Dr. Alessandro Acquisti
Associate Professor
Heinz College, Carnegie Mellon University
4800 Forbes Av
Pittsburgh, PA 15213
acquisti@andrew.cmu.edu

July 20, 2012

Future of Privacy Forum

919 18th Street, NW, Suite 901

Washington, D.C. 20006

Dear Members of the scientific committee for the Privacy Papers for Policy Makers 2012 compilation:

Please find attached our submission to your digest. The attachment contains two versions of a manuscript discussing privacy, face recognition, and online social networks. The first version was accepted and published in the proceedings of *BlackHat USA 2011*. The second is an extended and more detailed version of the former, prepared for journal review.

Thank you for your consideration.

Warm regards,

[by email]

Alessandro Acquisti

Faces of Facebook:
Or, How The Largest Real ID Database In The World Came To Be

Alessandro Acquisti, Ralph Gross, and Fred Stutzman

Carnegie Mellon University

Proceedings of BlackHat USA 2011

Las Vegas, Nevada, August 2011

In Steven Spielberg's movie rendition of "Minority Report" (a short story by author Philip K. Dick), advertising technology has become so advanced by 2054 that remote retina scans can be used for personalizing electronic billboards to match the interests and backgrounds of passers-by. That future may be much closer than imagined, and perhaps even more ponderous. The research we describe in this white paper investigates how the combination of online social network data and commercially available off-the-shelf facial recognition applications can be used to successfully identify individuals online (for instance, across different sites, such as a social network and a dating site) and offline (for instance, on the street), as well as to infer -- in real-time -- additional, sensitive information about those individuals. The application of such re-identification techniques to brick-and-mortar and electronic commerce may be enthralling; at the same time, its privacy implications are unnerving.

Our research is based on the near-future, inevitable convergence of two trends: 1) the slow but steady improvements in computer facial recognition algorithms, and 2) the avalanche of personal photos that Internet users post publicly online, often in an identified format. For instance, Facebook has become the largest repository of photos on the Internet. Since Facebook has been enforcing (albeit unevenly) a verified, single identity policy (under which Facebook users are required to create profiles under their real first and last names, and the usage of pseudonyms can lead to one's account deletion), Facebook profiles may soon become the largest identity database in the world; a sort of *de facto* "Real ID" that markets and IT, rather than government and regulation, have created.

The existence of such a large and semi-openly accessible database of identities makes it plausible to consider scenarios whereby members' profile data can be used to re-identify individuals both online (for instance, on websites where their photos are uploaded without their names) and offline. We designed three experiments to test the feasibility and effectiveness of using social network profiles for individual re-identification. The first two experiments tested the possibility of identifying individuals both online and offline. The last experiment tested the possibility of inferring even more personal and sensitive information about a stranger merely by combining, in real time,

facial recognition algorithms and access to online resources through a simple mobile device.

In the first experiment, we used images from Facebook profiles that were publicly accessible directly via popular search engines (such as Google), and successfully re-identified a significant proportion of pseudonymous profiles on a dating site popular in the United States.

In the second experiment, we used publicly available images from a social networking site popular among college students to identify individuals walking around the campus of a North-American academic institution. Passers-by were invited to participate in the experiment by sitting in front of a webcam for the time necessary to take three photos, and then by completing a short survey. While a participant was completing her survey, her photos were uploaded to a computing cluster and matched against a database of images from profiles on the social networking site. Thereafter, the participant was presented with the images that the facial recognizer had ranked as the most likely matches for her photograph. The participant was asked to complete the survey by indicating whether or not she recognized herself in each of the images. Using this method we re-identified a significant proportion of participants.

In the third experiment, we inferred personal information from a subject's social network profile in real time, after recognizing her face through an application installed on a common mobile phone device. We then linked to her, through her face, additional personal information found (or inferred, through data mining) online, and displayed that information on the phone. This example of an "augmented reality" application embodies both the promises and the significant perils raised by the upcoming combination of facial recognition, social networks data, cloud computing, and mobile devices.

Faces of Facebook: Privacy in the Age of Augmented Reality

Alessandro Acquisti [*] **and Ralph Gross** [†] **and Fred Stutzman** [‡]

[*]Carnegie Mellon University, Pittsburgh, PA 15213, United States,[†]Carnegie Mellon University, Pittsburgh, PA 15213, United States, and [‡]Carnegie Mellon University, Pittsburgh, PA 15213, United States

PRELIMINARY JOURNAL DRAFT – CONFERENCE VERSION AVAILABLE IN THE PROCEEDINGS OF BLACKHAT USA 2011

We investigate the feasibility of combining publicly available Web 2.0 data with off-the-shelf face recognition software for the purpose of large-scale, automated individual re-identification. Two experiments demonstrated the ability of identifying strangers online (on a dating site where individuals protect their identities by using pseudonyms) and offline (in a public space), based on photos made publicly available on a social network site: combining face recognition and publicly available data from Facebook profiles, we re-identified 10% of users of a popular dating site and over 30% of students walking on the campus of a North-American college. A third proof-of-concept experiment illustrated the ability of inferring strangers' personal or sensitive information (their interests and Social Security numbers) from their faces, by combining face recognition, data mining algorithms, and statistical re-identification techniques: starting merely from photos of their faces, we correctly identified interests for all the students who had participated in the previous experiment, and, in some cases, the first five digits of their Social Security numbers. The results highlight the implications of the inevitable convergence of face recognition technology and increasing online self-disclosures, and the emergence of "personally predictable" information. They raise questions about the future of privacy in an "augmented" reality world in which online and offline data will seamlessly blend.

Statistical Re-Identification | Face Recognition | Privacy | Social Network Sites | Augmented Reality | Visual Searches | Personally Predictable Information | Social Security Numbers

In 1997, the best computer face recognizer in the US Department of Defense's Face Recognition Technology program scored an error rate of 0.54 (the false reject rate at a false accept rate of 1 in 1,000); by 2006, the best recognizer scored 0.01 — a rate almost two orders of magnitude smaller [28]. In 2000, of 100 billion photographs shot worldwide [20], a negligible portion found their way online; by 2010, 2.5 billion digital photos *a month* were uploaded by members of Facebook alone [14]. Often, these photos depicted people's faces, they were tagged with real names, and they were shared with friends and strangers alike. This manuscript attempts to forecast the consequences and implications of the inevitable convergence of these trends: the increasing public availability of facial, digital images; and the ever-improving ability of computer programs to recognize individuals in them.

Research in computer face recognition has been around for over thirty years [23], cycling between promising breakthroughs and recurrent realizations that its successes remain limited under real world conditions [32]. Although computer face recognizers may never replicate human ability to identify people, their accuracy has improved so consistently that the technology has found its way into end-user products, and in particular Web 2.0 services. Following the acquisition of Neven Vision in 2006, and of Like.com more recently, Google has offered Picasa users face recognition to organize photos according to the individuals they depict [17]. Apple's iPhoto has employed face recognition to identify faces in a person's album since 2009 [9]. With Face.com's licensed technology, Facebook has employed face recognition to suggest "tags" of individuals found in members' photos [12].

So far, end-user Web 2.0 applications are limited in scope. They are constrained by, and within, the boundaries of the service in which they are deployed. Before being acquired by Apple, for instance, Swedish startup Polar Rose worked on a smart-phone "augmented ID" application that allowed users to point the camera at a person and identify her social media information — but only "[p]roviding the subject has opted in to the service and uploaded a photo and profile of themselves" [1]. (The application was never released, and no external review of its performance was published.) Face.com has developed face recognition services for Facebook users — but "if you choose to hide your Facebook tags, [their] services will get blocked out when attempting to recognize you in photos" [15]. In early 2011, CNN reported that Google was working on a mobile application that allowed users to snap pictures of people's faces to access their personal information — but only for people who "check[ed] a box agreeing to give Google permission to access their pictures and profile information" [2]. (Google later challenged the story, stating that face recognition would not be added to its search products unless the company could "figure out a strong privacy model for it" [5].)

Implicit in the above reports is the suggestion that face recognition is a technology that can be controlled, either by opt-in, or by limiting to a specific category of individuals the database against which a face will be matched. For instance, law-enforcement agencies in the US may soon use hand-held facial-recognition devices, but only to be used against "a database of people with criminal records" [6].

The genie, however, may already be out of the bottle. In recent years, massive amounts of identified and unidentified facial data have become *publicly* available through Web 2.0 applications, and so have the infrastructure and technologies to navigate through those data in real time, matching individuals across online services, independently of their knowledge or consent. In the literature on statistical re-identification [31, 24], an identified database is pinned against an unidentified database, in order to recognize individuals in the latter and associate them with information from the former. Many online services make available to visitors identified facial images: social networks such as Facebook and LinkedIn, online services such as Amazon.com profiles, or organizational rosters. Most Facebook users, for instance (estimated at over 750 million worldwide [13], with a collective 90 billion up-

Reserved for Publication Footnotes

loaded photos [29]), use photos of themselves as their primary profile image. These photos are often identifiable: Facebook has aggressively pursued a 'real identity' policy, under which members are expected to appear on the network under their real names under penalty of account cancelation [3]. Using tagging features and login security questions, Facebook has nudged users to associate their and their friends' names to uploaded photos. Facebook photos are also frequently publicly available. Primary profile photos *must* be shared with strangers under Facebook's own Privacy Policy ("Facebook is designed to make it easy for you to find and connect with others. For this reason, your name and profile picture do not have privacy settings"[1]) Many members also make those photos searchable from outside the network via search engines. Similarly, LinkedIn profiles — which are almost unfailingly associated with individuals' real first and last names — contain photos that can be perused by a visitor without logging onto the service or even accessing the site (since they are cached by search engines).

Unidentified facial images, on the other hand, can be found across a range of services, often sensitive, where members use pseudonyms to protect their privacy. Pseudonyms are common on photo sharing sites such as `flickr.com` or the more risque `tumblr.com`; on dating sites such as `match.com` or `manhunt.com`;[2] on adult sites such as `ashleymadison.com` or `adultfriendfinder.com`; or on sites where members report sensitive financial information, such as `prosper.com`.

Of course, unidentified faces are those of the strangers we walk by on the street. A person's face is the veritable conduit between the offline and online worlds. This manuscript examines how someone's face can become the link across different databases that allows strangers to be identified, and the trails of data associated with their different persona to be connected.

In three IRB-approved experiments, we investigated whether the combination of publicly available Web 2.0 data and off-the-shelf face recognition software may allow large-scale, automated, end-user individual re-identification. We identified strangers online (across different online services: Experiment 1), offline (in the physical world: Experiment 2), and then inferred additional, sensitive information about them, combining face recognition and data mining, thus blending together online and offline data (Experiment 3). Finally, we developed a mobile phone application to demonstrate the ability to recognize and then predict someone's sensitive personal data directly from their face in real time. We summarize the design and results of the experiments in this manuscript. The technical details are available in the Appendix.

Experiment 1: Online re-identification

In our first experiment, we used publicly available photos uploaded to a popular social network site to re-identify the members of an online dating site.

Materials and Methods. Our target population consisted of members of one of the most popular dating sites in the US ("DS") who lived in a North American city ("the city"). We chose a dating site as target due to the popularity of those services (over 10 million Americans were estimated to be member of one in 2006, and the number was reported as growing in 2008 [16]) and their sensitivity: While dating sites' operators actively encourage their users to include images of themselves, they also warn them of the risks of providing identifiable information (none of the profiles we used in our study contain, in fact, real names, phone numbers, or addresses).

Our source population consisted of Facebook ("FB") members from the same city. While members of the dating site choose pseudonyms to protect their privacy, an overwhelming majority of Facebook users join the service using their actual first and last names ([19] estimated in 2005 that 89% Facebook profiles on a campus network were identified with the owner's actual name; see below more recent data). In addition, as noted above, many users open their profiles to search engines.

Our goal was to estimate how many "matches" we could find, using face recognition, between the set of FB members and the set of DS members. We defined a "match" as a correct linkage between an identified face on FB and an unidentified face on DS. A match, therefore, makes possible the identification of an up-till-then anonymous DS user.

Preliminary Survey. To provide a backdrop to the actual experiment, which is described below, in November 2010 we recruited 429 U.S. adult Amazon Mechanical Turk users for a survey about their "usage of Web 2.0 applications" (Survey 1). We do not consider this sample nationally representative, but merely a source of an order-of-magnitude approximation of the scenario we were exploring. Specifically, since we had no *a priori* estimate of the amount of overlap between the FB and the DS members sets, we asked our subjects, anonymously, about their membership in FB and DS. Among all subjects, 85.08% claimed to be current FB members, and 3.73% claimed to be current DS member (17.25% claimed to have been its members at some point in the past). The overwhelming majority, but not the entirety, of current DS members were also current FB members (87.50%). (In abstract terms, the number of DS members who are also on FB represents the theoretical upper bound to our ability to re-identify members of the former via images from the latter.) We also asked subjects who had claimed to be FB members whether they used their real first and last names on their FB profile; 89.91% answered yes (lending support to [19]'s results from an different and older sample of FB users). While anyone posting facial images of themselves on the Internet must realize that they may be recognized by strangers or friends, the possibility might seem remote, and worrisome, to many. We asked DS past and current members how uncomfortable they would feel if a stranger could identify their name simply looking at their dating site profile. On a 1 to 7 Likert scale, the modal answer was 7 ("very uncomfortable"; 41.05% of subjects reported high, 6, or very high, 7, discomfort levels; mean: 4.53; sd: 2.07).[3]

Experiment. In early 2011, we used Google API to search FB profiles of users likely to be located in or near the city. Since FB no longer organizes users around geographical networks, our search strategy consisted in a combination of queries: searching for profiles that listed the city as "current location," profiles that merely listed the name of the city, and profiles that listed universities or major institutions related to the city. This strategy is a noisy approximation of the set of FB users in the city: Profiles from the city may not appear in the searches, while profiles not from the city may.[4] Using this strategy, we identified 277,978 FB profiles of users likely to be located in

[1] From `http://www.facebook.com/policy.php`, accessed July 22, 2011.

[2] In 2010, Manhunt raised privacy concerns by making changes that made it "easier for people to see profiles without being members" [4].

[3] We ran a second, similar survey focusing on U.S. adults from the city in which we ran Experiment 1. The results are equivalent to those presented here, although the sample size was significantly smaller.

[4] Sixty-five percent of FB users among the participants in Survey 1 claimed that they openly listed their "current location" on Facebook. Hence, we believe that a sizable proportion of actual FB users from the city was captured through our search strategy.

the city, and then downloaded each profile's name and primary photo directly from the search engine. For virtually all profiles (274,540), a primary profile photo was available. We then applied a commercially available face detector and recognizer (PittPatt; [25]) to the set of photos found on the search engine. PittPatt found one face in 80,040 profiles (29.2%) and multiple faces in 23,137 profiles (8.4%), detecting a total of 110,984 unique faces (or "templates"). Those formed our source set.

While the search for FB profiles was based on keywords associated with the city, the search for DS profiles relied directly on geographical metadata. We queried the DS for all profiles located within 50 miles from a ZIP code approximately centered in the city's Metropolitan Statistical Area, applying the additional criteria that profiles indicated interest in either opposite sex or same sex partners and an age between 18 and 95. We then filtered the 18,550 discovered DS profiles, keeping only those who listed boroughs included within the city's Urbanized Area. This reduced the number of DS profiles to 5,818. PittPatt detected at least one face in 4,959 (85%) of these profiles. They comprise our target set. When multiple photos were included in the profile, we used PittPatt's clustering algorithms to create composite models of a profile owner's face based on highly similar faces across photos within the same profile.

Results. We ran PittPatt recognizer to find matches between the DS and FB sets. PittPatt produces matching scores between -1.5 (a sure non-match) and 20 (a sure match — usually representing cases where the very same photo was found in the two sets). We used a cloud computing cluster with four computing cores to calculate matching scores for slightly more than 500 million DS/FB pairs.[5] To evaluate the results, we picked the highest-scored pair for each DS profile, and recruited human coders to independently grade the anonymized pairs on a 1 to 5 Likert scale (from "Definitely the same person" to "Definitely not the same person").[6] To ensure reliability, we only kept grades of coders who graded at least 30 pairs without mis-grading any of the test pairs we inserted for validation purposes (either sure matches, or sure non-matches). We also eliminated coders with more than 30% score "deviations" (defined as situations where the majority of graders considered a pair a match but the grader considered it as non-match, or viceversa). We had multiple coders grade each pair, with no fewer than five per pair.[7] We classified as likely matches pairs that were graded by at least two-thirds of the graders as either a definite or likely match. Those represented 369 of 5,818 profiles in our target set, or about 6.3%. Including also pairs that the *majority* of graders classified as a definite or likely match, the number raises to 610 of 5,818, or 10.5%. We manually validated these results by checking cases in which any grader had suggested a non-match. Figure 1 shows the distribution of PittPatt scores across all pairs, as function of the human coders' evaluation.

The results imply that about one out of ten DS members could be identified starting from search engine searches of Facebook profiles in the same geographical area. The computation of the more than 500 million pairs took the face recognizer about 15 hours (roughly 0.00019 seconds per pair, or about 21 seconds per DS profile). For comparison, the human graders took on average 14 seconds to review each *pair*. If a single individual had to grade all the 500 million pairs, the task would have required almost 2 million hours to complete.

Comments. Experiment 1's results are optimistic in some ways — i.e. the possibility that even a majority of human graders could misidentify a face — and conservative in others. Our approach to re-identifying pseudonymous DS users was conservative because we accessed FB identified photos only through a search engine (that is, without even logging onto the social network); we used one single identified photo per potential target subject; and we established whether or not we had found a match for a given target individual using only the identified source image with the *highest* matching score to the unidentified target image. In other words, we only considered the face recognizer's single best prediction, and disregarded instances in which the recognizer may have actually found the right FB profiles of a DS user, but assigned to it the second or third highest matching score.

Furthermore, our target and source data sets do not fully overlap (not every DS member will be a FB member), our search patterns for DS and FB profiles in the city by necessity differed, and both of our DS and FB sets likely did not contain all profiles of individuals actually located in the city and members of either DS or FB, or both. Any increase in the overlap between the two sets would be reflected in an increase ability of recognizing DS members.

Additionally, we did not consider false negative cases where the human graders may have not recognized a match, due to changes in the appearance of the person across photos — a plausible scenario, given differences in the performative nature of dating and social network sites.

Experiment 2: Offline re-identification

Experiment 2 extended Experiment 1 in a number of directions. Rather than focusing on online-to-online re-identification, Experiment 2 used social network data to re-identify individuals offline, in the physical world. Furthermore, Experiment 2 estimated how the ability to identify strangers improves when using more than a single photo for both source and target subjects, and when considering a *set* of high-scoring matches found by the face recognizer, rather than the single top match. Focusing on a set of high-scoring matches reflects an attacker model where face recognition is used to restrict the set of possible identities of a stranger from an arbitrarily large set to a set sufficiently small (such as ten potential matches) that a human can easily evaluate.

Materials and Methods. Individuals walking by the foyer of a building on the campus of a North American college ("the college") were approached and invited to participate in the experiment. They were asked to sit in front of a laptop equipped with a $35 webcam for the time necessary to have three shots taken (one frontal, and two with the subject's face slightly tilted towards either side). Then, each subject was asked to complete a short survey (Survey 2) on another laptop. While the subject was completing her survey, her shots were uploaded to a cloud computing cluster and matched against a database of photos from a social network site. By the time the subject reached the third page of the survey, she would find it populated with a sorted set of social network photos that the recognizer had ranked as the highest-scoring matches

[5] Each computing core comprised 3.25 EC2 Compute Units, where one Compute Unit provides the equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor.

[6] Most face recognition studies in the literature are based on a $known$ ground truth (the "true" correspondence between target and source images). We used human coders to assess the face recognizer's performance because the very point of our experiment was that such ground truth was not directly accessible: DS members protect their privacy by not revealing their identities. Note that [27] compared the performance of computer face recognition algorithms versus humans's performance, albeit in a scenario with known ground truth.

[7] After removal of inaccurate coders, as defined above, 454 graders completed the task. The average number of pairs graded by a coder was 149.79. The maximum number of pairs graded by one single grader was 1,987. Collapsing definite and likely matches together, versus unsure, likely, or definite non-matches together, Fleiss's kappa coefficient was 0.40.

against either of the three shots taken of the subject. For each photo, the subject indicated, directly on the survey, whether or not she could recognize herself in it. (We augmented these subjects-provided evaluations with additional analysis after the experiment was completed.)

The images had been found in the profiles of members of the Facebook's college network. We used a profile in the same network to analyze publicly available information about the friend connections of the network members; traversing their graph, we identified 25,051 profiles as members of the network. Of them, 82.2% contained one image (20,597), 14.6% contained multiple images (3,655), and 3.2% did not contain any image (799), for a total of 261,262 images.[8] The overall number of faces detected by the recognizer across all images was 114,745. (The recognizer detected one or multiple faces in 36.00% of the *main* profiles images; as noted earlier, main profile images are by default visible also outside the network, and often indexed by, and searchable through, external search engine searches.)

Results. Ninety-three subjects participated in the experiment on two consecutive days in late 2010.[9] Based on Survey 2's results, all were students at the college where the experiment took place, and all had a FB profile. However, only 68.97% of them were definitely members of the college FB *network*; 10.34% were not members, and 20.69% were not sure. (Due to our profile search strategy, this implies that we could expect, at best, to identify no more than about 90% of our subjects.) Eighty-four percent of subjects claimed to use as main profile image a photo of themselves. Almost one of two (51.72%) incorrectly believed that they had *not* made their main profile photos available to everyone else on FB (as noted, Facebook forces primary profile photos to be public), reflecting a misconceived perception of protection. Only 10.34% of the subjects actually made the *rest* of their profile information available to strangers — denoting more elevated privacy concerns over the textual information provided in a profile than over its primary profile photo. Looser concerns over the public disclosure of one's identified photo, however, may be compared, for context, to the significant discomfort expressed by the majority of participants in Survey 1 with the scenario in which a stranger identify them as users of a dating site using their profile photos.

Each of the three shots we took of a target subject was compared by the recognizer against the set source faces. [update number] percent of the subjects recognized themselves in the one of the ten highest-ranked matches found by the recognizer. We analyzed the data again few months after the experiment was conducted, using an upgraded version of the face recognizer and more source data. We were interested in measuring the improvement in accuracy in the space of few months of evolution in the recognizer's algorithms. Two independent graders labeled the results.[10] Using a conservative measure of success (all graders agreeing on a match), the recognizer found a FB photo that matched the source target within the highest-ranked matches for 31.18% of the subjects. Figure 2 shows an example of a successful match between a shot of a subject taken in the foyer of the college building, and an image found online depicting the same individual.

A correct match between a subject's shot taken during the experiment and a photo found on a FB profile creates a link between the (up till then, anonymous) Experiment 2's subject and that FB profile's (often identified) information. Through that link, it becomes feasible to infer the identity of the target subject (see Experiment 3). It took about 2.89 seconds to find the set of highly-scored matches for any given subject. Extrapolating, Experiment 2 suggests that the identity of about one third of subjects walking by the campus building may be inferred in a few seconds combining social network data, cloud computing, and an inexpensive webcam.

Experiment 3: Sensitive inferences

Experiment 3 consisted in a proof-of-concept test of the power of online self-disclosures and face recognition technologies to create linkages between the offline and the online worlds. It attempted to answer the question: can we predict personal, and even sensitive, information about strangers starting from a single, anonymous piece of information about them — their faces?

Answering that question involves exploiting a chain of inferences in a process of data "accretion" [26]. First, face recognition links an unidentified subject (such as a face among many in the street) to a record in an identified database (such as an identified photo of the subject on Facebook, or LinkedIn, or on Amazon). Once the link has been established, any online information associated with that record in the identified database (such as names and interests found in the subject's Facebook profile; or demographic data found on `Spokeo.com` - a social network and data aggregator - after searching for the subject's name) can in turn be linked to the unidentified subject. Lastly, through mining and statistical re-identification techniques, such online information can be used for additional and more sensitive inferences (such as sexual orientation [21] or Social Security numbers [7]), which in turn can be linked back to the originally unidentified face. Sensitive data is therefore linked to an anonymous face through some transitive property of (personal) information; it becomes "personally predictable information."

Materials and Methods. Experiment 3 predicted two pieces of information associated with subjects who took part in Experiment 2: their interests and their Social Security numbers. The subjects' interests were obtained from the subjects' Facebook profiles — which were found via the photos matched by the recognizer during Experiment 2. The subjects' SSNs were predicted combining the subjects' demographics (from their FB profiles) with the algorithm described in [7], which combines individuals' dates and locations of birth with data publicly available from the Death Master File.

The experiment consisted of the following steps:

- We designed an algorithm that, given a person's face detected by the recognizer in a photo found on a given FB profile, predicts the most likely profile's *name* of that person — that is, her actual identity.[11] Trivial for a human, the task is challenging for a computer: a face in a photo found on my profile may not depict me, but another person related to me, or even a stranger. The algorithm consists of a weighted combination of various criteria: a) whether the face found in the profile's photo was tagged (usually, tags in FB photos accurately link to the actual profile of the tagged person); b) whether the face was found in the primary profile photo of a profile (which raises the likelihood of the photo depicting the profile's owner); c) whether the face was clustered by the recognizer together with a set of faces that included one found in the profile's primary profile photo; d) whether the face was clustered together with a set of faces whose relative majority was tagged with

[8] The average number of images per profile was 10.4; considering only profiles with multiple images, the average number was 65.9.

[9] This number excludes subjects we used to test and pilot the experiment.

[10] Cohen's kappa: ; divergences in the graders' scores were manually resolved by a third grader.

[11] The algorithm assumes that the person is using her real name on her FB profile.

the same profile name; e) whether the face was clustered together with the largest cluster of faces within a given profile; and, finally, f) in which profile the photo containing the face was found. We tested the algorithm over a random subset of 500 images coming from the set of FB images that constituted the source data for Experiment 2. The ground truth (the actual profile of the person depicted in the photo where the person's face was found) was manually coded by human graders. The script accurately predicted [update this number]% of the profiles found by the graders.

- We then applied the script to the highest-ranked matches found by the recognizer for each subject in Experiment 2, in order to infer the most likely FB profile's associated with that source photo.
- From the profiles that made such information available, we then inferred names, interests, dates of birth, and hometowns of Experiment 2's subjects. For foreigners, we manually estimated time and location of arrival in the United States,[12] which [7] have shown to be highly correlated with the likely date of SSN application.[13] We fed the demographic information gathered in the previous step into the algorithm described in [7], and statistically predicted the most likely SSNs assigned to the subjects.
- Finally, we invited the subset of Experiment 2's subjects who had been correctly identified by the recognizer as the top-ranked templates in Experiment 2 to participate in a survey (Survey 3). Survey 3 asked subjects to evaluate our predictions of their interests and their SSNs. The survey was hosted on a secure server and designed so that the subjects' answers to questions about their SSNs could only be analyzed in the aggregate, and could not be linked back to individual survey participants — thus preserving the subjects' privacy.

Results. Out of the subjects who participated in Experiment 2, 29 that we identified using face recognition, and for whom we found publicly available demographic information, were invited by email in July 2011 to participate in Survey 3. Eighteen of them completed the survey (one subject started it, but did not complete it). For each subject, we had prepared a list of five personal interests, inferred from their FB profiles identified through face recognition. We correctly inferred at least one interest for *all* the subjects and, on average, 3.72 interests (out of 5) per subject — or about 75% of all interests. For the SSN predictions, we focused on whether we could predict with a few attempts the first five digits of the target subject's SSN. As discussed in [7], knowledge of the first five digits of a target victim is sufficient for effective brute force identity theft attacks. We correctly predicted the subjects' first five digits for about 16.67% of the subjects with two attempts, and 27.78% with four attempts. Although the sample size of subjects who participated in Experiment 3 was by necessairy small, the accuracy is significant: the probability of correctly guessing by random chance the first five digits of just a single person's SSN with two attempts would have been 0.0028%.

Experiment 3's subjects were very concerned about the scenario the experiment depicted, and surprised by its results. Before being presented with the actual predictions and our questions about them, the subjects who participated in Survey 3 were asked about their degree of expected discomfort if a stranger on the street could know their interests and predict their SSNs. On a Likert scale from 1 ("Not at all uncomfortable") to 7 ("Very uncomfortable"), the modal scores across the subjects were, respectively, 6 (mean: 5.11) and 7 (mean 6.17). In the open-ended boxes at the end of the survey, *after* their predicted interests and SSNs had been presented on the screen, some subjects expressed additional concerns: "the Social Security concerns (and the possibility of linking my face to credit card information, etc.) is very worrisome"; "surprised & shocked with the accuracy of the options"; "[t]his is freaky. [...] Makes me re-assess what I should ever reveal on the internet."

Discussion

Experiment 3 was a proof-of-concept test of sensitive inferences through face recognition. The test was asynchronous: source photos to be matched against the subjects' live shots had been downloaded prior to the experiment, while the prediction (and evaluation) of subjects' interests and SSNs was completed subsequently to the recognition of people's faces. To illustrate the possibility of real-time identification, we developed a smart phone demo application that captures the image of a person and then overlays on the screen her predicted name and SSN. The application is an example of augmented reality [11], in which offline and online data blend together.

Underneath the application, various components silently interact on a remote server, replicating in real time what Experiment 3 did in asynchronous fashion in a controlled experimental environment. The application transmits the captured shot of someone's face to a server that contains a database of source photos from identified FB profiles, as well as a running version of the face recognizer. The recognizer creates a model of the captured shot and calculates its matching scores against each of the source templates in the database. The highest-matching template is selected, and the algorithm described in the previous section is invoked to predict the most likely FB profile of the person depicted in the shot. If available from the identified database, the person's name is then inferred. Another script then uses the name to query, still in real time, online people search services (such as `zabasearch.com` and `usa-people-search.com`) to infer the presumptive date of birth and previous residences of the target subject. From the earliest state of residence the presumptive state of birth is predicted.[14] The demographic information thus inferred is fed into [7]'s algorithm, which also resides on the server. Its SSN prediction, together with the presumptive name of the target, is passed back, encrypted, to the smart phone — which displays it on the screen over the person's face (Figure 3).

From Face Recognition to Personally Predictable Information. On the one hand, described above is but one of many possible combinations of components and their resulting inferences. The application could interface with data from a voter registration list, instead of the target subject's FB profile; or, it may attempt to predict the subjects's health data [31], instead of her SSN. The constant element in the process we described is the accretion of more and more sensitive data, starting from a face, which the combination of increasing online self-disclosures, consumer-end face recognizers, faster cloud computing, and more accurate data mining make possible: a world of personally predictable information, linkable from someone's face, through end-users' devices connected to the Internet.

[12] Based on the first college institution frequented, or first job worked in the US, as reported on the profiles.

[13] Two years after [7]'s showed that SSNs were predictable from public data, the Social Security Administration changed their assignment by randomizing it [30]. However, since the hundreds of millions of SSNs issued under the previous scheme have not been re-issued, they remain, theoretically, predictable.

[14] If the query returns multiple records for the same name, the current demo version of the application naively chooses one of the records in the same state as the current GPS location of the smart phone.

While face recognizers have long been in the arsenals of governments and corporations, the synergy of those technologies will "democratize" surveillance by making peer-to-peer face recognition cost effective and available. This will affect both demand and supply: The set of target subjects will no longer be limited to well-definable groups to which only authorized entities have access (convicted criminals, legal aliens entering the country, or DMV visitors), but will include the universe of individuals whose photos are public online. The set of *users*, in turn, will include anyone with devices as ordinary as mobile phones, as the capital and technological requirements for real-time face recognition become increasingly affordable.

Limitations. On the other hand, various constraints *currently* affect the scalability of the process we described. Mass face recognition is limited by the availability of (correctly) identified facial images, which is itself function of legal constraints (Web 2.0 photos may be copyrighted, or shielded by the Terms of Service of the site where they are found) and technical constraints (the ability to download, and analyze, massive amounts of digital images). Inferences, of course, are limited by the percentage of individuals for whom facial images can be found, and then (if found) exploited to infer further personal data. The accuracy of face recognizers is also function of the quality of subjects' photos (Experiment 1 and 2 relied on frontal photos, either uploaded by the subjects themselves to their dating site profiles, or captured by the researchers on campus). It is also function of the geographical scope of the set of source subjects (Experiments 1 and 2 were confined to geographically restricted communities: the "city" and the "college"). Frontal shots may be harder to capture in the street, and as the set of source subjects expands, computations get more time consuming and false positives increase.

Facial Searches. Technological and social trends make it plausible to infer that the constraints and limitations we just espoused will keep loosening. Due to default privacy settings in social network sites, social norms on self disclosures, and the existence of search engines that index social networks data, identified facial images are already publicly available for increasing numbers of individuals. Furthermore, tagging self *and* others in pictures has become socially acceptable.[15] Recurrent acquisitions of face recognition start-ups by large Silicon Valley players provide evidence of the significant business interest in this space. Two possible business developments seem plausible: first, some of the largest players, which are already amassing ever-increasing databases of identified images (much larger than what we used in our experiments), may start selling identification services to other entities — such as governments, corporations, or the shop on the corner of the street; second, "facial searches" — in which facial images are pre-processed and indexed by search engines the same way search engines currently index textual data — may become more common; soon, searching for a person's face, online may not seem as farfetched as searching for all instances of someone's name on the Internet may have sounded 15 years ago, before the arrival of search engines.[16] Once an identity is found, demographic information may be available from multiple sources (voter registration lists, people search services, social networks [7]). Cooperative subjects may not be needed for frontal pictures once wireless camera are cheaply deployed in — for instance — glasses, instead of mobile phones. Finally, face recognizers will keep improving in terms accuracy (including scenarios where frontal shots are not available), and cloud computing services are likely to keep offering more speed at cheaper prices, making it possible to run face recognition on larger sets of source subjects.

Privacy and Augmented Reality. The commercial implications of the convergence of social networks' data and face recognition will likely be far reaching. For instance, ecommerce strategies such as behavioral advertising and personalized offers will become possible for the up-till-then anonymous shopper on the street. The privacy concerns raised by these developments may be ominous, too [33]. The instinctual expectation of privacy we hold in a crowd — be that an electronic or a physical one — is challenged when anybody's mobile devices, or online searches, can recognize us across vast sets of facial and personal data in real time. Research in behavioral economics has already highlighted the hurdles individuals face when considering privacy trade-offs [8]. Those hurdles may be magnified by these technologies, not just because we do not expect to be so easily recognized by strangers, but because we are caught by surprise by the additional inferences that follow that recognition.

It is not obvious which solution may balance the benefits and risks of peer-based face recognition. Google's Eric Schmidt once observed that, in the future, young individuals may be entitled to change their names to disown youthful improprieties [22]. It is much harder, however, to change someone's face. Blurring of facial images in databases, k-anonymization of photos, or opt-ins, are all ineffective when re-identification can be achieved through already publicly available data. Although the results of two of our surveys (Survey 1 and 3) suggest that most individuals loathe the possibility of being identified by strangers on the street, many of them nevertheless disclosed online identified photos that will make that sort of identification possible. Notwithstanding Americans' resistance to a Real ID infrastructure, as consumers of social networks we have consented to a *de facto* "Real ID" that markets and information technology, rather than government and regulation, have created.

In addition to its privacy implications, however, the age of augmented reality and personally predictable information may carry even deeper-reaching behavioral implications. Through natural evolution, human beings have evolved mechanisms to assign trust in face-to-face interactions. Will we rely on our instincts, or on our tools, when mobile devices can make their own predictions about hidden traits of the person we are looking at? Will these technologies bring about new forms of discrimination? Or will they help combat existing ones?

ACKNOWLEDGMENTS. The authors gratefully acknowledge research support from the National Science Foundation under grant # 0713361, from the US Army Research Office under contract # DAAD190210389, from the Carnegie Mellon Berkman Fund, and from Carnegie Mellon Cylab. The authors thank Nithin Betegeri, Aravind Bharadwaj, Varun Gandhi, Markus Huber, Aaron Jaech, Ganesh Raj ManickaRaju, Rahul Pandey, Nithin Reddy, and Venkata Tumuluri for outstanding research assistantship, and Laura Brandimarte, Samita Dhanasobhon, Nitin Grewal, Anuj Gupta, Hazel Diana Mary, Snigdha Nayak, Soumya Srivastava, Thejas Varier, and Narayana Venkatesh for additional assistantship.

References

1. *Technology Review*, 2010.
2. *CNN*, 2010.
3. *The Guardian*, 2010.

[15] One of our subjects claimed, before Experiment 2, that we would not be able to find him due to his profile privacy settings. That subject was found anyway because of a photo uploaded to one of his friends' profiles.

[16] Google has already announced the deployment of visual searches based on images (although not faces) pattern matching [18]. For an overview of "Internet Vision" (the intersection of Computer Vision and the Internet), see [10].

4. *Edge Boston*, 2010.
5. *PC World*, 2011.
6. *The Wall Street Journal*, 2011.
7. A. Acquisti and R. Gross. Predicting Social Security numbers from public data. *Proceedings of the National Academy of Science*, 196(27):10975–10980, 2009.
8. Alessandro Acquisti. Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the ACM Conference on Electronic Commerce (EC '04)*, pages 21–29, 2004.
9. Apple, 2011.
10. S. Avidan, S. Baker, and Y. Shan. Internet vision. *Proceedings of the IEEE*, 98(8):1367–1369, 2010.
11. R.T. Azuma et al. A survey of augmented reality. *Presence-Teleoperators and Virtual Environments*, 6(4):355–385, 1997.
12. Facebook, 2010.
13. Facebook, 2011.
14. Faceboook, 2010.
15. Face.com. Faq, 2011.
16. J.L. Gibbs, N.B. Ellison, and C.H. Lai. First comes love, then comes google: An investigation of uncertainty reduction strategies and self-disclosure in online dating. *Communication Research*, 38(1):70–100, 2011.
17. Google, 2011.
18. Google, 2011.
19. R. Gross and A. Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the ACM Workshop on Privacy in the Electronic Society*, pages 71–80. ACM, 2005.
20. M. Hilbert and P. López. The worlds technological capacity to store, communicate, and compute information. *Science*, 332(6025):60, 2011.
21. C. Jernigan and B.F.T. Mistree. Gaydar: Facebook friendships expose sexual orientation. *First Monday*, 14(10), 2009.
22. Holman W. Jenkins Jr. Google and the search for the future. *The Wall Street Journal*, (August 14), 2010.
23. M.D. Kelly. Visual identification of people by computer, 1970. Tech. rep. AI-130, Stanford AI Project, Stanford, CA.
24. A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *Proceedings of the IEEE Symposium on Security and Privacy*, pages 111–125, Oakland, CA, 2008.
25. M. Nechyba and H. Schneiderman. Pittpatt face detection and tracking for the clear 2006 evaluation. *Multimodal Technologies for Perception of Humans*, pages 161–170, 2007.
26. P. Ohm. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, 57:1701, 2010.
27. A.J. O'Toole, P.J. Phillips, F. Jiang, J. Ayyad, N. Pénard, et al. Face recognition algorithms surpass humans matching faces over changes in illumination. *IEEE transactions on pattern analysis and machine intelligence*, pages 1642–1646, 2007.
28. Phillips, 2007.
29. Quora, 2011.
30. Social Security Administration, 2011.
31. L. Sweeney. Weaving technology and policy together to maintain confidentiality. *Journal of Law, Medicine and Ethics*, 25(2-3):98–110, 1997.
32. W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.
33. Michael Zimmer, 2009.

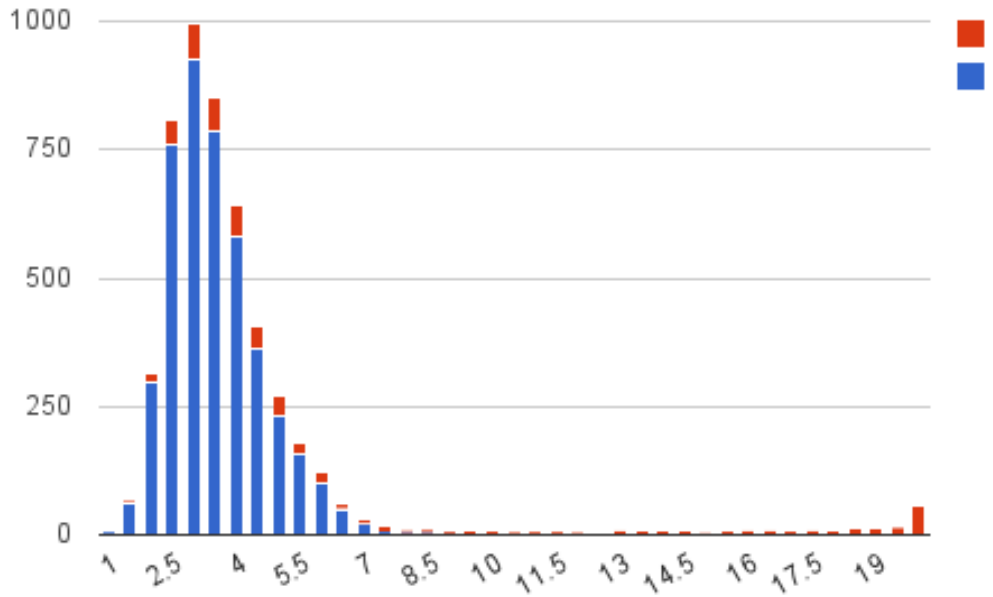

Fig. 1. Experiment 1: Distribution of PittPatt scores across all pairs, as function of the human graders' evaluation.

Fig. 2. Experiment 2: Exemplary target shot and matched source photo for one of the participant.

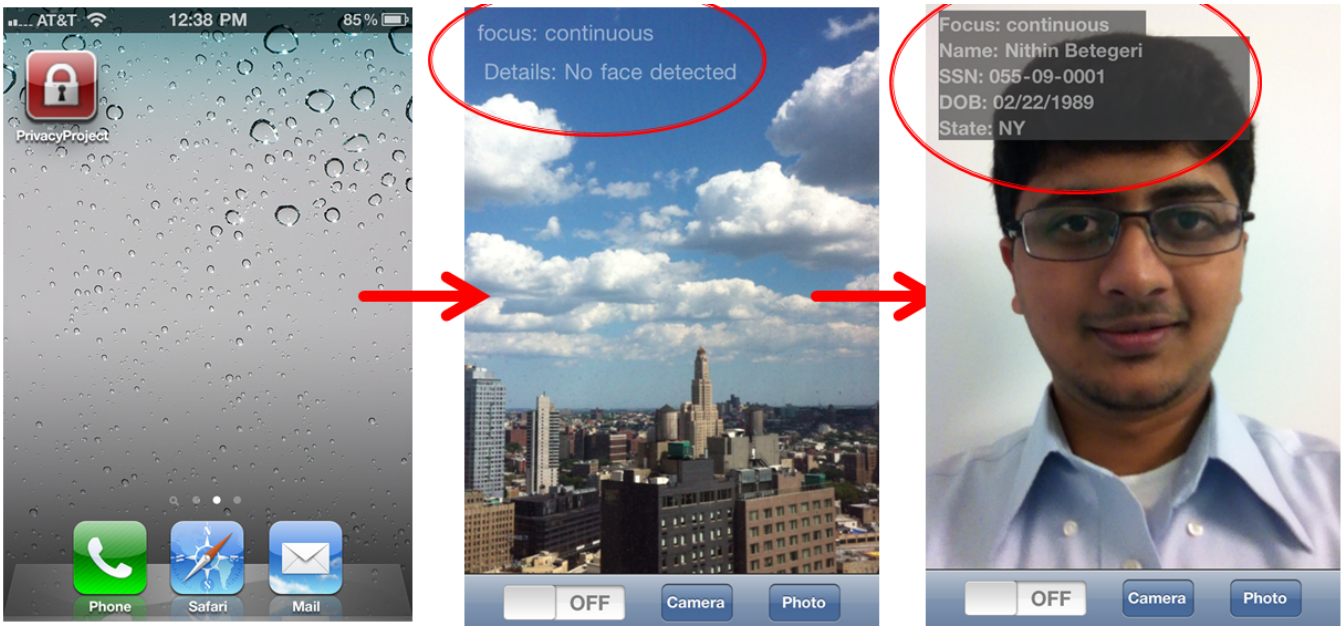

Fig. 3. Experiment 3: Screenshots from the real-time mobile phone application.