Journal of Privacy and Confidentiality (2014) **6**, Number 2, 1–20

Face Recognition and Privacy in the Age of Augmented Reality

Alessandro Acquisti*, Ralph Gross†, and Fred Stutzman‡

1 Introduction[1]

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 0.026 [1]. By 2010, the best recognizer scored 0.003 [2]—an improvement of more than two orders of magnitude in just over 10 years.

In 2000, of the approximately 100 billion photographs shot worldwide [3], only a negligible portion found their way online. By 2010, 2.5 billion digital photos *a month* were uploaded by members of Facebook alone [4]. Often, those photos showed people's faces, were tagged with their names, and were shared with friends and strangers alike.

This manuscript investigates the implications of the convergence of those two trends: the increasing public availability of facial, digital images; and the ever-improving ability of computer programs to recognize individuals in them.

In recent years, massive amounts of identified and unidentified facial data have become available—often publicly so—through Web 2.0 applications. So have also the infrastructure and technologies necessary 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 [5, 6], 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. Consider Facebook, for example. Most active Facebook users (currently estimated at 1.35 billion monthly active users worldwide [7], with over 250 billion photos uploaded photos [8]) use photos of themselves as their primary profile image. These photos are often identifiable: Facebook has pursued a 'real identity' policy, under which members are expected to appear on the network under their real names under penalty of account cancellation [9]. Using tagging features and login security questions, Facebook has encouraged 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 un-

*Heinz College, Carnegie Mellon University, Pittsburgh, PA, mailto:acquisti@andrew.cmu.edu.
†Heinz College, Carnegie Mellon University, Pittsburgh, PA, mailto:rgross@andrew.cmu.edu.
‡Eighty Percent Solutions, mailto:fred@fredstutzman.com.

[1]An earlier version of this work was presented at BlackHat 2011, under the title "Faces of Facebook: Privacy in the Age of Augmented Reality." This paper provides the details of the experimental design and results.

der Facebook's own Privacy Policy.[2] Many members also make those photos searchable from outside the network via search engines. Similarly, LinkedIn profiles—which are almost always associated with members' actual 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, including some sensitive ones, where members use pseudonyms to protect their privacy. Pseudonyms are common on photo sharing sites such as `flickr.com` or `tumblr.com`; on dating sites such as `match.com` or `manhunt.com`;[3] 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 also those of the strangers we walk by on the street. A person's face is the veritable conduit between her 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. Our results show 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, and then inferring additional and sensitive information about them. In doing so, the results highlight potential implications arising from the convergence of face recognition technology and increasing online self-disclosure, and raise questions about the future of privacy in an "augmented" reality world, in which online and offline data may seamlessly blend.

2 Background

Computer face recognition has been an active research topic for over forty years [11], alternating exciting breakthroughs with the recurrent realization that computers' successes at recognizing people remain limited under real world conditions [12]. In recent

[2] At the time of writing Facebook's Data Use Policy classifies the following as "Information that is always publicly available": name, profile pictures and cover photos, networks, gender, username and user ID at `http://www.facebook.com/policy.php`.

[3] In 2010, Manhunt raised privacy concerns by making changes that made it "easier for people to see profiles without being members" [10].

years, however, the accuracy of face recognizers has improved so consistently that the technology has found its way into end-user products, and in particular Web 2.0 services. With the acquisition of Neven Vision, Like.com, and then PittPatt, Google has started using facial recognition in applications such as Picasa, helping users organize photos according to the individuals they depict [13]. Apple's iPhoto has employed face recognition to identify faces in a person's album since 2009 [14]. After licensing, and then acquiring, Face.com's technology, Facebook has started employing face recognition to suggest "tags" of individuals found in members' photos [15].

The application of face recognition technologies in consumer software, however, has proved a contentious area. With the launch of Google Glass in 2013, the company developed a policy specifically banning facial recognition apps from the Glass marketplace [16], but not after developers at a hackathon developed a facial recognition app called MedRef [17]. In fact, developers have attempted to push the boundaries of identified face recognition with a variety of commercial apps, including Polar Rose, Klik, FaceLock, and FaceVault [18]. In 2014, the NameTag app was launched as the first commercial app that identifies individuals based on online profiles, including those hosted on sites such as Plenty of Fish and OKCupid [19]. The developers of the NameTag app were also putting pressure on Google to allow the app onto Google Glass, stating that "Google will eventually reconsider."[4] Of course, questions about the efficacy of commercial apps hinge not only on platform policies, but on access to databases with identified faces [20]. With partners like Plenty of Fish and OkCupid, the NameTag application seems to be moving in the direction of establishing such a platform.

Business executives and policymakers also had to adapt to the fast growth and adoption of face recognition technologies. In 2013, Facebook updated its policies to address the social network site's use of face recognition software; the move drew the attention of privacy regulators worldwide [21]. The software employed by Facebook, "DeepFace," has recently achieved near-human accuracy for identifying matched faces [22]. To address the privacy risks associated with face recognition technologies, regulatory agencies have formed working groups to better understand the implications of the technology. In the United States, the FTC has requested comments and held workshops related to face recognition technology [23]. More recently, the Department of Commerce initiated a multi-stakeholder process regarding commercial applications of the technology [24]. EU data regulators have also set up a working party on face recognition, issuing guidelines for mobile and online service providers [25].

Providers of face recognition services often address the concerns of policy makers and society at large by proactively limiting the scope and features of their systems, and by highlighting users' control over the recognition process. For instance, a smart phone "augmented ID" application by Swedish startup Polar Rose 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" [26]. Similarly, before being acquired by Facebook, Face.com had developed face recognition services for Facebook users, but stressed that "if you choose to hide your Facebook tags,

[4]Nametag's website, at `http://www.nametag.ws/`, accessed on March 15, 2014.

[their] services will get blocked out when attempting to recognize you in photos" [27].

It is possible, however, that the genie may be already out of the bottle: this manuscript investigates whether publicly available data and publicly accessible technologies can be used by third parties (researchers, peers, merchants) for large scale individual re-identification. In so doing, our work relates to a number of recent manuscripts in the field of face recognition. In particular, [28] gauges the maturity of face recognition technology in matching real-life, low quality face images of uncooperative subjects; while the social media photos we used in our experiments were actually willingly uploaded by the subjects, their quality—similar to that of the images in [28]—was often poor, due to resolution, pose, or occlusion. [29] considers the problem of learning from ambiguously labeled data, such as a photo from a social networking site with multiple faces in it, with a caption that does not clarify which names goes with each face. Similar to the proof of concept application we designed for one of the experiments described herein, [30] presents a system for face augmentation on mobile devices and trains it using Facebook images. In our study, however, the application was designed to link and re-identify users across different contexts (such as online and offline, or different online services) and infer sensitive information about them. [31] observes that personal photos are being captured in digital forms on social media platforms at an accelerating rate, and that the resources, structures, and contexts of online social networks may be leveraged to facilitate and improve face recognition accuracy (relatedly, [32] considers how the fusion of multiple biometrics can improve the performance, accuracy, and scope of biometrics authentication and identification). [33] investigates automated face recognition on a social network using a framework based on "loopy" belief propagation, which leverages information about the social graph of the subjects.

3 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.

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 online dating services (over 10 million Americans were estimated to be a member of one in 2006, and the number was reported as growing in 2008 [34]) 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 ([35] estimated in 2005 that 89% of Facebook profiles on a campus network were identified with the owner's actual name). In addition, many Facebook members leave their profiles indexable by search engines.

Our goal was to estimate how many "matches" we could find using available data and face recognition tools between the set of FB members and the set of DS members in that city. We define a match as a linkage between an identified face of a subject on FB and the unidentified face of the same subject on DS. A match, in other words, makes it possible to identify an up-till-then anonymous DS user.

We used Google API to search FB profiles of users likely to be located in the city. Since FB does not allow to search directly for all users within a given geographical area, our search strategy consisted of 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. (Naturally, the limitation of this strategy is that it produces only a noisy approximation of the set of FB users in the city: profiles of FB members who live in the city may not appear in the searches, while profiles of members who do not live in the city may appear.) Using this strategy, we identified 277,978 FB profiles of users potentially located in 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; [36]) to the set of photos found by 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 faces formed our target 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 made available by the dating site. We queried the DS for all profiles located within 50 miles from a ZIP code approximately centered in the city's Metropolitan Statistical Area. Of the 18,550 DS profiles we thus discovered, we then kept 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 query set.[5]

3.1 Results

We ran the PittPatt recognizer to find matches between the DS and FB sets. The version of PittPatt we used 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.[6] 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

[5]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.

[6]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.

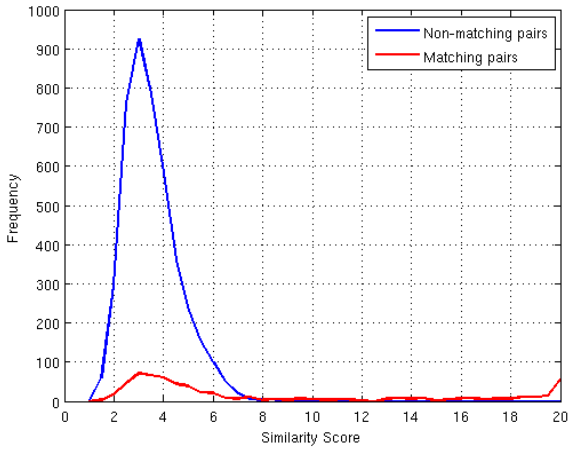

Figure 1: Experiment 1: Distribution of PittPatt scores across all pairs, as function of the human graders' evaluation.

"Definitely the same person" to "Definitely not the same person").[7] To ensure reliability, we retained only the 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 coders considered a pair a match but the coder considered it as non-match, or vice versa). Each pair was graded by multiple coders, with no fewer than five coders per pair.[8] We classified as likely matches pairs that were graded by at least two-thirds of the coders as either a definite or likely match. Those represented 369 of 5,818 profiles in our target set, or about 6.3%. Including pairs that the *majority* of coders 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 coder 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 of our experiment imply that about 1 out of 10 DS members could be identified starting from search engine searches of Facebook profiles in the same geographical area. This analysis involved over 500 million face pairs comparisons, and took

[7]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 our experiment did not have a known ground truth by design: DS members protect their privacy by not revealing their identities. Clearly, human coders' scores do not constitute or replace a ground truth (in fact, computer face recognizers can outperform humans under certain conditions [37]). Nevertheless, human coders' scores helped us interpret PittPatt's results.

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

about 15 hours (roughly 0.00019 seconds per pair, or about 21 seconds per DS profile). For comparison, the human coders took on average 14 seconds to review each *pair*. If a single individual had to grade all 500 million pairs, the task would have required almost 2 million hours to complete.

3.2 Discussion

Experiment 1's results may be optimistic in some ways (for instance, human coders may have considered a match a pair of photos of individuals who look similar but are not the same person), but conservative in various other ways. Specifically, our approach to re-identifying pseudonymous DS users was conservative because we accessed FB identified photos only through a search engine (that is, without logging onto the social network); hence, we used one single identified photo per potential target subject. Furthermore, we tested 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. In addition, our query and target datasets did not fully overlap (not every DS member was also a FB member), and our search patterns for DS and FB profiles in the city differed, making it likely that our DS and FB sets 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 increased ability of recognizing DS members.

4 Experiment 2: Offline Re-identification

Experiment 2 extended Experiment 1 in various directions. Whereas in Experiment 1 we explored online-to-online re-identification, Experiment 2 focused on online-to-offline re-identification. That is, we identified individuals in the physical world using pictures and data from the social network site. Experiment 2 also used multiple photos for re-identification, providing us information on 1) how re-identifiability increases when using more than a single photo for both query and target subjects, and 2) the value of considering a *set* of high-scoring matches found by the face recognizer, rather than the single best match. Instead of focusing on a single photo as we did in Experiment 1, we considered a scenario where the entity interested in re-identifying a stranger will peruse a sorted list of the best-possible matches found by face recognition software, reducing an arbitrarily large set of possible matches to a sufficiently small set that a human can easily evaluate.

During Experiment 2, 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 an experiment. The subjects were asked to sit in front of a laptop equipped with a 1.4 megapixel webcam for the time necessary to have three images taken (one frontal, and

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

two with the subject's face slightly tilted towards either side). Then, each subject was asked to complete a short survey on a second laptop. While the subject was completing her survey, their pictures were uploaded to a cloud computing cluster and matched against a database of photos from a social network site.

The images we used for re-identification purposes had been mined from the profiles of members of the college's Facebook network. We used a profile in the college network to analyze publicly available information about the friend connections of the network members; traversing their graph, we identified 27,371 profiles as members of the network. Of these profiles, 82.9% contained one image (22,703), 13.8% contained multiple images (3,765), and for 3.3% we were not able to access any images (903), for a total of 262,130 images.[9] We ran the face detector on these images; the total number of faces found across all images was 353,234. The detector found at least one face in 43.1% and exactly one face in 36.6% of the *main* profile images of all profiles; 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. Therefore, our re-identification dataset contained 353,234 images against which we compared the webcam photographs.

4.1 Results

We conducted Experiment 2 during two week days in the Fall semester, collecting 93 unique subject sessions. Ninety-two percent of subjects were students at the college where the experiment took place, and all but one had a FB profile. However, only 65.17% of the subjects were sure to be members of the college FB *network*; 10.11% were not members, and 23.6% were not sure. Due to our profile search strategy, this implies that we could expect, at best, to identify no more than about 89% of our

[9]The average number of images per profile was 9.9; considering only profiles with multiple images, the average number was 63.6.

subjects. Furthermore, 83.15% of subjects reported that their main profile image was a photo of themselves. Almost one of two subjects (47.19%) believed that they had *not* made their main profile photos available to everyone else on FB—a misconception, considering that since 2009 Facebook made all primary profile photos public.

We evaluated the results of the face recognition experiment using two approaches.

The first approach was similar to the evaluation approach used in Experiment 1. Three independent graders evaluated the results, comparing the photos of subjects taken with the webcam during the experiment to the ranked list of the 10 best-matching FB profile images produced by the face recognizer, and coding each pair as a match or non match. A match between a subject's image 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. Using a conservative measure of success (all graders agreeing on a match), the recognizer found a FB profile photo that matched the photo of an experimental subject for 31.18% of the subjects. Figure 2 shows an example of a successful match between an image of a subject taken in the foyer of the college building and a photo found online depicting the same individual. It took on average 2.9 seconds of pure computation time to find the set of highly-scored matches for any given subject.

The second approach we used to evaluate the results took place in real time, during the experiment itself. When a subject reached the third page of their survey, he/she would find it populated with a sorted list of social network photos that the face recognizer had ranked as the highest-scoring matches against either of the three images taken of the subject. For each photo, the subject was asked to indicate, directly on the survey, whether or not they recognized themself in the photo. Subjects identified themselves 33.3% of the time in matches found by PittPatt—a result very similar to the one obtained under the first evaluation approach we presented above. However, we should note that these real-time comparisons were completed using a smaller gallery of target images than the over 350,000 we used for the first evaluation approach.

Extrapolating, Experiment 2 suggests that the identity of about one third of subjects walking by the campus building could be inferred in a few seconds combining social network data, cloud computing, and a facial photograph taken by a webcam.

5 Experiment 3: Sensitive Inferences

Experiment 3 was a proof-of-concept test of the potential of online self-disclosures and face recognition technologies to create linkages between the offline and the online worlds. In conducting this experiment, we 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 this question involved exploiting a chain of inferences in a process of data "accretion" [38]. First, face recognition links an unidentified subject (e.g., the face

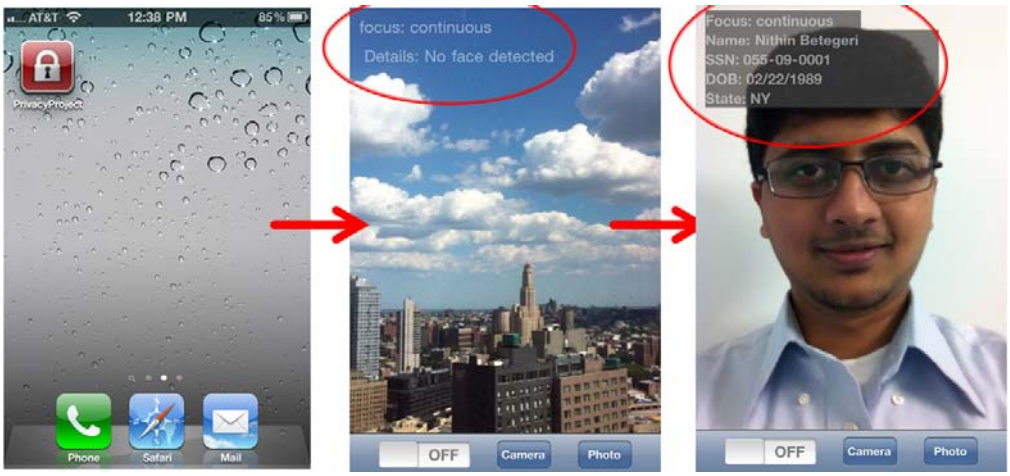

Figure 3: Experiment 3: Screenshots from the real-time mobile phone application (personal information has been replaced with fake data).

of a man on the street) to a record in an identified database (the subject's photo—and profile—on Facebook or LinkedIn). Once the link to the identified database 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) can in turn be linked to the subject. This information can then be supplemented through queries to data aggregators such as `Spokeo.com` by querying the subject's name. Finally, with the application of data mining and statistical re-identification techniques, online information discovered can be used to make additional, sensitive inferences about the individual (such as sexual orientation [39] or Social Security numbers [40]), 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."

More specifically, in Experiment 3 we attempted to predict the interests and Social Security numbers of the subjects who took part in Experiment 2. The subjects' interests were obtained from the subjects' FB profiles — which were found based on the photos matched by the face recognizer in Experiment 2. The subjects' Social Security numbers were predicted combining the subjects' demographics (from FB profiles) with the SSN prediction algorithm described in [40]. The algorithm combines individuals' dates and locations of birth with data publicly available from the Death Master File to produce an estimated SSN.

Experiment 3 consisted of the following steps:

- We focused on the highest-ranked matches found by the recognizer for each subject in Experiment 2, in order to infer the most likely FB profiles associated with that source photo.

- From the profiles that made such information available, we then obtained 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,[10] which [40] has shown to be highly correlated with the likely date of SSN application. We used the demographic information gathered in the previous step as input to the algorithm described in [40], and statistically predicted the most likely SSNs assigned to the subjects.

- Finally, we invited by email the subset of Experiment 2's subjects who had been correctly identified by the recognizer as the top-ranked templates (faces) in Experiment 2 to participate in a new survey. In this new survey, we 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 aggregate form, without us being able to link a given answer back to individual survey participants—thus preserving the subjects' privacy.

5.1 Results

Of the subjects who participated in Experiment 2, the 29 that we identified using face recognition—and for whom we found publicly available demographic information—were invited by email seven months after the initial experiment to participate in a new online survey. Participants were offered a $10 gift card, as well as the opportunity to win a $50 gift card in a raffle. Eighteen subjects responded to the invitation and completed the survey (a response rate of 62%). For each subject, we 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 the first five digits of the target subject's SSN. As discussed in [40], knowledge of the first five digits of a target victim's SSN is sufficient for brute force identity theft attacks. We correctly predicted the subjects' first five digits for 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 small, the results are better than random chance by several orders of magnitude: the probability of correctly guessing the first five digits of a 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 the results. Before being presented with the actual predictions and our questions about them, the subjects who participated in the survey 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 ("Very comfortable") to 7 ("Very uncomfortable"), the modal scores across the subjects were, respectively, 6 for

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

predicted interests (μ: 5.11) and 7 for predicted SSNs (μ: 6.17). In the open-ended question at the end of the survey, *after* their predicted interests and SSNs had been presented on the screen, some subjects expressed additional concerns, including: "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 [*sic*]."

5.2 Face Recognition and Augmented Reality

Experiment 3 provided a proof-of-concept test of the ability to infer personally predictable, sensitive information through face recognition. Our test was asynchronous: target photos to be matched against the subjects' live images 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 demo iPhone app that captures the image of a person and then overlays her predicted name and SSN on the screen. The application is an example of augmented reality [41], with offline and online data blending together through cloud computing.

In the application, different components interact on a cloud server, imitating in real time what Experiment 3 did in asynchronous fashion in a controlled experimental environment. The application transmits the captured image of someone's face to a face recognition server that contains a database of target photos previously downloaded from FB profiles. The face recognizer locates and encodes the face in the captured image and calculates its matching scores against each of the target templates (faces) in the database. The highest-matching template is selected, and the associated profile from which the photo was mined is tentatively picked as the presumptive profile of the subject.[11] If available from the identified database, the person's presumptive name is then inferred. Another script then uses the name to query, in real time, 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.[12] The demographic information thus inferred is fed into [40]'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 phone — which displays the results on the screen on the person's face (Figure 3).

[11] We had designed a more sophisticated algorithm for asynchronous usage that leverages social network data to predict the most likely profile's *name* of a person, starting from the photo in which the matching face was found, and the profile from which that photo was mined. However, we did not develop a real time version of that algorithm to synch with our iPhone application.

[12] 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 smartphone.

6 From Face Recognition to Personally Predictable Information

We have presented a series of experiments illustrating the ability to re-identify individuals online and offline, and infer sensitive information about them, by combining facial recognition and social media data. The experiments we described are merely examples of combinations of software and data and the many possible inferences that those combinations make possible. Key to that process is the "accretion" of increasingly sensitive information, starting from an anonymous face, by combining data from different databases.

The experiments also illustrate a "democratization" of surveillance: while face recognizers were for a long time the domain of governments and large corporations, the availability of Web 2.0 data is rendering peer-to-peer face recognition possible and cost effective. Still, the frequent acquisitions of face recognition start-ups by large Silicon Valley companies provide evidence of the significant business interest in this space. In light of that, two possible future 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.

On the other hand, various limitations *currently* affect the scalability of the processes we described. Mass face recognition is limited by the availability of (correctly) identified facial images, which is itself a 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 a function of the quality of subjects' photos (Experiment 1 and 2 relied on frontal photos with favorable lighting, either uploaded by the subjects themselves to their dating site profiles, or captured by the researchers on campus). It is also a 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.

Technological, commercial, and social trends, however, make it plausible to argue that the constraints and limitations we just espoused will keep loosening over time.

Due to increasingly public default visibility settings in social network sites, evolving social norms toward self disclosures, and the existence of search engines that index social

network site data, identified facial images will arguably become publicly available for an increasing numbers of individuals.[13] Once an identity is found, demographic information may be available from multiple sources (voter registration lists, people search services, social network sites [40]). Furthermore, all technical components of the system described here are constantly improving, implying that better results will be obtained in lesser time and at a lower costs. As an example, the largest compute unit offered by Amazon's AWS service at the time of writing has more than 30 times the computing power of the configuration used at the time of our experiments, suggesting that what took 15 hours previously could now be done in less than 30 minutes. This enables running facial searches on larger sets of source subjects. Similarly, the quality of cameras likely to be used for images found on social network sites keeps improving. As an example, the camera found in Apple's popular iPhone evolved from using a 2 megapixel sensor (no flash and aperture of f/2.8; original iPhone and iPhone 3G) to an 8 megapixel back-illuminated sensor (True Tone flash and aperture of f/2.2; iPhone 6) with vastly improved low-light performance. Cooperative subjects may not be needed for frontal pictures once wireless cameras are cheaply deployed in—for instance—glasses, instead of mobile phones. Finally, face recognizers will keep improving in terms of accuracy as researchers increasingly focus on more challenging image conditions [42].

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 [43]. 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 [44]. 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.

Finding a technological or policy solution that can balance the benefits and risks of peer-based face recognition is not easy. Google's former CEO, Eric Schmidt, once observed that, in the future, young individuals should be entitled to change their names to disown youthful improprieties [45]. It is much costlier, however, to change someone's *face*. Blurring of facial images in databases, k-anonymization of photos, or opt-ins, are ineffectual tools when re-identification can be achieved through already publicly available data. Although the results of one of our surveys (the one conducted for Experiment 3) suggest that most individuals loathe the possibility of being identified by strangers on the street, many of them nevertheless share identified photos online that make this sort of identification possible. Notwithstanding Americans' resistance to a

[13] As a matter of fact, tagging one's self *and* others in pictures has become socially acceptable. Before Experiment 2, one of our subjects claimed, 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.

Real ID infrastructure, as consumers of social network sites we have already 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, 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 and under grant #1327992, from the US Army Research Office under contract # DAAD190210389, from the Carnegie Mellon Berkman Fund, from Carnegie Mellon Cylab, and from the IWT SBO Project on Security and Privacy for Online Social Networks (SPION). The authors thank Nithin Betegeri, Aravind Bharadwaj, Varun Gandhi, Markus Huber, Aaron Jech, 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, Sonam Samat, Soumya Srivastava, Thejas Varier, and Narayana Venkatesh for additional assistantship.

References

[1] Phillips, P. J., Scruggs, W. T., O'Toole, A. J., Flynn, P. J., Bowyer, K. W., Schott, C. L., and Sharpe, M. (2007). FRVT 2006 and ICE 2006 large-scale results, *National Institute of Standards and Technology, NISTIR*, 7408.

[2] Grother, P. J., Quinn, G. W., and Phillips, P. J. (2011) Report on the evaluation of 2D still-image face recognition algorithms, National Institute of Standards and Technology. NIST Interagency Report 7709.

[3] Hilbert, M. and López, P. (2011). The world's technological capacity to store, communicate, and compute information, *Science*, 332(6025):60.

[4] Facebook. (2010). Faster, simpler photo uploads. Available at `http://www.facebook.com/blog.php?post=206178097130`.

[5] Sweeney, L. (1997). Weaving technology and policy together to maintain confidentiality, *Journal of Law, Medicine and Ethics*, 25(2–3):98–110.

[6] Narayanan, A. and Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In *Proceedings of the IEEE Symposium on Security and Privacy*, Oakland, CA. 111–125.

[7] Facebook. (2014). Facebook Newsroom – Company Info. Available at `http://newsroom.fb.com/company-info/`.

[8] Facebook, Ericsson, and Qualcomm (2013). A focus on efficiency. Available at `https://fbcdn-dragon-a.akamaihd.net/hphotos-ak-prn1/851575_520797877991079_393255490_n.pdf`.

[9] Branigan, T. (2010). Facebook's 'Real Name' policy attacked by Chinese blogger, *The Guardian*, March 9.

[10] Cassels, P. (2010). Manhunt policy change raises questions anew about Internet privacy, *Edge Boston*, November 22.

[11] Kelly, M. (1970). Visual identification of people by computer. Tech. report AI-130, Stanford AI Project, Stanford, CA.

[12] Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. (2003). Face recognition: A literature survey, *ACM Computing Surveys*, 35(4):399–458.

[13] Google. (2011). Add name tags in Picasa. Available at `http://picasa.google.com/support/bin/answer.py?answer=156272`.

[14] Apple. (2013). iPhoto '09 & iPhoto '11: Improving face recognition results. Available at `http://support.apple.com/kb/ht3442`.

[15] Facebook (2010). Making Facebook photos better. Available at `http://www.facebook.com/blog.php?post=403838582130`.

[16] Velazco, C. (2013). Google won't approve Glass apps that recognize people's faces...for now, *Techcrunch*, May 31.

[17] Kelly, S. M. (2013). Facial recognition comes to Google glass, *Mashable*, May 13.

[18] Tatheer, (2013). Top 5 face recognition apps for iPhone in 2013, *Maypalo*, September 15.

[19] Starr, M. (2014). Facial recognition app matches strangers to online profiles, *CNET*, January 7.

[20] Hartzog, W. and Selinger, E. (2014). I see you: The databases that facial-recognition apps need to survive, *The Atlantic*, January 23.

[21] Tsukayama, H. (2013). Facebook facial recognition policy draws attention from German privacy regulator, *Washington Post*, August 30.

[22] Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. Available at `http://static1.trouw.nl/static/asset/2014/deepface_4361.pdf`. Tech. Report.

[23] Federal Trade Commission. (2011). FTC announces agenda panelists facial recognition workshop. Available at `http://www.ftc.gov/news-events/press-releases/2011/11/ftc-announces-agenda-panelists-facial-recognition-workshop`.

[24] National Telecommunications and Information Administration. (2014). Privacy multistakeholder process: Facial recognition technology. Available at `http://www.ntia.doc.gov/other-publication/2014/privacy-multistakeholder-process-facial-recognition-technology`.

[25] eForum. (2012). EU privacy watchdog sets out facial recognition principles. Available at `http://www.eu-forum.org/item/51-eu-privacy-watchdog-sets-out-facial-recognition-principles`.

[26] Dillow, C. (2010). Augmented identity app helps you identify strangers on the street, *Technology Review*, February 23.

[27] Face.com. (2011). Faq. Previously available at `http://developers.face.com/docs/faq/`.

[28] Klontz, J. C. and Jain, A. K. (2013). A case study of automated face recognition: The Boston marathon bombings suspects, *Computer*, 46(11):91–94.

[29] Chen, Y -C., Patel, V. M., Pillai, J. K., Chellappa, R., and Phillips, P. J. (2013). Dictionary learning from ambiguously labeled data. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. 353–360.

[30] Dantone, M., Bossard, L., Quack, T., and Van Gool, L. (2011). Augmented faces. In *IEEE International Workshop on Mobile Vision (ICCV 2011)*. IEEE. 24–31.

[31] Stone, Z., Zickler, T., and Darrell, T. (2010). Toward large-scale face recognition using social network context. In *Proceedings of the IEEE*, 98(8):1408–1415.

[32] Bhanu, B. and Govindaraju, V. (2011). *Multibiometrics for Human Identification*. Cambridge University Press.

[33] Wu, T., Phillips, P. J., and Chellappa, R. (2013). Propagation of facial identities in a social network. In *Proceedings of the Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE. 1–8.

[34] Gibbs, J., Ellison, N., and Lai, C. (2011). First comes love, then comes Google: An investigation of uncertainty reduction strategies and self-disclosure in online dating, *Communication Research*, 38(1):70–100.

[35] Gross, R. and Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of the ACM Workshop on Privacy in the Electronic Society*. ACM. 71–80.

[36] Nechyba, M. and Schneiderman, H. (2007). Pittpatt face detection and tracking for the CLEAR 2006 evaluation, *Multimodal Technologies for Perception of Humans*, 161–170.

[37] Phillips, P. J. and O'Toole, A. J. (2014). Comparison of human and computer performance across face recognition experiments, *Image and Vision Computing*, 32(1):74–85.

[38] Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization, *UCLA Law Review*, 57:1701.

[39] Jernigan, C. and Mistree, B. (2009). Gaydar: Facebook friendships expose sexual orientation, *First Monday*, 14(10).

[40] Acquisti, A. and Gross, R. (2009). Predicting Social Security numbers from public data, *Proceedings of the National Academy of Science*, 196(27):10975–10980.

[41] Azuma, R. et al. (1997). A survey of augmented reality, *Presence-Teleoperators and Virtual Environments*, 6(4):355–385.

[42] Beveridge, J. R., Phillips, P. J., Bolme, D. S., Draper, B. A., Given, G. H., Lui, Y. M., Teli, M. N., Zhang, H., Scruggs, W. T., Bowyer, K. W., et al. (2013). The challenge of face recognition from digital point-and-shoot cameras. In *Proceedings of the 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. 1–8.

[43] Zimmer, M. (2009). Photo finder: Automated facial recognition on Facebook. Available at `http://www.michaelzimmer.org/2009/03/25/photo-finder-automated-facial-recognition-on-facebook/`.

[44] Acquisti, A. (2004). Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the ACM Conference on Electronic Commerce (EC '04)*. 21–29.

[45] Jenkins, H. W. J. (2010). Google and the search for the future. *The Wall Street Journal*, August 14.