COLORADO LAW

UNIVERSITY OF COLORADO LAW SCHOOL

June 30, 2010

Public Comment to the Federal Trade Commission:

Re: COPPA Rule Review P104503

Submitted electronically via website

INTRODUCTION

I appreciate the opportunity to submit these comments to the Federal Trade Commission ("FTC" or "Commission") regarding its implementation of the Children's Online Privacy Protection Act ("COPPA" or "the Act") through the Children's Online Privacy Protection Rule ("COPPA Rule"), in response to the request for public comments published in the *Federal Register* on April 5, 2010.[1] I am a law professor at the University of Colorado Law School in Boulder, Colorado. At the invitation of the FTC staff, on June 2, 2010, I participated in the roundtable relating to this request for public comment, and I write now to elaborate and expand upon what I and others said at the roundtable.

In these comments, I draw on several areas of my experience and expertise. First, I specialize in information privacy and cyberlaw, having authored more than a dozen law review articles and essays dealing directly or tangentially with issues that the FTC is now considering.[2] Second, I have a rich understanding of computer science and information technology drawn from undergraduate degrees in computer science and electrical engineering and from several years working as a computer programmer, network systems administrator, and IT specialist. Third, I served for four years as a federal prosecutor for the U.S. Department of Justice's Computer Crime and Intellectual Property Section in Washington D.C., where I specialized in the search and seizure of information on computers and computer networks. Finally, I have recently written a law review article surveying data anonymization and reidentification,[3] work that inspired these comments.

---

[1] 75 Fed. Reg. 17,089 (April 5, 2010)

[2] For a complete list of my publications, *see* http://paulohm.com/scholarship.shtml.

[3] Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. ___ (forthcoming 2010), draft available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006.

As I did at the roundtable, I focus now on the definition of the phrase "personal information." I argue that the FTC should define this phrase in light of recent advances in computer science that call into question our faith in the ability of data anonymization to protect privacy. These advances have rendered underinclusive the traditional regulatory approach of creating lists of categories of information that raise privacy risks. Although the FTC should continue to make lists like these, I argue it should also embrace a new, less rigid approach, one better suited to modern privacy concerns.

Specifically, I urge the FTC to define "personal information" in the COPPA Rule, in part, to mean:

> Personal information means individually identifiable information about an individual collected online including: . . . (h) any collection of more than twenty-five distinct categories of information about a user.

I describe, justify, and discuss the impact of this proposal in three Parts. In Part I, I survey the recent advances in computer science that call into question the ability of data anonymization to protect privacy, and I explain how these advances should alter the way the FTC regulates privacy. In Part II, I explain the proposed new definition of personal information and examine some of the implementation details and impacts. Finally, in the Legal Appendix, I construe COPPA, concluding that Congress meant to give the FTC the breadth of authority for defining personal information necessary to accommodate my recommendation.

I. THE FAILURE OF ANONYMIZATION

COPPA defines "personal information" as "individually identifiable information about an individual collected online," listing five specific examples: "a first and last name; a home and other physical address including street name and name of a city or town; an e-mail address; a telephone number; [and] a Social Security number."[4] This part of COPPA, like many other privacy laws, rewards companies that engage in what is known as data anonymization, the filtering of databases to remove only what the law deems to be harmful data while leaving behind supposedly benign data, which the database owner may continue to exploit for advantage or profit. COPPA reflects a pervasively held faith in anonymization, one shared by every privacy law ever written, in the U.S. and abroad.[5]

The problem is that recent developments in computer science have caused us to rethink our faith in anonymization and thus our faith in laws like COPPA. Computer scientists have begun to demonstrate that even after we anonymize data, what we leave behind can often be used to "reidentify" individuals.

---

[4] 15 U.S.C. § 6501(8)(A – E) (2006). The FTC is empowered to add new identifiers to this list. 15 U.S.C. § 6501(8)(F). I explore this power in depth in the Legal Appendix.

[5] Ohm, *supra* note 3, at 33-34 ("In addition to HIPAA and the EU Data Protection Directive, almost every single privacy statute and regulation ever written in the U.S. and EU embraces—implicitly or explicitly, pervasively or only incidentally—the assumption that anonymization protects privacy.").

A. A PRIMER ON THE NEW SCIENCE OF REIDENTIFICATION[6]

Anonymization works by removing data that supposedly can be tied to identity while leaving behind data that supposedly cannot. On the contrary, computer scientists have repeatedly discovered pockets of surprising uniqueness in the data that anonymization leaves behind. Just as human fingerprints left at a crime scene can uniquely identify a single person and link that person with supposedly anonymous information, so too do data subjects generate data fingerprints—combinations of values of data shared by nobody else in the database. And researchers have found these data fingerprints in supposedly-anonymized data with much greater ease and much more quickly than even experts would have predicted. Consider some of the surprising results.

How many other people in the United States share your specific combination of ZIP code, birth date (including year), and sex? According to a landmark study, for 87 percent of the American population, the answer is zero; these three pieces of information uniquely identify each of them.[7] How many users of the Netflix movie rental service can be uniquely identified by when and how they rated any three of the movies they have rented? According to another important study, a person with this knowledge can identify more than 80 percent of Netflix users.[8] Prior to these studies, nobody would have classified ZIP code, birth date, sex, or movie ratings as personally identifying. As a result, even after these studies, companies have disclosed this kind of information connected to sensitive data in supposedly anonymized databases with impunity.

Consider another reidentification study that bears more directly on COPPA, the Electronic Frontier Foundation's ("EFF") Panopticlick study.[9] EFF researchers demonstrated how easily websites can track the behavior or users across repeat visits even when they are forced to ignore traditional tracking tools like cookies and IP addresses. They can do this by tracking the fingerprints of their users' web

---

[6] In this Subpart, I provide a very brief overview of the computer science research. For much more detail, please consult my forthcoming article. Ohm, *supra* note 3, at 15-25, pt. I.B.

[7] Latanya Sweeney, *Uniqueness of Simple Demographics in the U.S. Population*, (Laboratory for Int'l Data Privacy, Working Paper LIDAP-WP4, 2000). More recently, Philippe Golle revisited Dr. Sweeney's study, and recalculated the statistics based on year 2000 census data. Dr. Golle could not replicate the earlier 87 percent statistic, but he did calculate that 61 percent of the population in 1990 and 63 percent in 2000 were uniquely identified by ZIP, birth date, and sex. Philippe Golle, *Revisiting the Uniqueness of Simple Demographics in the US Population*, 5 ACM WORKSHOP ON PRIVACY IN THE ELEC. SOC'Y 77, 78 (2006).

[8] Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets*, PROC. OF THE 2008 IEEE SYMP. ON SECURITY AND PRIVACY 111, 121.

[9] The study itself involved a website, Electronic Frontier Foundation ("EFF"), Panopticlick, http://panopticlick.eff.org/ (last visited June 29, 2010). The EFF published the results from the study at Peter Eckersly, *How Unique is Your Web Browser?*, http://panopticlick.eff.org/browser-uniqueness.pdf (May 17, 2010). The EFF researchers cite and build upon the earlier work of Jonathan Mayer. Jonathan Mayer, "Any person . . . a pamphleteer": Internet Anonymity in the Age of Web 2.0, Undergraduate Senior Thesis, Princeton University (2009), available at http://www.stanford.edu/~jmayer/papers/thesis09.pdf.

browsers derived from the dozens of characteristics a web browser reveals whenever a website asks.[10]

According to the study, browsers reveal data that uniquely identify them with startling frequency.[11] The researchers asked people to visit a website that displayed the vast collection of information it could coax their browsers to reveal as well as the number of prior visitors whose browsers had presented the same fingerprint.[12] After running the website for twenty days, the researchers analyzed all of the fingerprints for trackability. The researchers could uniquely identify 83.6 percent of the more than 400,000 browsers that visited the site by their browser fingerprints alone and they could narrow down another 8.2 percent of the visitors to a pool of fewer than ten browsers each bearing the same signature.[13]

The Panopticlick results reveal a surprising betrayer of identity: A user's font collection.[14] Many browsers with Javascript and Flash or Java installed will reveal a list of all fonts installed on a computer when asked.[15] As it happens, our computers' font collections act almost like Social Security numbers, uniquely identifying our computers by reflecting the idiosyncratic choices we have made about which computer programs to install. Websites that can track our fonts can also often track our identity and activity, again without cookies or IP addresses.[16]

These results may defy intuitions, but they mark a trend not a quirk. We have only just begun to understand the many ways that websites can track us. Already, vendors are offering web browser fingerprinting services,[17] perhaps catering to customers unable to use cookies because of privacy laws like COPPA.[18] If our privacy laws and regulations are to keep up with the new power of reidentification and web tracking, we need a new approach.

B. How the Power of Reidentification Changes Privacy Law

COPPA is an example of what is called the personally identifiable information ("PII") approach to statutory privacy, which tries to protect privacy through the categorization of data. Lawmakers adopting this approach make lists of specific types of data they conclude can hurt privacy, implicitly deciding that data not

---

[10] Eckersley, *supra* note 9, at 3-4. Web browsers reveal all of this information for benign, indeed useful, reasons; it enables the customized web. It is generally a good thing that the New York Times website knows that my screen measures 768 pixels across and the Cambria font can be found in my fonts folder.

[11] Eckersley, *supra* note 9, at 9.

[12] EFF, *supra* note 9.

[13] Eckersley, *supra* note 9, at 2.

[14] *Id.* at 9, 11 fig. 3 (showing identifying power of specific browser measurements and concluding that "plugins and fonts are the most identifying metrics"). The other characteristic that revealed the most identity was the collection of browser plugins. *Id.*

[15] *Id.* at 5 tbl 1.

[16] *Id.*

[17] *Id.* at 2 ("There are several companies that sell products which purport to fingerprint web browsers in some manner and there are anecdotal reports that these prints are being both used for analytics and second-layer authentication purposes."); *see also id.* at 6 n.4.

[18] 16 C.F.R. § 312.2 (defining "personal information" to include "A persistent identifier, such as a customer number held in a cookie or a processor serial number, where such identifier is associated with individually identifiable information").

appearing on the lists is less likely (or unlikely) to hurt privacy. By providing a complete safe harbor for those who purge listed data, COPPA encourages companies to engage in the anonymization of a narrow and static list of data categories.

The power of reidentification and failure of anonymization disrupt the PII approach. The PII approach assumes that lawmakers can evaluate the inherent riskiness of data categories, assessing with mathematical precision whether or not a particular data field contributes to the problem enough to be regulated. Now we understand that PII is an ever-expanding category. Ten years ago, almost nobody would have categorized movie ratings as PII, and as a result, no law or regulation did either. Today, four years after computer scientists exposed the power of movie ratings data to identify, no law or regulation yet treats them as PII.

Easy reidentification makes PII-focused laws like COPPA underprotective by exposing the arbitrariness of their intricate categorization and line drawing. Although the 1999 COPPA rule treats seven or eight categories of information as individually identifying, it excludes from this list the kind of data used by the Panopticlick study—such as font information—that a child-directed website can use to defeat anonymization.

But the point is not that the FTC should add font information to COPPA's list. This response would miss the point entirely. The principal lesson for regulators is that they need to do more than make lists of PII. Given the potential for *any* type of information to serve an identifying role, lists of data categories are underinclusive as soon as they are written and will become more underinclusive over time.

To supplement traditional PII list making, regulators should find other ways of expressing and defining the kind of information likely to raise the risk of harm they seek to avoid. For example, regulators should specify characteristics of information collection that increase the odds that a website will be able to reidentify people in anonymized data. I propose a rule that does precisely this in the next Part.

II. How the FTC Should Define "Personal Information" in COPPA

In light of how the power of reidentification has disrupted the PII, list-making approach to privacy regulation, I propose that the FTC add a new subdefinition for personal information to the COPPA Rule that takes into account how the power of reidentification brings identifiability to classes of data we never before would have called personally identifiable.

A. A Quantity-based Definition of Personal Information

I propose specifically a *quantity-based* definition of personal information. Computer scientists recognize that the likelihood of successful reidentification increases with the amount of information possessed by a data owner.[19] These computer scientists formalize this idea through the concept of entropy.[20] Entropy

[19] Ohm, *supra* note 3, at 54.
[20] *Id.* at 41-42, pt. III.A.3.

measures how close a database owner is to connecting a given fact to an individual.[21]

Consider entropy in the children's game Twenty Questions. At the start of a game, the Answerer thinks of a subject the Questioner must discover through yes or no questions. Before any questions have been asked, entropy sits at its maximum because the Answerer can be thinking of any subject in the world. With each question, entropy decreases as each answer eliminates possibilities. The item is a vegetable; it is smaller than a breadbox; it is not green. The Questioner is like the reidentifier, connecting outside information to the anonymized database, reducing entropic uncertainty about the identity of his target.

Roughly speaking, each additional field of information in a database decreases the entropy of the information, increasing the odds that the information can be linked to a specific individual.[22] For example, if website owner A knows a user's IP address and website owner B knows that same user's IP address and state of residence, website owner B's data contains less entropy than website owner A's. Likewise, if website owner C adds behavioral information—perhaps website owner C knows that the user recently clicked on an advertisement for a particular kind of toy—the entropy is even less.

As a website gathers and stores information about its users, entropy decreases, and the chance that the website can identify a user increases. This is true whether or not the website purposely omits the information in the current COPPA Rule's list, such as name, e-mail address, home address, and Social Security number.[23] Above some quantity threshold, the information held about a particular user should qualify as sufficiently identifying.

The difficulty is trying to specify precisely how much information should count as the critical threshold. Whether a given collection of information counts as uniquely identifying in any given dataset turns on variables about the type of information held, how much users share characteristics in common, and what kind of outside information the website owner can access to correlate with the data it holds. If COPPA restricted the meaning of personal information only to information that was perfectly, uniquely identifying we would be hard-pressed to come up with an administrable quantity standard.

Fortunately, COPPA does not insist that personal information be uniquely identifiable. As only one example, under the statute a child's "street name and . . . city" alone qualify as "individually identifiable information,"[24] even though a street name and city narrow down a user to a group of individuals in most cases and, for many streets in many cities, to a huge group of individuals.[25] Congress did not insist upon perfection; information that narrows down a child to a group is enough.[26]

---

[21] *Id.*
[22] *Id.*
[23] 16 C.F.R. § 312.2 (defining "personal information").
[24] 15 U.S.C. § 6501(8)(B).
[25] For example, consider how many people in Manhattan live on a given avenue.
[26] I provide a much more thorough analysis of the broad scope of COPPA's mandate to the FTC in a "Legal Appendix" appearing at the end of these comments. *See infra* Legal App.

I thus propose the following regulation:

> Personal information means individually identifiable information about an individual collected online including: . . . (h) any collection of more than N distinct categories of information about a user.

What remains is to set a value for N. As a starting point, consider how much information a website collects in a typical web server logfile, the file that tracks visitors to a website. Consider specifically the Apache web server, which is, according to most studies, the leading web server.[27] Apache keeps a file called access.log, with each entry in the file corresponding to a single transaction between a user and the website.[28] For each entry, Apache stores by default approximately seven pieces of information—originating IP address, identd information, userid, date/time, page requested, status code, and amount of data returned.[29] Apache also permits website owners who seek even more information to select a richer "combined log format" for this logfile, which stores two additional fields, referer, and user-agent string.[30] Because the user-agent string provides several pieces of additional information about the user's browser, it counts as at least three or four additional pieces of information itself.[31] Thus, Apache, even in a richer-than-default setting, stores only twelve pieces of information about each visit.

With Apache's twelve as a baseline, I propose the FTC initially set N at twenty-five. When a website collects more than twenty-five categories of information about a user, we can draw two conclusions. First, the collection of the data the website collects has much lower entropy than even the expanded Apache web log file. Second, the website owner who knows twenty-five things about a user is well on the path to reidentifying that user, if it should so choose, even if it intentionally anonymizes important identity information.

Thus, I propose the following new definition:

> Personal information means individually identifiable information about an individual collected online including: . . . (h) any collection of more than twenty-five distinct categories of information about a user.

B. Implementation Details

Finally, I provide a few implementation details, along the way responding to some possible objections to the proposal that may arise.

---

[27] The Apache server serves 54.02% of the world's hostnames according to the most recent data. Netcraft Web Server Survey, http://news.netcraft.com/archives/category/web-server-survey/ (June 16, 2010).

[28] I say "transaction" and not "visit" or "page view." A single page view typically generates multiple entries into the access.log file. For example, every image on a web page generates an access.log entry.

[29] Apache Software Found., Access Log, Log Files: Apache HTTP Server Version 2.0, http://httpd.apache.org/docs/2.0/logs.html#accesslog (2009).

[30] *Id.*

[31] Eckersley, *supra* note 9, at 5 tbl 1 (listing four categories of information in a User Agent string).

1. *How Many Websites Will be Affected?*

Although I have not undertaken any empirical research about the impact of this proposed rule, I imagine the impact will be slight, because the vast majority of websites will be able to ignore it. My recommendation focuses solely on the definition of "personal information" and would change nothing about the meaning of "directed at children" or "website or online service."[32] Websites not directed at children that can ignore COPPA today can probably continue to ignore COPPA under the new regulation.

And even a website directed at children could avoid the proposed rule simply by limiting the amount of information it collects. The rule's threshold should be set high enough to leave unaffected the many websites for children that collect only the Apache (or equivalent) logfile information and even a bit more. A website that collects a modest amount of personalization information about its users—say through a survey of four or five questions—will still fall below the threshold value of twenty-five and can safely ignore the new rule.

If the FTC worries about the number of websites covered by the rule, it can tune the threshold value upward, perhaps to thirty or forty. The post-PII approach to privacy regulation works like a risk assessment not like a bright line. Regulators should treat privacy rules like a tuning knob for risk, trying to set the knob at a place that protects privacy enough without burdening desirable behavior too much.

2. *What an Affected Website Must Do to Comply*

A website that thinks it may be covered by the regulation would need to analyze its data in a way most websites tend not to do today. But do not misunderstand the complexity of the task described: The only thing a website needs to do is count. The regulation would require affected websites to engage in a lightweight audit of their data, counting the number of categories of information they collect about their users.[33] Websites that collect very little information will not need to count at all, because they can safely assume they fall below the threshold. The proposed regulation does not demand a precise count; a website needs to calculate only whether it is above or below the numerical threshold. The precise count is not required.

I concede that most websites probably do not count their data in this way today, so the regulation will require some websites to expend modest new resources to

---

[32] The proposed definition will have only a minimal impact on the meaning of the critical phrase, "actual knowledge that it is collecting personal information from a child." § 6502(a)(1). By including the modifier "actual," Congress intended to impose a higher standard than a mere knowledge standard, in particular not triggering the statute in cases of constructive knowledge. Thus, what the proposed definition will do is trigger the requirements of the statute and rule whenever a company collects more than twenty-five categories of information about a child with actual knowledge. It is probably safe to assume that any company that does so already deals with COPPA compliance concerns.

[33] The rule does present one potential ambiguity, albeit one the FTC can easily define away. What counts as a category? I have not defined this term, which I think is fairly self-defining, but the FTC may decide it better to provide a more detailed definition.

comply. Moreover, every time a website decides to collect new categories of information from users, it needs to recalculate its count.

But although the regulation will impose a new burden on a small number of websites, the new burden is consonant with both the goals of COPPA and sound public policy. The rule will require websites to think about the information they collect along a new dimension, quantity. Privacy rules too often focus only on the quality of information, disregarding quantity, even though large quantities of information often put people at risk of privacy harm.[34] What the regulation would require is a new metric of transparency, one that is simple to calculate.

3. EXAMPLE IMPLEMENTATIONS

To better understand how the proposed regulation will operate, consider two hypothetical websites that are likely covered by the rule. First, a website already "directed at children" or with "actual knowledge that it is collecting personal information from a child,"[35] will trigger the requirements of the COPPA rule whenever it collects more than twenty-five categories of information about users. If the website already collects Apache's combined access.log file, which gathers twelve pieces of information,[36] and it also collects four more categories of information from each child user, say his or her age, date of birth, country of residence, and gender, its total data collection would still fall below the quantity threshold. As the website collected more information—favorite food, favorite band, name of school, name of best friend—its data would inch closer to the numeric threshold. Finally, once the website collected more than twenty-five pieces of information, it would trigger the requirements of COPPA. This is so even if the website could have escaped the strictures of the prior COPPA rule by relying on anonymization—say by deleting names, e-mail addresses, phone numbers, and the other identifiers in the list.

Again, this reflects sound public policy in light of the newly revealed power of reidentification. The more information a website collects about a child, the more likely it is the website can identify and contact the child.

Second, a company that stores massive amounts of information about individuals—such as large data brokers, credit agencies, and internet search engines—will automatically fall within COPPA whenever it decides to operate a website directed at children or with actual knowledge that it is collecting personal information from a child. Once again, this is a sound result. Companies that amass large databases cannot credibly contend that they lack the ability to identify people in anonymized databases. Accordingly, websites with massive databases should no longer be given a free pass for deleting categories of information like e-mail addresses and Social Security numbers. For companies like these, such acts of anonymization are empty gestures, so the choice under COPPA should be a simple

---

[34] Ohm, *supra* note 3, at 54.

[35] 15 U.S.C. § 6502(a)(1).

[36] The proposed rule counts *categories* of information but not *pieces* of information within a category. Thus, the Apache log file counts as only twelve categories of information even though for any one category of information (say, web page requested), a website may collect many different pieces of information for that category for a given user (say, all of the web pages requested in the past ninety days). The FTC may wish to clarify this in the definition.

one: Comply with the statute and the requirements of the rule (most importantly, obtain parental consent) or stop targeting children.

CONCLUSION

The approach I outline above marks a departure from how the FTC has exercised its COPPA authority in the past. It shifts the FTC's approach to protecting the privacy of children from specific to general and from qualitative to quantitative. These are necessary departures in light of new developments in computer science that enable new threats to privacy. I hope the FTC agrees and I look forward to answering any questions the proposal may inspire.

Sincerely,

Paul Ohm
Associate Professor

LEGAL APPENDIX: THE FTC'S BROAD POWER TO DEFINE "PERSONAL INFORMATION"

COPPA defines the key term "personal information,"[37] and empowers the FTC to add to the statutory definition.[38] In Part II, I urged the FTC to exercise this power a bit more broadly than it has in the past but still consistently with its statutory authority, by enacting a new quantity-based meaning to the definition. In this legal appendix, I explain how this proposal is consistent with COPPA, as revealed by the text and structure of the Act.

COPPA defines "personal information" as "individually identifiable information about an individual collected online," listing five specific examples: "a first and last name; a home and other physical address including street name and name of a city or town; an e-mail address; a telephone number; [and] a Social Security number."[39] Subsection 6501(8)(F) empowers the FTC to add to this list "any other identifier that the Commission determines permits the physical or online contacting of a specific individual." Finally, subsection 6501(8)(G) defines information that "combines with" other personal information as personal information.

The definition I proposed in Part II falls within the FTC's statutorily defined power because, as I read the statute, the FTC may define "personal information" to include any piece or class of information that can be used (1) directly to contact an

[37] 15 U.S.C. § 6501(8).
[38] *Id.* at § 6501(8)(F).
[39] *Id.* at § 6501(8)(A – E).

individual; (2) indirectly as an identifier or set of identifiers that can be used to link databases of information about a person together; or (3) indirectly by narrowing down the pool of people the information may potentially describe. Let us consider each of these categories in turn.

First, the FTC may include identifiers that, like telephone numbers or email addresses, may be used directly to contact a child.[40] In the FTC's current COPPA Rule, enacted first in 1999, it added "instant messaging user identifier" to the definition, which falls within this direct contact category.[41] It would be a mistake to conclude, however, that Congress meant to restrict the FTC to including only information useful for direct contact.

Congress suggested that it meant more than just direct contacting by using the word "permits". *Id.* The dictionary provides many definitions of permit, one of which is "[t]o give leave or opportunity for something; to provide the right conditions for something; to make something possible."[42] At least under this definition, something permits contacting by "provid[ing] the right conditions" for contacting, without necessarily directly enabling. Thus, identifiers that bring a "website or online service"[43] ("website") closer to a child's identity can "permit the . . . contacting" of the child, even if the information alone does not suffice for making contact.

Further support for this interpretation comes from Congress's instruction to the FTC to look for "other identifier[s]" that "permit[] the physical or online contacting of a specific individual."[44] By using the word "other," Congress directs the FTC to look to the categories of information in 6501(8)(A – E) for interpretive guidance.[45] Only two of these five categories permit direct, immediate contact with a person: "(C) e-mail address" and "(D) telephone number." The other three categories do not: "(A) first and last name"; "(B) home and other physical address including street name and name of city or town"; and "(E) Social Security number." Perhaps "other" refers only to the former two categories, but the more natural reading, and the one I urge the FTC to embrace, interprets "other" to refer to all five of the categories.

Two of Congress's five categories dictate the second and third parts of my formulation of the FTC's authority. The second part, "information that can be used . . . indirectly as a common identifier or set of identifiers that can be used to link databases of information about a person together," follows directly from Congress's decision to include Social Security numbers.[46] Obviously, a Social Security number serves no direct role in communication. Congress nevertheless sensibly included Social Security numbers in the list because a Social Security number can be used to join information about a single person from two or more different databases by serving as a common identifier in the data. Congress focused

---

[40] *Id.* at § 6501(8)(F) (empowering the FTC to list "any other identifier" that "permits the physical or online contacting of a specific individual.").
[41] *Id.* at 16 C.F.R. § 312.2 (defining "personal information").
[42] Oxford English Dictionary Online, http://dictionary.oed.com (equivalent to 3d ed., current as of June 2010).
[43] 15 U.S.C. § 6501(10)(A).
[44] *Id.* at § 6501(8)(F).
[45] *Id.*
[46] *Id.* at § 6501(8)(E).

COPPA not only on direct communication but also on the ability of database owners to link disparate databases, combining incomplete information from each into a richer whole.

The FTC has already recognized its power to add identifiers like these to the definition of "personal information" by including in the 1999 Rule, "[a] persistent identifier, such as a customer number held in a cookie or a processor serial number, where such identifier is associated with individually identifiable information."[47] The FTC should continue to consider other identifiers that can be used to link databases like a Social Security number or a customer number can.

Finally, the third part of my formulation, "information that can be used . . . indirectly by narrowing down the potential pool of people the information may describe" derives directly from Congress's decision to include a "home or other physical address including street name and name of city or town" in the definition of personal information.[48] This subdefinition conspicuously omits several key pieces of information—street number, state, and ZIP code—that one needs to send a letter to a child. The omission of street number is especially telling, because without it, one would have no hope of narrowly targeting a communication to a particular household, much less a particular child. Still, Congress considered a mere street name plus city to be "individually identifying" enough to "permit the . . . contacting" of a child; for example, knowing that a child lives on Kittredge Loop Road in Boulder, Colorado, alone triggers COPPA's responsibilities. Like the decision to include Social Security numbers, this decision was wise, because Congress must have realized the risk to privacy when a website collects enough information to take steps along the road to identification even before it can uniquely identify an individual.

Thus, Congress provided a broad grant of authority to the FTC to define "personal information," and the FTC has exercised this broad authority in its 1999 Rule, which it renewed in 2006.[49] This authority is broad enough to support the definition I proposed in Part II, especially as the advances in reidentification described in Part I.A has deepened our understanding of databases, uniqueness, and identity in ways that, properly understood, expand the scope of the FTC's power under COPPA even more.

[47] 16 C.F.R. § 312.2 (defining "personal information").
[48] 15 U.S.C. § 6501(8)(B).
[49] 71 Fed. Reg. 13,247 (March 15, 2006).