service-and-utility* framing is built on two dichotomous metaphors: "customized/generic" and "improvement/stagnation." This framing is consistent in all ten privacy policies in the sample. A couple of instances illustrate the framing. Yahoo! provides a blanket statement that it reiterates throughout its privacy policy:

> 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.

Amazon.com provides one of the most consistent *service-and-utility* policy framings:

> The information we learn from customers helps us personalize and continually improve your shopping experience at Amazon.com...
> Cookies are alphanumeric identifiers that we transfer to your computer's hard drive through your Web browser to enable our systems to recognize your browser and to provide features such as 1-Click purchasing, Recommended for You, personalized advertisements on other Web sites (e.g., Amazon Associates with content served by Amazon.com and Web sites using Checkout by Amazon payment service), and storage of items in your Shopping Cart between visits...
> 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.

Amazon.com (and others) build on the *service-and-utility* metaphor by nominally providing information about process that affect user privacy, but couched in "all-or-nothing" statements such as above. This effectively limits the user's choice set. The user is then faced with two bundles (as it relates to tacit data collection): limit tacit data collection but give up service or trust Amazon to act in the user's best interest regarding service and customization, but give up control of privacy.

Of the OSPs in the sample, Amazon does provide the finest-grain access to the aggregate image, albeit in a limited fashion. Amazon provides tools that allow users to select which *recently viewed* items contribute to Amazon's image of the user [4]. For example, if a user typically browses books on statistics, they may find substantial utility in the Amazon's recommendation services, which recommend books based on similar category and what others browsing the same books viewed and bought. On the other hand, the user may not want their image distorted by spurious searches that do not reflect their genuine interests. For example, if a friend of our statistics user comes over and they are talking about the prices of baby strollers and they browse the selection of baby strollers recreationally (rather than as serious buyers), but the statistics user runs the risk of polluting her aggregate image (at least temporarily). Although this is a relatively harmless example, the instances from observed evidence in Section 1 can damage one's reputation (implication of a drinking problem) or be disconcerting (weight loss advertisements).

To correct this situation, the statistics user must remember to remove these recently viewed items before they are committed permanently to the aggregate image. If they are committed, the only recourse available is to delete the entire aggregate image, effectively losing any genuine preferences contained therein. Although the ability to remove recent items is a useful feature, it places a burden on the user to remember to do this. Moreover, at the time of writing, this feature was "buried" at the bottom of the page for each item viewed. Section 4 will return to this discussion, highlighting the processes and tradeoffs regarding how to better integrate signals indicating that mixing is occurring more prominently into the user's workflow.

Across the board, OSP's privacy policies satisfice to notice requirements, but the information regarding the tacitly collected "non-PII" that contributes to behavioral profiling is typically a description of the technical methods being used and is couched the *service-and-utility* framing. This framing is a reflection of OSP interests. More specifically, it is in the interest of the online service provider for the user select the "trust us" bundle, allowing OSPs to make decisions regarding

appropriateness (and dissemination) for contexts hosted in the environments they have architected.

3.2 Mixed Contexts

Privacy policies provide evidence of the mixed context, but insufficient information to quantify "how mixed" the context may be (Section 3.4). Providing the information necessary to identify these implications is referred to as mixed context disclosure. Mixed context disclosure is fundamentally based on OSPs 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. Beneath the veneer of the *service-and-utility* framing are two mechanisms that facilitate dissemination via mixing. The first is tacit data collection via technologies such as cookies and beacons by OSPs that allow third parties to directly embed these technologies in OSP architected environments. The second is inferring segment information based on the targeting criteria; this form of dissemination could theoretically allow an advertiser to collect any and all of the attributes of a user that the OSP uses to target advertisements. Evidence of these processes are derived from roles and relationships between OSPs and third parties described in the policy sample.

3.2.1 Third Parties

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.

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 is that most support services fall along the lines of billing, logistics management, and specialized analytics. OSPs 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. [4]

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. [23]

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.

Table 1 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 associated 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, the *service-and-utility* framing and focus on PII may actually conceal re-identification privacy risks by demoting non-PII to a second-class category of information with respect to information security requirements.

Advertisers (and advertiser networks) comprise the second category of third party actors described in the policy sample. All privacy policies make some mention of advertisers. In addition to indicating that their privacy policy binds only to the OSP, a second trend is that all OSPs in the sample make at least one reassurance that information shared with advertisers, either directly by the OSP, through segment inferences, or collected directly via cookies or other instrumentation, is not personally identifiable. In the case of direct data collection via cookies or other instrumentation, the OSP architected environment facilitates direct access to the user's be-

OSP	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 1: **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 [14] 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," [25] implying, but not specifying, a contractual enforcement mechanism.

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

TABLE 2: **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 *explicit* 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.

haviors. Moreover, direct access allows the advertiser to use the content of the page being viewed to further contribute to the aggregate image.

OSPs do reaffirm advertisers do not have direct access to PII. Some OSPs (summarized in Table 2) do make mention of advertisers using segment information, such as age range or region, when advertisements are served to users. With the limited exception of LinkedIn, disclosure of the implications of advertiser data aggregation is a simple statement about segmenting.

Disclosing the practice of segmenting and describing the implications is very different. The former satisfices to requirements that OSPs give notice that information is collected. Describing the implications requires a meaningful articulation of the benefits and harms that contributes to the user's understanding of (or experience with) the environment. 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 collect[14]. 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. [40]

This particular statement is somewhat abstract. It satisfices to notice requirements, but "targeting criteria" may not be especially meaningful to the casual user. Moreover, it does not convey that targeting criteria may vary across visits, allowing advertisers to collect a variety of user interests over time and across environments. 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). [4]

In terms of developing an understanding of the implications of information collected by advertisers, the jump from the abstract or basic information about segmenting to the implications of (re)targeting based on a comprehensive aggregate image constructed across multiple contexts is an exercise left to the user.

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 OSPs such as Facebook,

14. The FTC, in [17], 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.

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

TABLE 3: **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* is included as a point of reference and indicates that the privacy policy indicates platform developers are contractually obligated to protect users privacy 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 nor 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.

LinkedIn, and MySpace that provide developer APIs. 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 3.

The implications of these differentiated obligations are, like the dilemma with advertisers, that the user is 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, at least in its privacy policy, it does not support (or encourage) using these applications:

> 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 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. [27]

MySpace and other OSPs that support the development of embedded applications provide platforms that facilitate developing applications as a competitive feature of their product, but distance themselves from both the abuse and the burden of policing these applications.

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, application C 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 3, 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. [23]

Assuming the user considers 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 may distort 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 [35], [31]. Bounded rationality indicates that humans have a limited capacity for reasoning about a situation in a finite amount of time, thus limiting pure rational choice that assumes perfect information and sufficient time to process all necessary information [35]. A consequence of bounded rationality is rational ignorance [31]. 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 OSP.

3.2.2 Dynamic Contexts

The problem of mixed context is further confounded by the fact that OSP architected environments comprises content contributed by a dynamic set of actors (observers, in the sense that they can each collect information about users). 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 Nissenbaums terms, each of these particular configurations is a source of privacy norms. Thus, each configuration would comprise a unique (static) set of actors and may elicit a unique privacy response from the user. Rather, while the OSP's privacy policy remains the same, the set of third party content elements is not guaranteed to originate from the same static set of actors upon every engagement.

For instance, advertisements are presented based on the inferred preferences of the user based on the aggregate image and advertising segments associated with the primary OSP content being viewed. Moreover, an OSP may have contracts with multiple advertisers and/or advertising networks. Further still, the OSP maintained aggregate image may change. The result is that there are at least two factors that may lead to a dynamic configuration. First, the OSP may (re)target based on an updated aggregate image. As the aggregate image maintained by the OSP changes, associated interestes may change, finding the user continually exposed to a different set of third party content providers. Second, to satisfy all its advertisers, the OSP may show the user an advertisement targeted to one segment (girls interested in weight loss) from one content provider on visit i, another weight loss related content provider on visit $i+1$, and yet others on subsequent visits. The result is that a particular, unique mixed context is ephemeral, and may reoccur nondeterministically, if at all.

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, especially considering the set of third parties is neither finite nor static. The second is difficult because combinations are ephemeral and unpredictable and do not occur with sufficient frequency to develop expectations even if the user was aware of their existence and had the motivation and knowledge to successfully reason about each configuration's implications as they occur. A possible solution to this problem is presented in the next section, recommending tools that concisely convey the reputations of contributing actors contributing to a mixed context.

Although this discussion highlights the information flows across environments that may host different contexts, not all environment spanning flows are necessarily context violations. Recall from the definitions of environment and context that contexts may legitimately span multiple environments. As a step towards understanding mixing and the environment, the next section describes mixing within a single OSP architected environment and across different environments.

3.3 Mixing and the Environment

OSPs and advertisers garner competitive advantage by having a more complete image of the user. Two conceptual types of mixing were identified in the policy sample: mixing by an actor (typically by the OSP) within a single environment and mixing by a single actor across multiple environments. Most of the outcomes of mixing discussed thus far focus on mixing across environments. The following excerpts and instances from policies illustrate how user data is disseminated. The following describes the distinction between contexts created by users using OSP-supplied privacy tools relative to the mixing performed underneath the hood by the same OSP across these contexts.

OSPs indicate that they collect information from a variety of sources to supplement the user profile. One instance is the case of Facebook:

> 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. [15, Emphasis added]

The list of "other sources" covers both mixing within (platform developers and information from other users of Facebook) and across environments (blogs, instant messaging, and newspapers).

Although Facebook does give users notice it is supplementing user profiles with information from a variety of sources, it does not necessarily provide users with the ability to audit the content of such an aggregate image or validate the provenance of this information. Taken alone, this disclosure gives examples of broad contexts from which Facebook draws information to construct its aggregate image of a user. Depending on the user's preferences, the use of information from other users of Facebook may violate contextual integrity. In the case of social network applications, the objective is to construct a community of users, a space for social interactions.

In effect, it is an environment for constructing contexts. Tools for enforcing privacy expectations for users have been developed and refined to allow custom groups and contexts. As this environment has taken off, tacit data collection tools have proliferated, but tools for reifying contextual boundaries between users do not necessarily affect tacit data collection by either the OSP or third parties.

For instance, limiting access by school or workplace based "networks" can be considered a coarse-grained implementation of contextual boundaries *between users*. OSPs do not indicate whether the OSP-developed aggregate image (sometimes referred to as a profile) will respect these boundaries. Some social networks allow users to create custom contexts (for instance limiting access to photos to a custom set of users rather than just within a particular network or just for "Friends"), there is no guarantee the information shared in these contexts will not be incorporated into the aggregate image of the user. Thus, although not personally identifying in the conventional sense, because OSPs uses this information for (re)targeting advertisements and the potential for advertisers to make inferences and link them to unique identifiers, whatever contextual boundaries users set up between themselves via intra-environment privacy tools do not necessarily affect the flow of the same information (minus conventional PII) to advertisers and other third parties. Moreover, advertisers may have precise segmenting information based on how advertising (re)targeting is negotiated between the advertiser and OSP or may have less precise information and rely on inferences. In either case, data initially mixed within the environment by the OSP may be available to advertisers. The caveat under which the flow of information to third parties occurs is that behavioral profiles are not linked to conventional PII even though it is linked to a unique identifier by both the OSP and third parties. Thus, even though the user may have specifically limited access to certain information to a context comprised of a specific set of actors, this information may be inferred by third parties, especially advertisers with a financial interest in constructing the most comprehensive profile possible.

Another instance of mixing within a single OSP architected environment is a recent GMail policy regarding advertisements that are shown alongside e-mail messages. Previously, advertisements shown alongside GMail messages were based on the content of that message alone. An update in the help section of GMail describes the new policy:

> [But] sometimes, the ads related to a particular message aren't good enough. Rather than show less relevant ads, Gmail can now instantaneously serve ads based on another recent message on the same page of your inbox, helping make the ads more relevant to you. For example, if your friend sends you a message to say happy birthday, but there aren't any good ads to show related to birthdays, you might see ads related to another message in your inbox instead – like flights to Chicago. [18]

This statement implies that mixing occurs only within the same page of your inbox. The remainder of the statement implies that no additional information is stored. Also note the second sentence is yet another instance of the *service-and-utility* framing.

It is arguable that e-mails constitute snippets of various contexts a user may engage in. In the most innocent case, the combination maybe be incidentally disconcerting. If a ephemeral version of an aggregate image based on the "same page" of e-mail is constructed to serve advertisements about interests the user is engaged in very recently (as implied by the recency of the e-mails), this could be very disconcerting. Consider a variation of the instance of searching for information about a potential ailment; instead of mixing information from a web search the information is drawn from a recent e-mail. Mixing advertisements related to potential treatments into the contexts of other e-mails may be disconcerting to the user, regardless of how "relevant" they may be. A more subtle question is what information is used to determine how relevant an advertisement is. As per the OSP's choices regarding mixing described earlier (Section 3.2.2), it may be the case that GMail just does not have advertisers that have chosen the "birthday" segment. The other possibility is that advertisement relevancy is based on either the user's overall aggregate image or the temporary "same page" image. In either case, mixing may again cause a disconcerting feeling for the user.

Mixing across environments was the first harm identified in the policy sample. The instances presented in Section 1 are ultimately the outcomes of mixing across environments. The policy sample provides sufficient evidence of the processes of mixing within and across described here. The media has recently provided additional observed evidence. Angwin and McGinty reports that in a survey of the top 50 websites (ranked by visits), 3,180 "tracking files" (cookies, web beacons, and other tracking mechanisms) were used [5]. The average number of tracking files per site was 64; Dictionary.com, Comcast, and Microsoft's MSN.com topped out the list with more than 100 tracking tools each. Further, two-thirds of the tools were installed by businesses invested in user profiling; the top of this list included Google, Microsoft, and Quantcast. Angwin and McGinty also describe the types of information collected. Their investigation developed an "exposure index" to describe what they called "aggressive surveillance." Among their findings were that 121 tools installed via the same OSP did not exclude collecting information about financial or health data. This application of mixed context highlights the mixing of content from different contexts and the reach of third party actors that users are exposed to when engaging an OSP.

3.4 How Mixed?

Thus far, mixed context has highlighted the architecture that facilitates mixing and the relationships amongst the actors that maintain these architectures and benefit from aggregation. As implied earlier, to better understand the effects of mixing, a natural follow-up question is "How mixed?" Although articles such as Angwin and McGinty [5] provide valuable insights, they have their own framing in terms of "tracking," "surveillance," and "spying" which potentially colors the conclusions of the analysis. Quantifying mixed context will, in the tradition of Schön's analysis, avoid conventional dichotomies and identify value-free measures that highlight common interests.

One question regarding mixing is how compatible different data categories are when combined in a particular context. This is not a universal, objective function over a set of data categories that applies to all users. Understanding what is compatible is user specific and requires developing a better understanding of how to represent context and the wide variety of preferences. For example, one user may be happy to have advertisers mix information about their hobbies with travel searches; this may be a very valuable mix of compatible interests, especially if they are going on vacation. Facebook's experiment with publishing individuals online purchases to users' news feeds resulted in a backlash. Information about purchases from the relatively private context of gifting was not compatible with a public feed. In effect, Facebook made an error with regard to mixing.

Pushing farther to understand how mixed draws on the language used to describe relevant advertising. The discussion of GMail advertising implied some advertisements are "more relevant" than others and that there "aren't any good ads" for the particular content the user is viewing. In terms of "How mixed?" this speaks to what constitutes a qualitative difference between the content of the current context and the advertisement, the frequency with which this occurs, and how frequently a particular advertisement is placed in a different context because it is "better." Returning to the case of Caitlin and weight loss presented earlier, one way to begin to measure the qualitative difference is to determine whether advertisements for weight loss were shown only when browsing content on weight loss (traditional content-based targeting), when browsing adjacent topics (browsing sites on favorite foods, shopping for clothes, physical activity-based recreation), or frequently across all topics. The latter two have the potential for substantial, user-specific qualitative differences in the compatibility of the contexts to arise. Caitlin has voiced discomfort that she sees these advertisements all the time. Others may find it useful to mix their interest with weight loss, food contexts, and physical activities, but consider clothes shopping a bit too much. Others still may prefer advertisements remain content-based and that weight loss advertisements should only be shown when the context is clearly related to weight loss. Yet others still may prefer completely random advertisements. In promoting mixed context as a design metaphor, an objective is to understand how to quantify these differences.

Recent media explorations of user profiling have drawn attention to the volume and variety of information contained within an aggregate profile. While there may be benefits to having a single aggregate image, issues of relevance discussed above indicate that some users may prefer to have context-specific aggregate images, potentially maintained by different actors. As a coarse-grained instance, consider the difference between Facebook and LinkedIn. Facebook has a distinctly playful tone, punctuated by applications revolving around users playing zombies and the Farmville game. LinkedIn purposely develops a more professional tone:

> 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.* [23, Emphasis added.]

In terms of maintaining separate aggregate images, a user may trust Facebook to mix content about their personal interests (zombies and all) while preferring to have the aggregate image maintained by LinkedIn mix only professional interests. It may be considered inappropriate or harmful to transfer interests from the informal context of Facebook to the more formal context of LinkedIn.

Returning to the health examples, users may not want their health information to become attributes of a publicly traded aggregate image. The user may gain utility if information about physical recreation activities they enjoy can be integrated into treatment options. In this example, there is utility to mixing information from outside interests by a trusted steward of the health information context, *within* that context. The trick is to introduce mechanisms that disincentivize health information leaking out, as described by the inferencing mechanisms discussed in previous sections.

These questions and categories represent the *kinds* of data necessary that could provide empirical evidence supporting institutional arrangements that introduce accountability for user profiling. This information is not immediately available to a single category of actor and, if collected by one category of actor, the metaphors used to describe it may well be as one-sided as the *service-and-utility*-based framing presented by OSPs. As will be discussed in the next section, it is arguable that

privacy and advertising preferences are two sides of the same coin and that the techniques used to identify what people are and are not interested in may also be applied to limit exposure to what is discomforting or what they consider inappropriate. The next section will present institutional arrangements that may incentivize this transition.

4 Recommendations

The objective of mixed context, as a generative metaphor, is to highlight the information flows across contexts and provide insights into the tools that can meaningfully convey the beneficial and harmful implications of these flows to users. The actor relationships developed in the previous section were developed bottom-up from evidence in the policy sample and supplemented with recent observed evidence reported by the media. The following recommendations and analyses proceed top down, starting with the privacy implications of comprehensive aggregate images as an outcome of existing institutional arrangements between OSPs and third parties. Based on these relationships, institutional analyses, highlight the strategic technical and policy options available to actors that produce and consume information affecting user privacy and advertisement targeting. Complementary to studies of what is technologically possible, institutional analyses focuses on what data collection and collaborations are strategically tractable. Given this starting point, the objective is to leverage tools and collaborations to incrementally transition to an institutional arrangement that supports accountability for mixing practices.

The options presented here are certainly not exhaustive, but illustrate the process and types of data collection and sharing that can help answer some of the questions of "How mixed?" The institutional arrangement proposed here suggests working towards establishing collaborative privacy standards development forums. Such a forum is based on a feedback process that uses constructive conflict (over strategic options) 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. One general objective is to make the OSP architected environment behave more like an experience good. The actual operationalization and implementation of objectives and strategies described here will be a product of the negotiations amongst actors within the collaborative regime. Through a process of issue reframing and empirical feedback, shared concerns and strategies may be developed that give rise to mutually beneficial strategies rather than pursuing actor-specific, first-best solutions that characterize the current arrangement.

It is important to note that the process described here is not intended as a "blue sky" articulation of how policy "should" be made. Rather, recommendations start with current institutional arrangement and describe the benefits of each category of actor contributing to a collaborative regime. This includes a menu of strategic options available to each and how these may affect the development of privacy standards. The next section ties together the evidence presented thus far to draw a picture of the current institutional arrangement and incentives. In particular, this illustrates the power asymmetry between the OSP-advertiser structure and users, whose interests are loosely federated by civil society organizations, privacy activists, and regulatory agencies.

4.1 Current Arrangements

Currently, the primary categories of actors are the OSPs themselves, advertisers, other third parties, users, civil society organizations, activists, and regulators. The primary actors affecting the evolution of privacy tools available to users has been the OSPs and advertisers as they respond to threats of regulation by bodies such as the FTC. Early on, concerns regarding how conventional PII was handled and protected became and issue, resulting in requirements for OSPs to provide privacy statements that assure users their data is protected. This phase of online privacy enforcement (at least in the US) confounded security with privacy; the majority of privacy cases referenced by the FTC in [17] are "privacy" cases that focus on failure to provide adequate security protections. Only one recent case addresses use limitations. The EU Data Protection Directive attempted to harmonize the protection of PII in the EU and created an externality that gave rise to the US Safe Harbor agreement. The Safe Harbor agreement requires most OSPs with European users implement a variant of the FIPs, but this is arguably and empty formalism [37]. More recently, the issue of behavioral targeting has become increasingly controversial, eliciting recommendations of guidelines for self-regulation by the FTC [17].

Although this concise history does constitute a feedback loop, it is slow and the power to contribute to the design of privacy tools resides almost exclusively in the hands of OSPs and more recently advertisers. As implied by this brief history, the normatively liberal self-regulatory motif has dominated: harms are identified, regulations are threatened and/or developed, and OSPs stave off further regulation by satisficing to the black letter of regulations. Abstract norms such as the FIPs garner consensus in intergovernmental forums, but stated alone do not provide sufficient guidance for domain-specific standards development processes. By satisficing to the FIPs through equally abstract privacy policies, OSPs avoid disclosing precise details of data collection processes and purposes. This strategy has preserved OSPs' power to control the shape and direction of privacy policies and tools regarding precisely how user information is used and shared.

There are a number of problems under this institutional arrangement. The first is that the feedback mechanism is slow, allowing abuses of a particular technology

to become institutionalized components of emerging user information economies. The longer a lucrative technology is available for use, the greater the resistance to regulation and the greater potential for displacement of the behavior to adjacent, unregulated technologies. As an instance of displacement, the tracking functions conventionally performed by browser cookies have been replicated in Flash cookies, which have the potential to store more data and are not yet as well-known a tracking mechanism. It is arguable that a tighter feedback loop would have surfaced the implications of mixing and the aggregate image, facilitating balanced regulation (either legislated or through industry standards) of these practices before they migrated to other technologies.

Another problem is that the locus of control is in the OSP, but, based on both policy and observed evidence, advertisers have fewer incentives not to abuse the potential for mixing and the aggregate image than OSPs. A number of factors contribute to this. First, OSPs do have an investment in establishing a trust relationship with users. Unlike third parties, OSPs are the primary actor associated with the environment. Second, the caveat that privacy guarantees by the OSP do not bind to third parties allows OSPs to distance themselves from egregious abuse by third parties without placing burden on these partners. Third, advertisers are only indirectly, if at all, accountable to users with regard to how they (advertisers) handle the aggregate images they construct via mixing and inferencing.

A third concern for users and privacy advocates is that users have limited control over the elements contributing to the aggregate image. This is in part linked to OSP control of the development of privacy tools. It is not in the interest of OSPs (or advertisers) to allow users substantive control over the content of the aggregate image because this may result in a much more sparse image than the currently incentivized objective of creating the most comprehensive ("relevant") image possible. OSPs and some network advertisers provide users with an interface to (de)select coarse-grained interests, but not direct control over the discrete elements that contribute to the aggregate image itself (Amazon being a limited exception).

Consider a quote from Microsoft's personalized advertising management tool:

> 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. [26]

Microsoft is singled out here as the only OSP in the sample to specifically make the distinction between display of advertisements and contributions to the aggregate image. The other OSPs do not make it clear whether deselecting interests remove this category from the underlying image or simply keep those advertisements from being shown. In the latter case, the aggregate image is just as comprehensive, if not more so, than before.

Facebook is another instance. Facebook allows users to "like" or "dislike" items using a thumbs up or thumbs down button attached to elements presented within the environment. It is not clear whether the categories of information disliked are removed from an aggregate image or simply flagged for use in other relevance decisions but not for use in displaying advertisements for that particular category.

It is unclear how interest management tools such as those described above affect the fidelity of the aggregate image. It is also not clear whether direct access to the "raw" elements that contribute to the aggregate image is the most efficacious interface for users or their proxies to manage the aggregate image, either. Simply having access to the raw image does not guarantee users will be either motivated to sift through it to filter out elements that they (a) do not want to be part of their image or (b) that they feel are distorting. Further, the process itself is daunting and invites rational ignorance, as with the discussion of evaluating third party applications in Section 3.2.1. Allowing the user access alone does not disincentivize the aggregator from collecting the same or similar data again. Rather, access coupled with reputation based disincentives are recommended. The collaborations that can facilitate such mechanism and a potential candidate approach are discussed in the next section.

Currently, there is a dearth of meaningful signals that convey the fact that OSP environments are mixed contexts, that convey information is flowing into or out of a context, or that convey the implications of these flows. Absent these signals, environments do not have the beneficial characteristics of experience goods, requiring an extensive search process[15] akin to the analysis in Section 3 or inside information akin to the resources available to media investigations [39], [38] to surface the types of information collected, the actors involved, and how user information flows amongst these actors. In either case, this process places an undue burden on the individual user. The outcomes of a collaborative standards construction processes should be to develop meaningful policies and attendant tools that inform users of the implications of mixed contexts and the aggregate image in a way that preserves the benefits of customization but also highlights the potentially harmful implications.

4.2 Collaborative Privacy Standards Setting

Given the deficiencies of the existing institutional arrangement, the collaborative forum intends to incentivize sharing the kinds of information sufficient to convey meaningful signals to users regarding the benefits and implications of mixing. As indicated in Section 3.4, this will require substantive information regarding the

15. In the sense of Nelson [28], described in an earlier note.

subjective preferences of users. The information serves to both develop an understanding of how mixed context can be applied and to provider users with meaningful signals and choices. This effort will require data collection and analysis by multiple interested actors, negotiations over development, and ultimately feedback on the efficacy of efforts as the process evolves. The transition recommended here builds on the objectives of a collaborative forum, presented in the next section. The strategic options and attendant tools are presented in Section 4.2.2.

4.2.1 Objectives of a Collaborative Forum

A key objective in the proces proposed here is to develop collaborative tools for collecting data about users' privacy choices, the types of contexts they would like to construct, and which tools and signals are most efficacious. This in itself has privacy implications and would require explicit consent 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.

The institutional arrangement suggested in the next section presents an opportunity to continuously surface common privacy motifs from empirical evidence of privacy choices made *in situ*. The first step to transitioning to this arrangement is to develop categories of contexts assumed or created by users and evaluate where they do and do not find mixing to be appropriate. For instance, users create groups of friends to share information with: collecting information about 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.) can contribute to better understanding of when mixing is appropriate and when integrity violating mixing is occurring. Understanding these categorizations may facilitate operationalizing these motifs into templates of common privacy configurations that illustrate compatible and incompatible mixing common to subsets of users[16]. Such templates could provide an approximation that can be further customized to match unique user preferences.

Data collection and the attendant strategies for minimizing actor transaction costs is distinct from developing a logic for checking whether contextual integrity has been violated. The two processes are complementary. Barth et. al. have developed a linear temporal logic for determining whether contextual integrity has been violated [6]. In contrast to such a logic, this work proposes strategies that facilitate (1) collaboration amongst the actors necessary to collect this data and (2) the use of such data as an empirical basis for the substantive categories of data, contexts, and user roles a logic such as Barth's would operate over.

16. Templates are not expected to fall out of the data; substantial mining of user choices will be necessary to identify commonalities.

4.2.2 Incentivizing a Collaborative Regime

Based on the existing institutional arrangement, OSPs do not have substantive incentives to collaborate with regulators any more than necessary to avoid direct regulation. Beyond abstract public relations articulations such as those in the privacy policies, OSPs and advertisers have little incentive to share additional information with CSOs that may play the role of educating users. Transitioning to a collaborative regime will require introducing incentives that draw attention to the mixed context problem in terms accessible to users as they interact with the environment. Ultimately, this will be best served by collaboration amongst OSPs, advertisers, CSOs, and developers that shifts burdens to the actor with the minimal transaction cost to contribute. Government regulators are proposed as the regulatory authority of last resort in the event of deadlock amongst actors. This serves a number of purposes. The government backstop acts as a baseline "bootstrap" incentive for OSPs and advertisers to participate. Considering the recent response to the threat of FTC regulation [10], [12], [20], [19] and the aversion for static regulation, it is argued this backstop will also incentivize ongoing participation. Government regulators will also serve as a backstop against the disintegration of the forum. Other rules will be necessary 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 standards should be a product of non-government actors (as those closest to the necessary data and with the motivation to engage in the process). Regulators are present to ensure progress is made by actors directly involved in the process toward substantive privacy standards, but regulators *do not* directly contribute to specific standards 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 find it in their interest to perform these tasks.

Returning to the transition process, the process of aligning actors in a way that moves them closer to cooperatively drawing attention to the mixed context problem can be presented as a menu of strategic options available to various combinations of actors. The mixed context problem can be surfaced by introducing unobtrusive signals into the environment that highlight when a context is mixed, what elements of the environment contribute to the mix, and what information is being mixed. This will entail (1) designing tools to begin development of measures that contribute to understanding "How mixed?" (2) developing a typology of meaningful data categories and contexts, (3) determining which actors

should develop the vocabulary for describing these to casual users, (4) determining which actor should translate these into requirements for the attendant privacy tools, and (5) which actor(s) should implement the tools for the collection and display of mixed context signals. These steps will require collaboration on designing tools to collect data, which actor(s) should analyze it, and a process of reconciling conflicts. These steps will require negotiations amongst actors over both the substantive categories and the granularity of the categories. The decisions on categories, in turn, will impact the design and implementation of tools for signaling mixed contexts have been encountered and the subsequent tools for integrating user responses into changes in the aggregate image.

As per the forum objectives, information about categories should be based on empirical data regarding what kinds of information users actually choose to share or keep private within and across contexts. One strategy for collecting the information necessary for operationalization 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 and exploratory analyses. As may be obvious, the collection of data for this sample is intrinsically private and will require explicit consent from users. Another factor in the data collection, directly related to burden, is the technical means for collecting this information and who will bear the burden of developing this means.

It is not expected that any of the individual categories of actors will be clamoring to take up this task. The criteria for decomposing the task and allocating subtasks should be based on both minimizing overall transaction costs and balancing strategic interests. Presenting the actors with a menu of options, each representing a potential set of trade-offs, may facilitate constructive conflict and negotiation around processes of collecting data, analyzing it, and developing standards.

One strategy for surfacing (collecting) 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. This may mean providing information based on existing databases of aggregate images coupled with information about what interests have been deselected as a starting point. Another strategy that reduces the burden on OSPs is for OSPs to provide an API for accessing (tacit) category information collected by the OSP during users' browsing sessions. Platform developers or third party browser extension developers (perhaps supported by civil society organizations) could build data collection tools for users to install in the process of participating in this experiment. Combinations of civil society organizations and OSPs 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. A commercial endeavor to reconstruct what advertisers think users's are interested is already being developed by Bynamite [24]. Yet another option is for regulators to sanction the data collection process and contract out the "experiment" to companies like Bynamite. 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. Scenario four is appealing because it is sanctioned and monitored by a legitimate government regulator. The fifth scenario is undesirable for a number of reasons: online service providers and advertisers will argue that it stifles innovation; it is a static solution that may not weather underlying technological changes; it is not empirically grounded and thus may be just as insufficient as vague, overly broad categories.

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, domain specific characterizations of contexts 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 give rise to a grounded (re)framing of data categories, (mixed) context, and the aggregate image.

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 potentially 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 based on understanding whether mixing is appropriate or not becomes a mechanism for eliciting actual preferences regarding privacy *and* advertising. In this sense, understanding and designing around a mixed context metaphor may develop into a shared concern that has valuable returns for OSPs, advertisers, and privacy advocates.

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 [11]. Based on the vague categories that characterize the policy sample [37], an immediate question is whether the current institutional arrangement is sufficient to generate such an icon with meaningful information necessary to make privacy decisions or whether this is a tactic to satisfice to recent FTC queries and stave off further regulation. Taken up in the forum described here, this type of signaling could be evolved into a mechanism for realigning incentives and introducing accountability to third parties. In addition to targeting information, mixed context information may also be embedded with this icon, providing meaningful measures of the sources of mixing, how this advertisement contributes to mixing, and other previously discussed measures of how mixed. 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. Building on understanding how mixed, 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 based on context and/or advertiser reputation for mixing. Following the theme of instrumenting these tools, data may be collected regarding how frequently these responses occur, on what grounds, and what the (inevitable) deficiencies of these tools are. For instance, users that choose to minimize customization rather optimize or correct their aggregate image may face penalties to incentivize participation necessary to preserve the advertising revenue model while also maintaining an accurate aggregate image.

Finding the appropriate balance in such an incentive structure would require ongoing analysis of how users react and the impact on OSP and advertising revenues. Ars Technica recently performed an experiment where they blocked users that used a certain advertising blocking extension [16]. According to the article, once Ars explained the dependence on advertising revenue, many users decided that Ars content was worth a few advertisements and added Ars as an exception to their advertisement blocking software. While this is an admittedly simple experiment, it does provide evidence that users may be willing to establish more sophisticated trust relationships with specific information stewards. 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. In particular, this may inform the earlier discussion of the utility of mixing within while disincentivizing segment-based inferencing that gives rise to incompatible mixing across.

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 to the user that the OSP may not be vetting the privacy policies of advertisers and networks effectively. Under such as scheme, Dictionary.com may have quite a red tint. A surfeit of red indicators damages the user's perception of that OSP. Taking the example a step further, civil society organizations may develop browser extensions that collect more precise statistics on the underlying measures of how mixed. Compiling this information and publishing it as a kind of consumer reports on OSP vetting practices may be a means to give OSPs reputation scores, as well. In effect, this can signal whether an OSP frequently accommodates advertisers or advertiser networks with poor privacy practices.

This strategy is not intended as a one time rating process that gives rise to adversarial relationships. Couched in a continuous re-evaluation process, OSPs should have the chance to improve their vetting process. OSPs may in turn monitor their scores to help them better vet advertisers and/or remove those whose policies have changed to the detriment of the OSPs reputation. Taken as a reputation indicator, this may actually help OSPs manage relationships with advertisers, balancing the power asymmetry between both OSP and advertiser and between OSP-advertiser structures and the user. For instance, to reduce the adversarial potential of such a rating system, OSPs may register with the actor that maintains OSP ratings and negotiate receiving messages indicating a sudden change in the ratings of advertisers it has relationships with before these ratings are published in a more public forum. This type of forewarning may allow the OSP to handle the problem between itself and the offending advertiser(s) rather than suffer the reputation blow. Demonstration of prompt response to misbehaving advertisers of data brokers may be incorporated as a positive factor in reputation of OSPs.

Under this scenario, the coloring scheme (or some other visual mechanism) provides meaningful signals to individuals. Such a mechanism contributes to resolving two of the deficiencies identified with OSP environments. First, it contributes to both the knowledge and experience with the environment and its formerly invisible constituents, allowing users to immediately investigate potentially conflicting elements as they see fit. Second, it allows users to make immediate decisions based on information about the privacy implications of a particular environment, reducing the cost of search-based evaluation. With regard to a CSO supported browser extension, this helps resolve certain aspects of collective action problems and is a strategy that may better incentivize genuine cooperation by OSPs and advertisers. Such approaches would also have the effect of incentivizing *OSPs* 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 access to user information. In either case, the reputation incentive acts on OSPs, the category of actor that arguably has the closest relationship with advertisers and thus perhaps the most influence.

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 location-based advertisements and supporting technologies continue to mature and are 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. As new technologies emerge, it is argued this forum can help identify the related implications for mixing before harmful practices become institutionalized.

5 Conclusions and Future Work

Mixed context has been presented as an alternative to existing framings of OSP privacy practices. Application of mixed context has focused attention on cross context information flows rather than appealing to existing metaphors of control (rooted in security) or value-laden metaphors rooted in spying or surveillance. The conceptual roots of mixed context, built upon Nissenbaum's contextual integrity framework, were extracted completely from the *service-and-utility* framing of a sample of ten large OSP's privacy policies. Recent investigations by the media provide confirmatory instances of the outcomes extrapolated from conceptual evidence in the privacy policy. Mixed context has provided insights into the kinds of tools that may better convey data sharing implications to users and the institutional relationships that give rise to the harms identified in this work.

As implied earlier, the claim of the generative metaphor is not that it identifies novel behaviors that had been previously undocumented. While the notion of mixed context is a novel application of contextual integrity (to the author's knowledge), the application of mixed context "selects for attention" the process of mixing, the tools that facilitate mixing, and the actor incentives that perpetuate mixing that were arguably already known. In this sense, mixed context has focused the designer of privacy policies and tools on what has been argued here as the root of the problem. As per the evidence developed in Sections 3 and 4, the key roots of the problem are the mutually reinforcing architectures that hide mixing processes and the incentive structures that characterize the existing arrangements between OSPs and third party advertisers. This is not intended to vilify either party; rather, unlike value-laden metaphors, mixed context attempts to reconcile actors interests in an effort to help identify means to balance benefits and harms of mixing information from different contexts.

As implied by Section 3.4 and the collaborative data collection processes recommended, there is still quite a bit of work to be done to understand the dynamics of mixed context problems. Efforts such as those by Bynamite are encouraging, indicating that some of the transition strategies recommended in Section 4.2.2 are already emerging. Future work on mixed context will focus on quantifying the concepts developed here. One research direction will focus on designing the policy experiments necessary to better understand how to collect representations of the complex of subjective user

preferences and context. Another direction will focus on understanding how mixed an environment is by developing web robots to collect statistics on the types of advertisements and sources of mixing.

REFERENCES

[1] A. Acquisti. Protecting Privacy with Economics: Economic Incentives for Preventive Technologies in Ubiquitous Computing Environments. Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.58.4845.

[2] ——, "Privacy in Electronic Commerce and the Economics of Immediate Gratification," in *Proceedings of the 5th ACM conference on Electronic commerce*, 2004, p. 29.

[3] A. Acquisti and J. Grossklags, "Privacy and Rationality in Individual Decision Making," *IEEE Security & Privacy*, vol. 3, no. 1, pp. 26–33, 2005.

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

[5] J. Angwin and T. McGinty, "Sites Feed Personal Details To New Tracking Industry," *wsj.com*, Jul. 2010, retrieved fom http://online.wsj.com/article/SB10001424052748703977004575393173432219064.html on 10 August 2010. [Online]. Available: http://online.wsj.com/article/SB10001424052748703977004575393173432219064.html

[6] A. Barth, A. Datta, J. Mitchell, and H. Nissenbaum, "Privacy and Contextual Integrity: Framework and Applications," in *2006 IEEE Symposium on Security and Privacy (S&P'06)*, Berkeley, CA, USA, 2006, pp. 184–198. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1624011

[7] S. Bellman, E. Johnson, S. Kobrin, and G. Lohse, "International differences in information privacy concerns: a global survey of consumers," *The Information Society*, vol. 20, no. 5, pp. 313–324, 2004.

[8] B. Berendt, O. Günther, and S. Spiekermann, "Privacy in e-commerce: stated preferences vs. actual behavior," *Commun. ACM*, vol. 48, no. 4, pp. 101–106, 2005.

[9] R. Chellappa and R. Sin, "Personalization versus privacy: An empirical examination of the online consumer's dilemma," *Information Technology and Management*, vol. 6, no. 2, pp. 181–202, 2005.

[10] S. Clifford, "Industry Tightens Its Standards for Tracking Web Surfers," *The New York Times*, Jul. 2009, retrieved fom http://www.nytimes.com/2009/07/02/business/media/02adco.html on 5 February 2010. Published 1 July 2009. [Online]. Available: http://www.nytimes.com/2009/07/02/business/media/02adco.html?_r=1

[11] ——, "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. [Online]. Available: http://www.nytimes.com/2010/01/27/business/media/27adco.html

[12] ——, "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. [Online]. Available: http://mediadecoder.blogs.nytimes.com/2010/01/11/ftc-has-internet-gone-beyond-privacy-policies/

[13] J. Dewey, *Logic - The Theory of Inquiry*. Henry Holt and Company, Nov. 1938.

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

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

[16] K. 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. [Online]. Available: http://arstechnica.com/business/news/2010/03/why-ad-blocking-is-devastating-to-the-sites-you-love.ars

[17] FTC. FTC Staff Revises Online Behavioral Advertising Principles. [Online]. Available: http://www.ftc.gov/opa/2009/02/behavad.shtm

[18] Google.com. Ads in GMail. Retrieved from http://mail.google.com/support/bin/answer.py?hl=en&answer=6603, August 2010.

[19] A. Greenberg, "Congress Mulls Online Privacy Law," *Forbes.com*, 2009, retrieved fom http://www.forbes.com/2009/06/18/google-yahoo-privacy-technology-advertising.html. [Online]. Available: http://www.forbes.com/2009/06/18/google-yahoo-privacy-technology-advertising.html

[20] ——, "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. [Online]. Available: http://www.forbes.com/2009/06/17/online-advertising-privacy-technology-security-congress.html

[21] I. Hann, K. Hui, T. Lee, and I. Png, "Online information privacy: Measuring the cost-benefit trade-off," in *23rd International Conference on Information Systems*, 2002.

[22] C. Jensen, C. Potts, and C. Jensen, "Privacy practices of internet users: Self-reports versus observed behavior," *International Journal of Human-Computer Studies*, vol. 63, no. 1-2, pp. 203–227, 2005.

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

[24] S. Lohr, "You Want My Personal Data? Reward Me for It," *The New York Times*, Jul. 2010, retrieved from http://www.nytimes.com/2010/07/18/business/18unboxed.html on 17 July. [Online]. Available: http://www.nytimes.com/2010/07/18/business/18unboxed.html?_r=1

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

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

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

[28] P. Nelson, "Information and Consumer Behavior," *The Journal of Political Economy*, vol. 78, no. 2, pp. 311–329, Apr. 1970, ArticleType: primary_article / Full publication date: Mar. - Apr., 1970 / Copyright © 1970 The University of Chicago Press. [Online]. Available: http://www.jstor.org.libproxy.mit.edu/stable/1830691

[29] H. Nissenbaum, "Privacy as contextual integrity," *Washington Law Review*, vol. 79, no. 1, 2004.

[30] D. North, *Institutions, Institutional Change, and Economic Performance*, ser. Political Economy of Institutions and Decisions. Cambridge University Press, 1990.

[31] M. Olson, *The Rise and Decline of Nations*. New Haven, CT: Yale University Pres, 1982.

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

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

[34] D. A. Schön, "Generative Metaphor: A Perspective on Problem-setting in Social Policy," in *Metaphor and Thought*, A. Ortony, Ed. Cambridge University Press, 1993, ch. 9.

[35] H. A. Simon, *Models of Bounded Rationality*. MIT Press, 1982.

[36] D. Solove, "Identity theft, privacy, and the architecture of vulnerability," *Hastings Law Journal*, vol. 54, p. 1227, 2003. [Online]. Available: http://ssrn.com/paper=416740

[37] J. Sowell, "Deficiencies in Online Privacy Policies: Factors and Policy Recommendations," Master's thesis, Massachusetts Institute of Technology, September 2010.

[38] S. Vanek-Smith, "Hey Baby, What's Your Cluster?" *Marketplace*, retrieved fom http://www.publicradio.org/columns/marketplace/business-news-briefs/2010/07/hey_baby_whats_your_cluster.html on 29 July 2010. [Online]. Available: http://www.publicradio.org/columns/marketplace/business-news-briefs/2010/07/hey_baby_whats_your_cluster.html

[39] wsj.com. (2010, Aug.) What They Know Series. Retrieved fom http://online.wsj.com/public/page/what-they-know-digital-privacy.html on 11 August 2010. [Online]. Available: http://online.wsj.com/public/page/what-they-know-digital-privacy.html

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