Faces of Facebook: Privacy in the Age of Augmented Reality

Alessandro Acquisti, Ralph Gross, Fred Stutzman

Heinz College & CyLab
Carnegie Mellon University

PLEASE NOTE: DRAFT VERSION
Final version to be presented at BlackHat USA on August 4,2011

Black Hat 2011

Background

- Computer face recognition has been around for a long time (e.g.: Bledsoe, 1964; Kanade, 1973)
- Computers still perform much **worse than humans** when recognizing faces
- However, automatic face recognition has kept improving, and has started being used in actual applications
  - Especially in security, and – more recently – Web 2.0

Background

- Face recognition in Web 2.0
  - Google has acquired Neven Vision, Riya, and PittPatt and deployed face recognition into Picasa
  - Apple has acquired Polar Rose, and deployed face recognition into iPhoto
  - Facebook has licensed Face.com to enable automated tagging
- *So, what is different about this research?*

What is different: The convergence of various technologies (1/2)

- Increasing **public self-disclosures** through online social networks; especially, photos
  - In 2010, 2.5 billion photos uploaded by Facebook users alone *per month*
- **Identified** profiles in online social networks
  - Individuals using their real first and last names on Facebook, LinkedIn, Google+, etc.
- Continuing **improvements** in face recognition accuracy
  - In 1997, the best face recognizer in FERET program achieved a false reject rate of 0.54 (at false accept rate of 0.001)
  - By 2006, the false reject rate was down to 0.01

What is different: The convergence of various technologies (2/2)

- Statistical **re-identification:** data mining allows surprising, sensitive inferences from public data
  - US citizens identifiable from zip, DOB, gender (Sweeney, 1997); Netflix prize de-anonymization (Narayanan and Shmatikov, 2006); SSN predictions from Facebook profiles (Acquisti and Gross, 2009)
- **Cloud** computing
  - Makes it feasible and economic to run millions of face comparisons in seconds
- **Ubiquitous** computing
  - Combined with cloud computing, makes it possible to run face recognition through mobile devices – e.g., smartphones

What this implies

- The converge of these technologies is **democratizing surveillance**
  - Not just Web 2.0 face recognition apps limited and constrained to consenting/opt-in users, but…
  - **….a world where anyone may run face recognition on anyone else, online and offline**

Why this matters

- Your face is the **veritable link** between your offline identity and your online identit(ies)
- Data about your face and your name is, most likely, **already publicly available online**
- Hence, face recognition creates the potential for **your face in the street** (or online) **to be linked to your online identit(ies)**, as well as to the sensitive inferences that can be made about you after **blending together offline and online data**

Why this matters

- This seamless merging of online and offline data raises the issue of what **“privacy” will mean in such augmented reality world**
  - Through social networks, have we created a *de facto*, **unregulated “Real ID”** infrastructure?

Our research focus

- Our research investigates the feasibility of combining **publicly available** online social network data with **off-the-shelf** face recognition technology for the purpose of **large-scale, automated, peer-based…**
  1. **individual re-identification**, online and offline
  2. **"accretion" and linkage of online, potentially sensitive, data** to someone's face in the offline world

Key themes in our research

- Democratization of surveillance
- Faces as conduits between online and offline data
- The emergence of PPI: "personally predictable" information
- The rise of visual, facial searches
- The future of privacy in a world of augmented reality

Experiments

- Experiment 1: Online-to-online Re-Identification
- Experiment 2: Online-to-offline Re-Identification
- Experiment 3: Online-to-offline Sensitive Inferences

In a nutshell

Un-Identified DB

- Personal Profiles on Match.com, Prosper.com, etc.
- Photo repositories (e.g., Flickr)
- Open web cams
- CCTVs
- **Your face on the street**
- […]

Identified DB

- Personal Profiles on Facebook.com, Linkedin, etc.
- Govt or corporate databases
- […]

[2]

[3]

- Additional sensitive inferences (e.g. sexual orientation, SSN, etc.)

[4]

[1]

- **Face recognition [1] allows to match a subject in an un-identified DB from data in an identified DB [2]**
- **Once that is done, sensitive data inferred from the un-identified DB [3] can be linked to the identity of the subject in the identified DB [4]**

Experiment 1

- Online to online
- We mined **publicly available images** from online social network profiles to re-identify profiles on one of the most popular dating sites in the US
  - We used PittPatt face recognizer (Nechyba, Brandy, and Schneiderman, 2007) for:
    - Face detection: automatically locating human faces in digital images
    - Face recognition: measuring similarity between any pair of faces to determine if they are of the same person

Experiment 1: Data

- Facebook profiles
  - We downloaded primary profile photos for Facebook profiles from a North American city using a search engine's API (i.e., **without even logging on the Facebook itself**)
  - "Noisy" profile search pattern: Combination of search strategies (current location, member of local networks, fan of local companies/teams, etc.)

Experiment 1: Data

- Dating site profiles
  - Profiles were members of one of the most popular dating sites in the US
  - Members use pseudonyms to protect their identities
  - However, facial images may make members recognizable not just by friends, but by strangers
    - **Unfeasible if done manually** (hundreds of millions of potential matches to verify), but quite **feasible using face recognition + cloud computing**

Experiment 1: Ground truth

- Overlap between our dating site data and Facebook data is inherently noisy (geographical search vs. keywords search)
- We ran two surveys to estimate Facebook/dating site members overlap
- Then, multiple human coders graded matched pairs to evaluate face recognizer’s accuracy

Experiment 1: Results

- One out of 10 dating site's pseudonymous members was identified
- Note:
  - In Experiment 1, we constrained ourselves to using **only a single Facebook** (primary profile) photo, and only considering the **top match** returned by the recognizer
    - However: Because an "attacker" can use more photos, and test more matches, ratio of re-identifiable individuals will dramatically increase
    - **See, in fact, Experiment 2**
  - Also: as face recognizers' accuracy increases, so does the ratio of re-identifiable individuals

Experiment 2

- Offline to online
- We used publicly available images from a Facebook College network to identify students strolling on campus

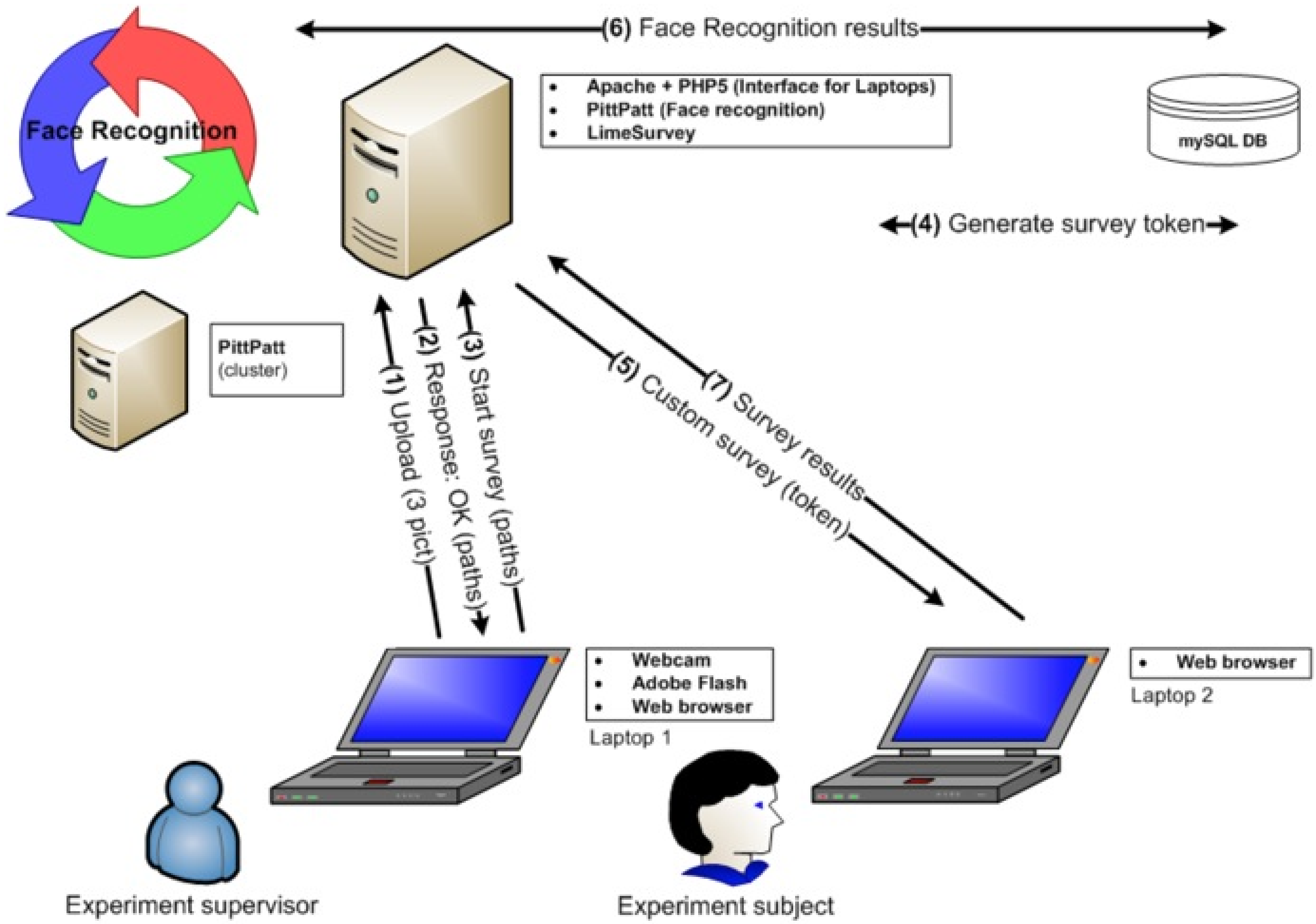
Experiment 2: Data

- College photos
  - We used a webcam to take 3 photos per participant
  - Photos gathered over two days in November

Experiment 2: Process and ground truth

- We ask students walking by to stop and have their picture taken
- Then, we asked participants to answer an online survey about Facebook usage
- In the meanwhile, face matching was taking place on an cloud computing service
- The last page of the survey was populated dynamically with the best matching pictures found by recognizer
- Participants were asked to select photos in which they recognized themselves

Experiment 2: Approach



Experiment 2: Results

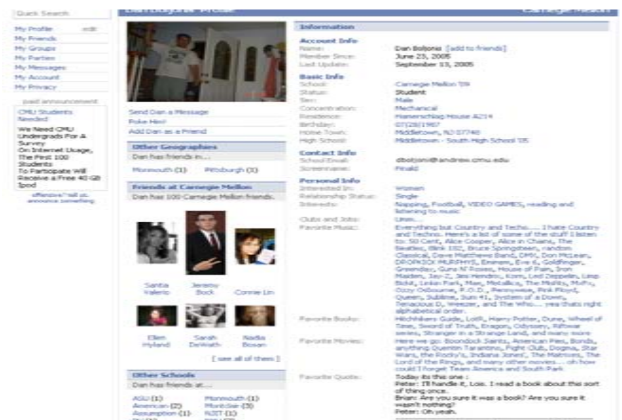
- Roughly one of out three subjects was identified
  - Average computation time per subject: less than three seconds

From Experiment 2 to Experiment 3

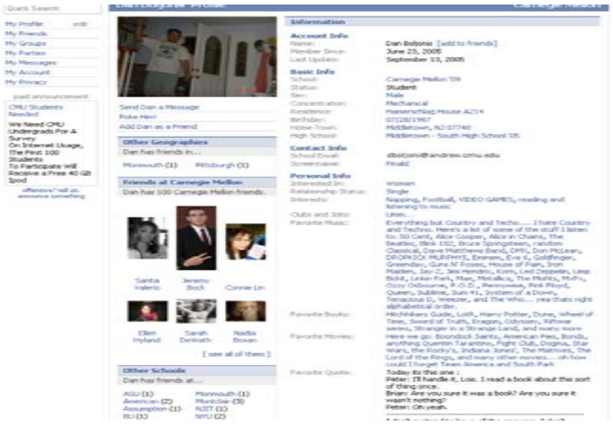
- In Experiment 2 we found the Facebook profiles containing images that matched the facial features of students working on campus
- But: in 2009, we used Facebook profile information to predict individuals' Social Security numbers
  - Acquisti and Gross, Predicting Social Security Numbers from Public Data, *Proceedings of the National Academy of Science*, 2009

What we have shown so far

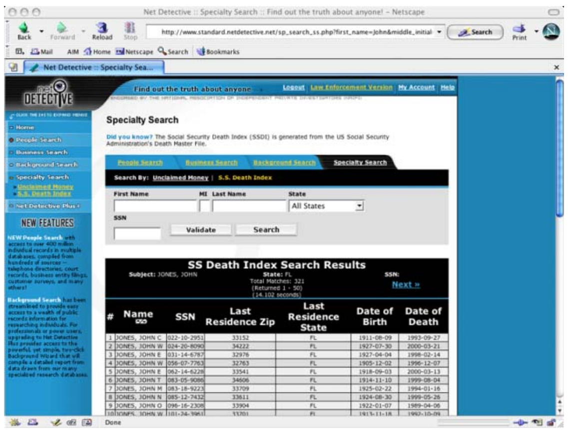
+ pittpatt =



What we had done before (Acquisti and Gross 2009)



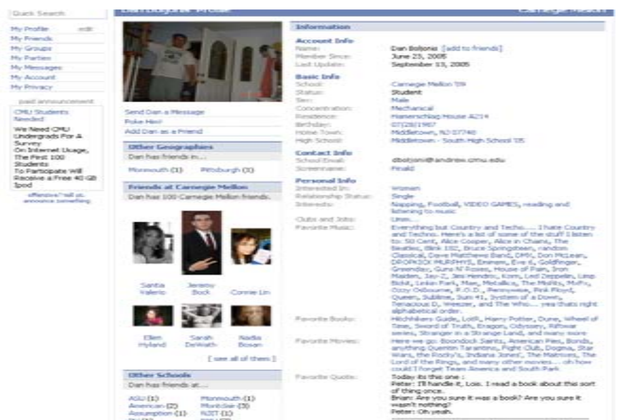
+



= SSN

Can you do 1+1?

+



= SSN

I.e., predicting SSNs from faces

Experiment 3

- Experiment 3 was about predicting personal and sensitive information… from a face
  - We trained an algorithm to automatically identify the most likely Facebook profile owner given a match between the Experiment 2 subjects' photos and a database of Facebook images
  - From the predicted profiles, we inferred names, DOBs, other demographic information, as well as interests/activities of the subjects
    - With that information, we predicted the participants' SSNs
  - We then asked participants in Experiment 2 whom we had thusly identified to participate in a follow-up study

Predicting SSNs from someone's face

- In the follow-up study, we asked participants to verify our predictions about their:
  - Interests/Activities (from Facebook profiles)
  - SSNs' first five digits (predicted using Acquisti and Gross, 2009's algorithm)
    - Note: **last 4 digits are predictable too** (see Acquisti and Gross, 2009). Prediction accuracy varies greatly, as function of state and year of birth, and can be correctly estimated only with **larger sample sizes** that what available in Experiment 3

The Age of Augmented Reality

Source: http://www.director-thailand.com/blog/what-is-augmented-reality

Real time demo

- Our demo smart phone app combines and extends the previous experiments to allow:
  - Personal and sensitive inferences
  - From someone's face
  - In real time
  - On a mobile device
  - **Overlaying information (obtained online) over the image of the individual (obtained offline) on the mobile device's screen**

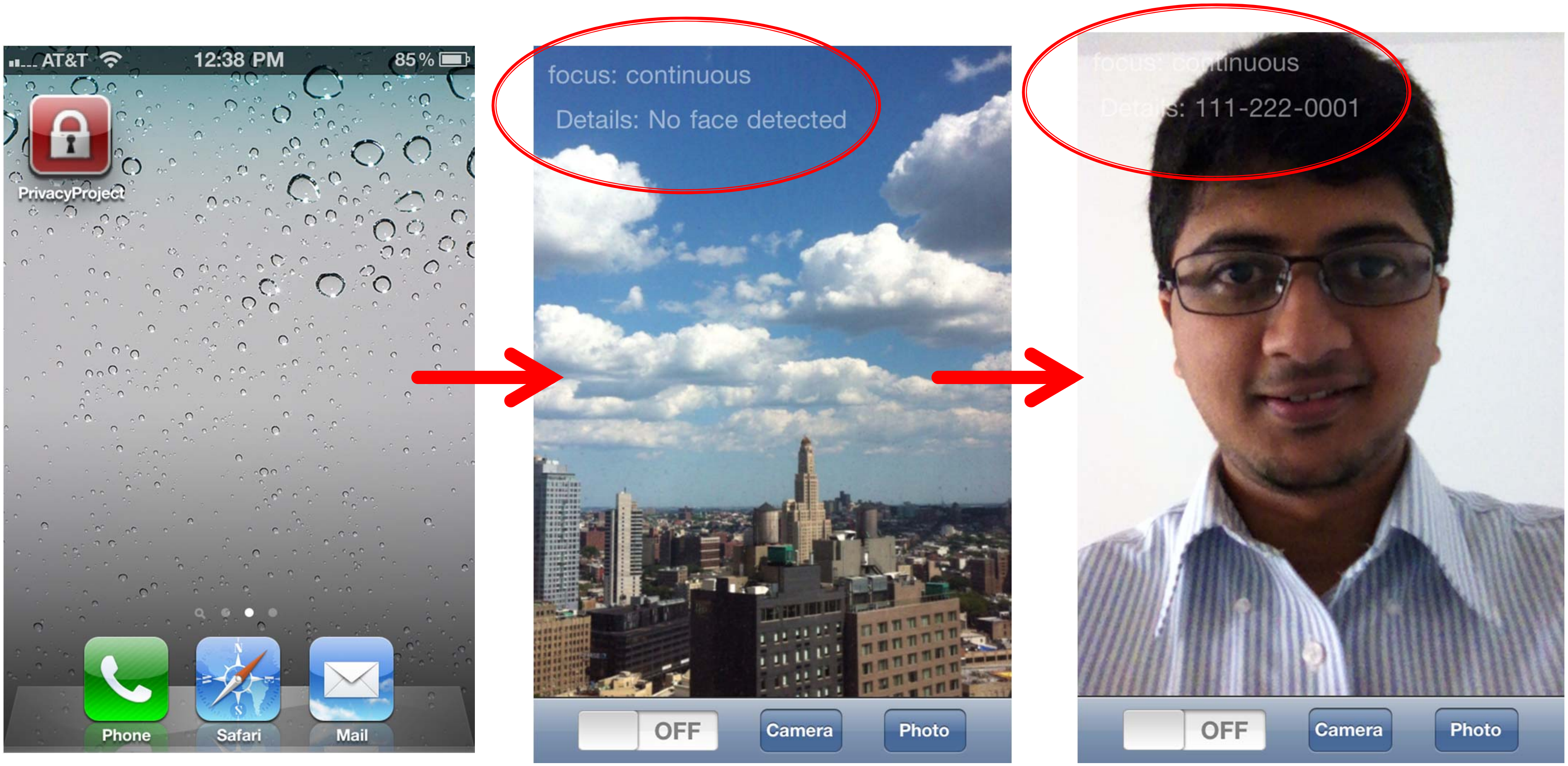
Real time demo

- Sources of online data can be Facebook (to identify someone's name), Spokeo (once someone's name has been identified)…
- … and then, the **sensitive inferences one can make based on that data** (e.g., SSNs, but also sexual orientation, credit scores, etc.)
  - That is: the emergence of **personally predictable information** from a person's face

Data accretion

- Overlaying information (obtained online) over the image of the individual (obtained offline) on the mobile device's screen
  - It's the **"accretion" problem**: "once any piece of data has been linked to a person's real identity, any association between this data and a virtual identity breaks the anonymity of the latter" (Arayanan and Shmatikov, 2007)
  - Or: "Once an adversary has linked two anonymized databases together, he can add the newly linked data to his collection of outside information and [...] unlock other anonymized databases. Success breeds further success" (Ohm, 2010)

Screenshots



Limitations

- Availability of facial images
  - Legal and technical implications of mining identified images from online sources
- Cooperative subjects
  - Face recognizers perform worse in absence of clean frontal photos
  - On the street, clean and frontal photos of uncooperative strangers are unlikely
- Geographical restrictions
  - Experiment 1 focused on City area (~330k individuals). Experiment 2 focused on College community (~25k individuals)
  - As the set of potential targets gets larger (e.g., nationwide), computations needed for face recognition get less accurate (i.e., **more false positives**), and take more time

Extrapolations

- Face recognition of everyone/everywhere/all the time is **not** yet feasible
- **However:** Current technological trends suggest that most current limitations will keep fading over time

Scalability: Availability of images (1/2)

- There exist legal and technical constraints to mining identified images from online sources
- However:
  - Many sources are publicly available (e.g., do not require login, such as LinkedIn profile photos; or can be searched through search engines, such as Facebook primary profile photos: **see Experiment 1**)
  - Face recognition companies are already collaborating with social network sites to tag "billions" of images (e.g., see Face.com recent announcement)
  - Tagging self, and others, in photos has become socially acceptable – in fact, widespread (thus providing a growing source of identified images)

Scalability: Availability of images (2/2)

- As search engines enters the face recognition space, **facial visual searches may become as common as today's text-based searches**
  - Text-based searches of someone's name across the WWW, which are common now, were unimaginable 15 years ago (before search engines)
  - From spidered & indexed html pages, to spidered & indexed photo
    - Google has already announced searches based on image (although not *facial image*) pattern matching
  - The number of Silicon Valley players entering this space in recent months demonstrates the **commercial interest in face recognition**

Scalability: Cooperative subjects

- What we did on the street with mobile devices today (requiring point-and-shoot and cooperative subjects), will be accomplished in less intrusive ways tomorrow
  - Glasses (already happening: Brazilian police preparing for 2014 World Cup)
  - How long before it can be done on…. *contact lenses*?
- Face recognizers will keep getting better at matching faces based on non-frontal images (compare PittPatt version 5.2 vs. version 4.2)

Scalability: Geographical restrictions

- As the set of potential targets gets larger (e.g., nationwide DB of individuals), the computations needed for face recognition get less accurate (more false positives) and take more time
  - However: databases of identified images are getting larger, with more individuals are in them (see previous slides)
  - Accuracy (number of false positives, number of false negatives) of face recognizers steadily increases over time – especially so in last few years
  - Cloud computing clusters will keep getting faster, larger (more memory available==larger target DBs feasible to analyze), and cheaper, making massive face comparisons economical

Implications (1/4)

- Web 2.0 profiles (e.g. Facebook) are becoming *de facto* **unregulated "Real IDs"**
  - See recent FTC's approval of *Social Intelligence Corporation*'s social media background checks
- Great potential for commerce and ecommerce…
  - Imagine "Minority Report"-style advertising…
  - … however, happening much earlier than 2054

Implications (2/4)

- But also: **ominous risks for privacy**
- These technologies challenge our expectations of **anonymity in a digital or a physical crowd**
- Especially risky, because:
  1. We do not anticipate being identified by strangers in the street/online
  2. We do not anticipate the sensitive inferences that can be made starting merely from a face
  3. No obvious solutions without risks of significant unintended consequences

No clear solution

- Opt-in is **ineffective** as protection, since most data is already publicly available
  - E.g., Facebook sets primary profile photos to be visible to all by default, and members to sign up to the network with their real identities

Implications (3/4)

- What **will privacy mean** in a world where a stranger on the street could guess your name, interests, SSNs, or credit scores?
- The coming **age of augmented reality, in which online and offline data are blended in real time**, may force us to reconsider our notions of privacy

Implications (4/4)

- In fact, augmented reality may also carry **deep-reaching behavioral implications**
  - Through natural evolution, human beings have **evolved mechanisms to assign and manage trust in face-to-face interactions**
  - Will we rely **on our instincts, or on our devices**, when mobile devices make their own predictions about hidden traits of a person we are looking at?

Key themes, again

- Democratization of surveillance
- Faces as conduits between online and offline data
- The emergence of PPI: “personally predictable” information
- The rise of visual, facial searches
- The future of privacy in a world of augmented reality

Thank you

- We gratefully acknowledge research support from
  - National Science Foundation under Grant 0713361
  - U.S. Army Research Office under Contract DAAD190210389
  - Heinz College
  - Carnegie Mellon CyLab
  - Carnegie Mellon Berkman Fund

Thank you

- Main RAs: Ganesh Raj ManickaRaju, Markus Huber, Nithin Betegeri, Nithin Reddy, Varun Gandhi, Aaron Jaech, Venkata Tumuluri
- Additional RAs: Aravind Bharadwaj, Laura Brandimarte, Samita Dhanasobhon, Hazel Diana Mary, Nitin Grewal, Anuj Gupta, Snigdha Nayak, Rahul Pandey, Soumya Srivastava, Thejas Varier, Narayana Venkatesh

For more info

- Google: economics privacy
- Visit: http://www.heinz.cmu.edu/~acquisti/economics-privacy.htm
- Email: acquisti@andrew.cmu.edu