Comments on "Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers"

Vasant Dhar[1], Jessy Hsieh[2], Arun Sundararajan[3]

New York University, Leonard N. Stern School of Business[4]

February 18, 2011

The purpose of this document is to respond to selected questions for comment on the proposed framework in the FTC report "Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers" (December 1st, 2010). Our responses are based on our ongoing research about online privacy and data risk at NYU Stern School's Center for Digital Economy Research. Our findings are described further in Dhar, Hsieh and Sundararajan (2011).

These are the specific questions proposed for comment that our response addresses.

1. **Are there practices that should be considered "commonly accepted" in some business contexts but not others?**
2. **Under what circumstances (if any) is it appropriate to offer choice as a "take it or leave it" proposition, whereby a consumer's use of a website, product, or service constitutes consent to the company's information practices?**
3. **What types of disclosures and consent mechanisms would be most effective to inform consumers about the trade-offs they make when they share their data in exchange for services?**
4. **Should access to data differ for consumer-facing and non-consumer facing entities?**
5. **Should consumers receive notice when data about them has been used to deny them benefits? How should such notice be provided? What are the costs and benefits of providing such disclosure?**

[1] Professor and Daniel P. Paduano Faculty Fellow. vdhar@stern.nyu.edu
[2] Research Scientist. jhsieh@stern.nyu.edu
[3] Associate Professor and NEC Faculty Fellow. arun@stern.nyu.edu
[4] We are submitting these comments on our own behalf. The views expressed here should not be interpreted as representing an official view of New York University, the Stern School of Business, or any of its affiliated entities.

Summary and key points

- There is *privacy risk* inherent in any form of consumer data acquisition, retention and use by a firm. Some parts of this risk are *non-systemic* and can be reduced by firm and/or policy action, while other parts are *systemic* and cannot be altered without widespread changes in social norms.
- Our framework for examining the non-systemic privacy risk associated with the acquisition, retention and use of consumer data is based on the relative extent to which the acquiring party (typically a firm or government) and the providing party (typically a consumer) perceive that they *own* the data in question, which in turn is based on the *intention* of the consumer.
- Much like intellectual property, data is non-rival, and in the absence of appropriate technological or legal controls, is non-excludable. Any assessment of ownership must be based on some notion of the *division of rights* between a data provider (consumer) and a data acquirer (firm).
- Currently, both the data acquirer and the data provider seem to have *equal* and *comprehensive* rights over the exchanged data (in principle) independent of the context of data exchange. By default, this status quo gives excessive ownership rights to the acquirer (the firm) and generates excessive non-systemic data risk for both parties.
- Consumer perception of data ownership is associated with the *intent* of the consumer when transferring data to the firm. Specifically, the consumer inherently presumes a *lower* granting of rights to the firm if the consumer *did not intend* to transfer the data in exchange for a service by the firm.
- Our framework for classifying consumer intent during data transfer identifies two dimensions that can be used to assess perceived ownership: whether the data transfer was *explicit* or *implicit*, and whether the data transfer was *voluntary* or *required* for service provision.
- Specifying the intent associated with consumer data transfer on the Internet is complex because some kind of data transfer is *necessary* for *every action* during an electronic interaction, irrespective of whether the action explicitly involved intentional data transfer. This is because a consumer may be unaware that data not core to the interaction are being transferred.
- The *distinction* between explicit and implicit data transfer can be partially disambiguated by examining the corresponding actions that would need to be taken by the firm and the consumer if such a transfer was occurring in a non-electronic (or “physical world”) setting. Physical world norms are a useful benchmark because they have evolved over a long time and provide a baseline for gauging intent.

Discussion of the main points

Consumers generate a tremendous and growing volume of data as a by-product of their electronic interaction with firms and with each other. Furthermore, this gathering of consumer data and its associated risks are growing exponentially as the number of electronic touch points (mobile devices, social media, and payment systems) increases. A number of firms use such data for activities as diverse as production planning, marketing and new product development. The more electronic the spaces of interaction between agents, and the more this interaction involves the transaction of digital goods, the richer are the data trails created, and the greater the possibility of their intelligent interpretation and summarization in real time. What is knowable about individuals, business, and society increases tremendously as such data become available (Dhar and Sundararajan, 2007). As a consequence, firms frequently acquire and retain consumer data based on a belief that this data will provide them with future business returns, but without sufficient assessment of the risks they expose themselves to by retaining and/or using the data. In our prior and current research, we develop frameworks for the assessment of such risk (Dhar and Sundararajan, 2010; Dhar, Hsieh and Sundararajan, 2011).

What is particularly alarming is the potential for *connecting* these various streams of data in ways where identities of individuals can be inferred without their knowledge or consent (Ohm, 2010; Acquisti, 2009). We refer to this type of risk as *systemic risk* in that it is driven by advances in information technologies and is inherent in their use. In other words, the world continues to become less private in general for anyone who engages in the normal use of current technology. While such risks might be altered globally they can require changes to current norms or processes, some of which have been in place for decades, such as how social security numbers are generated (Acquisti, 2009).

A more pragmatic goal for the next decade would be to focus on addressing the significant *non-systemic* risks, those that are controllable by intelligent firm and government policy. It is this class of risks that we focus on in our research and in this response. For example, a cell phone service provider who tracks the location of its users and sells that information to retailers is taking on significant risk of invasion of privacy or causing potential harm to its customers[5]. This risk can be altered by firm policy (the provider can reduce their risk by choosing not to sell the data, and reduce it even further by choosing not to even record and store this data electronically, thereby insulating itself from the negative fallout of an unintended breach of its databases) or by public policy action (a government could reduce this risk by making such data sales illegal).

A significant fraction of the risk discussed above is associated with perceived violations of consumer privacy. We believe that a discussion of privacy issues associated with electronic consumer data should therefore begin with some notion of who *owns* this data.

[5] A physical world analogy of this would be a service provider employing private detectives to follow its customers and record their every move, and subsequently selling the consolidated data to retailers. Clearly, this is not practical in the physical world due to the enormous transaction costs involved, but if it were somehow possible, it might even be considered akin to stalking.

Consider the following extreme scenarios. One policy might be to allocate complete and exclusive ownership of consumer data to the consumer in question, who is, after all the "creator" of the data. Such a policy might mandate that a firm who wishes to use this data for any purpose must license specific rights to using this data from the consumer. While conceptually appealing, the transaction costs associated with such a policy are sufficiently onerous to render it non-viable[6]. Another policy might be to grant complete and exclusive ownership of any transferred consumer data to the acquiror of the data (in most cases, a firm). Under this extreme scenario, a consumer cedes all rights to ownership and usage of their data upon transfer, and must "license" it back from the firm if they wish to use it in any way.

While the latter scenario seems as biased as the former, it is distressingly close to today's reality. Currently, being on a firm's website is akin to being "on their property," implicitly giving the firm ownership of the data. While the consumer jointly owns this data in principle, the reality is quite different. For example, Amazon owns the browsing and purchase history of their consumers, and while it shares a consumer's purchase history on request, consumers still require Amazon's cooperation in receiving this data. It is impossible for a user to get access to their browsing history from Amazon. Google owns its consumers' search and subsequent browsing/click-through history on Google's web site. An individual can choose to be able to view his or her history of search queries with Google's cooperation, but cannot as of today access their subsequent browsing history[7]. A Facebook user can access their interactions with any of their friends that involved explicitly taking an action (a post, a comment, a "like"), but cannot access the history of their browsing activities. Facebook, on the other hand, owns all the data about each user's implicit and explicit interactions on their web site.

That consumer data ownership is joint in principle is partially on account of data being *non-rival*, wherein use by one party does not prevent simultaneous use by another. In principle, such data is also *non-excludable*, but unless consumers invest in meticulous and expensive tracking of their own electronic behaviors, a firm can prevent a consumer from accessing their own data, thus making such data excludable in practice. These properties of data suggest that one might consider treating its ownership in the same way as akin to that of *intellectual property* (IP). This parallel has been drawn many times in past writing (see for example, Samuelson, 2000; Ipeirotis, 2008).

Rather than proposing a framework for IP-like treatment of data ownership, our goal in revisiting this parallel is more modest. We argue that there are two principles underlying IP law that can inform how a firm might treat consumer data. First, IP law is pragmatic, in that it does not specify a division of rights that is *technologically infeasible* to enforce. In the past, for example, law that specified a buyer of a physical book only be able to read the book once would be senseless since the "technology" of printed

[6] Besides, one might argue that data generated by a consumer's use of a firm's web site might be more appropriately considered joint product.

[7] In theory, a consumer might record each click and the associated time spent by merging their browsing histories across their web browsers and maintaining a database of this over time. However, we have no evidence that this practice is prevalent, so in reality, the ownership is with the firm.

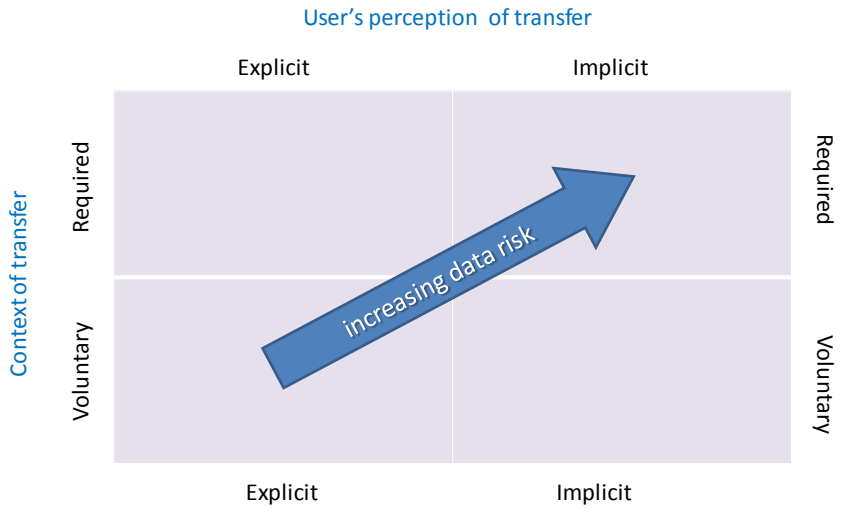

Figure 1

books makes this impossible to reasonably monitor[8]. Second, and perhaps more important, the transfer of rights from the creator of IP to its "acquirer" seems to map to the intent of the creator when authorizing the transfer. For example, an artist who sells a song to a listener intends to allow the listener to hear the song whenever they want, but does not intend to let them resell multiple copies of the song. Both of these intentions are reflected in copyright law.

In the absence of a practical and widespread set of laws governing the ownership of consumer data and the rights associated with consumer data transfer, we therefore believe that the "intent" of the creator of the data (the consumer) is a good basis on which to consider alternative policy actions.

In Dhar, Hsieh and Sundararajan (2011), we propose a data taxonomy that is based on user intent. Our goal is to provide a framework that reduces non-systematic risk by classifying data into categories and associating *intent* with each category of data. The taxonomy is illustrated in Figure 1 and consists of two dimensions of the user's perception of the exchange of data:

1. Whether the data are *required* for the transaction or whether their provision was *voluntary.* For example, buying a good online and having it shipped home isn't possible without providing a physical address. Nor is it possible with providing payment information. But it does not require providing gender information, which may nevertheless be collected by the merchant and voluntarily provided by the consumer.
2. Whether the user was *explicitly* aware of the data transfer during the interaction or whether the data exchange took place without the user being aware of it, that is, *implicitly.*

[8] Unfortunately, this law is circumvented for digital goods by the creative use of licensing contracts and digital rights management software, often to the detriment of consumers, and potentially to the detriment of firms. For an example of how excessive control over rights using DRM software can lower a firm's profits, see Sundararajan (2004).

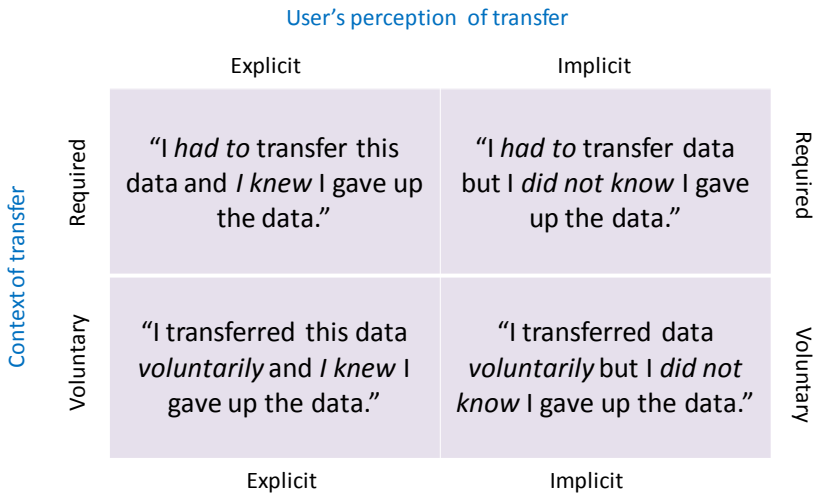

Figure 2

Consider an interaction between a consumer (data provider) and an online retailer (data acquirer) of DVD's. A data set including {title of movie disk} and {credit card number} is likely to be perceived by the consumer as both *required* and *explicitly* given. By clicking "purchase" next to an icon of the specific movie title and typing a credit card number into an online form, the consumer transfers data that is simultaneously required for the fulfillment of a purchase and explicitly provided by the consumer. In a similarly explicit way, a consumer may fill out an online review for a certain movie title. The data transferred in this interaction, for example {title of movie is a five star movie}, was given voluntarily by the consumer, since it was not asked for or required in exchange for a service. An example of data being "required" to complete a transaction without a user being aware could be in mobile communication wherein a user is communicating with others without being aware that data related to the location of their mobile device is being transferred to the carrier by its base stations. The latter data are required for the exchange to be possible, but something that the user is not cognizant about.

Examples of data transfer that are voluntary (in contrast with data required for the completion of a transaction) are increasingly common as the use of social media becomes more widespread. Users voluntarily provide data about themselves, for example, on Facebook, and are aware of this transfer. They usually expect something in exchange for providing this data, such as comments from other users or personal satisfaction. Finally, they may not be aware about providing information about themselves which isn't even required for the exchange.

While the contrast between required and voluntary seems quite clear, applying our framework to a variety of examples has led to the realization that there is often ambiguity about what might actually be perceived as an "explicit" transfer by a consumer, and correspondingly, what is implicit. Some cases are clear: for example, when a user fills out a form and clicks on a "submit" button, their intent to transfer the data they provided was clearly explicit. Clicking on an "I agree" button is arguably explicit although this often depends on context (Bakos, Marotta-Wurgler and Trossen, 2009). Similarly, if data is collected

by installing tracking software on a user's computer without their knowledge, this clearly represents an implicit transfer. In contrast, a user's browsing history during a session at a specific web site is presents ambiguity. Since the user took an action (clicking on a link, typing in a URL) for each web site they visited, it might be argued that the transfer was explicitly intended. However, one might also argue that the true intent of the user was simply to "visit" the web site the way one might visit a store, rather than to transfer any data whatsoever.

More generally, specifying the intent associated with consumer data transfer on the Internet is complex because some kind of data transfer is *necessary* for *every action* during an electronic interaction, irrespective of whether the action explicitly involved intentional data transfer. This is largely because a consumer may be unaware that data not core to the interaction are being transferred. We believe that this can be partially disambiguated by examining the corresponding actions that would need to be taken by the firm and the consumer if such a transfer was occurring in a non-electronic (or "physical world") setting.

To understand how this analogy might be useful, consider again the interaction between a consumer and an online retailer. A user may sign up for an account with the online retailer. Each time she makes a purchase with that online retailer, the consumer signs into her account. A result of these actions could be the data {Jane Doe purchased titles X,Y,Z}. Constructing an analogous data set in the physical would require the use of a frequent buyer or club card. With each swipe that the consumer makes before a purchase, the consumer is likely to know that her purchases are being recorded, and thus, it seems reasonable to classify this as an explicit data transfer.

Returning to the case of browsing histories, contrast this with the digital world actions that lead to the data {Jane Doe browsed titles X,Y,Z} . True, typing in a web site address or clicking on a link to it may be explicit acts, but the user's intent to transfer this specific data was not entirely so clear. Jane Doe clicked on the icons next to certain titles in search of movies she may be interested in, but she may be unaware that a server recalled her IP address or retrieved a cookie and recorded each title she viewed. The corresponding physical world actions that might generate this data would be considered quite unreasonable: they might require a salesperson to view the consumer's driver's license without their knowledge to establish identity, then to follow a consumer around the store and note down each title that was picked up by the consumer. Clearly, this would not correspond to an explicit or intended transfer of data from the consumer to the firm, and our argument based on the physical world analogy would therefore be that browsing history should be considered implicit rather than explicit.

Physical world norms are a useful benchmark because they have been refined over a long time and represent "reasonable" norms in society in general. While the online world is different in many obvious respects, when consumer intent is ambiguous online, the physical world provides a useful baseline for gauging this intent.

To summarize briefly, we recommend a focus on *non-systemic privacy risk* rather than systemic privacy risk, and posit that risk depends on a notion of data ownership which need not be explicitly specified as a division of rights, but can be based on an assessment of consumer *intent* to transfer the data. Our

framework classifies intent based on whether the data transfer was *explicit* or *implicit*, and whether the data transfer was *voluntary* or *required* for service provision, and uses physical world analogies to disambiguate settings in which "intent" in the digital world may not be clear.

References

1. Acquisti, Alessandro, and Gross, Ralph, 2009. Predicting Social Security Numbers from Public Data. *Proceedings of the National Academy of Sciences* 106 (27), 10975-10980.
2. Bakos, Yannis, Marotta-Wurgler, Florencia and Trossen, David, 2009. Does Anyone Read the Fine Print? Testing a Law and Economics Approach to Standard Form Contracts. NYU Law and Economics Research Paper No. 09-40. Available at *http://ssrn.com/abstract=1443256*
3. Dhar, Vasant and Sundararajan, Arun, 2007. Information Technologies in Business: A Blueprint for Education and Research. *Information Systems Research* 18, 125-141.
4. Dhar, Vasant and Sundararajan, Arun, 2010. Data as a Liability? A Risk-Based Model for Data Governance. Working Paper CeDER 10-06, Center for Digital Economy Research.
5. Dhar, Vasant, Hsieh, Jessy and Sundararajan, Arun, 2011. A Risk-Based Taxonomy of Online Data from a Consumer Privacy Perspective, Working Paper CeDER 11-01, Center for Digital Economy Research.
6. Ipeirotis, Panos, 2008. Privacy and Intellectual Property: Two Sides of the Same Coin? *http://behind-the-enemy-lines.blogspot.com/2008/02/privacy-and-intellectual-property-two.html*
7. Ohm, Paul, 2010. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review* 57, 1701-1777.
8. Samuelson, Pamela, 2000. Privacy as Intellectual Property. *Stanford Law Review* 52 (5), 1125-1173.
9. Sundararajan, Arun, 2004. Managing Digital Piracy: Pricing and Protection. *Information Systems Research* 15 (3), 287-308.