When Enough is Enough: Location Tracking, The Fourth Amendment, Mosaic Theory, and Machine Learning

Steven M. Bellovin

(Joint work with Renée Hutchins, Tony Jebara, Sebastian Zimmeck)

LAW?

- I'm a CS professor
- This is a data science class
- So why am I going to talk about law?

PATTERNS AND PREDICTIONS

- Machine learning can find all sorts of patterns
- Some uses of big data are fairly obvious, once we know how to do it
- Some aren't—like shaping legal doctrine
- For example: should the police need a search warrant to track someone's location?

AN OPEN LEGAL QUESTION!

- The Supreme Court has never ruled about tracking *people*
  - The closest they came was in *United States v. Knotts*, 460 U.S. 276 (1983)
  - That was about tracking a drum of chemicals
- They had a chance in *United States v. Jones*, 615 F. 3d 544 (2012)—but punted and issued a ruling on other grounds

THE FOURTH AMENDMENT

"The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized."

Searches do not always require a warrant, but they have to be *reasonable*

SHOULD POLICE NEED A WARRANT FOR GPS TRACKING?

- No: movements are public
  - Police could just follow someone
  - You have no "reasonable expectation of privacy" in public activities
- No: in *Knotts*, the Supreme Court said that putting a beeper on a chemical shipment for three days is ok
  - It tracked movements on public roads

SHOULD POLICE NEED A WARRANT FOR GPS TRACKING?

- Yes: One check on police abuse of their power is economic: they can't afford to trail very many people for a very long time

  ➔ Modern tracking is *much* cheaper

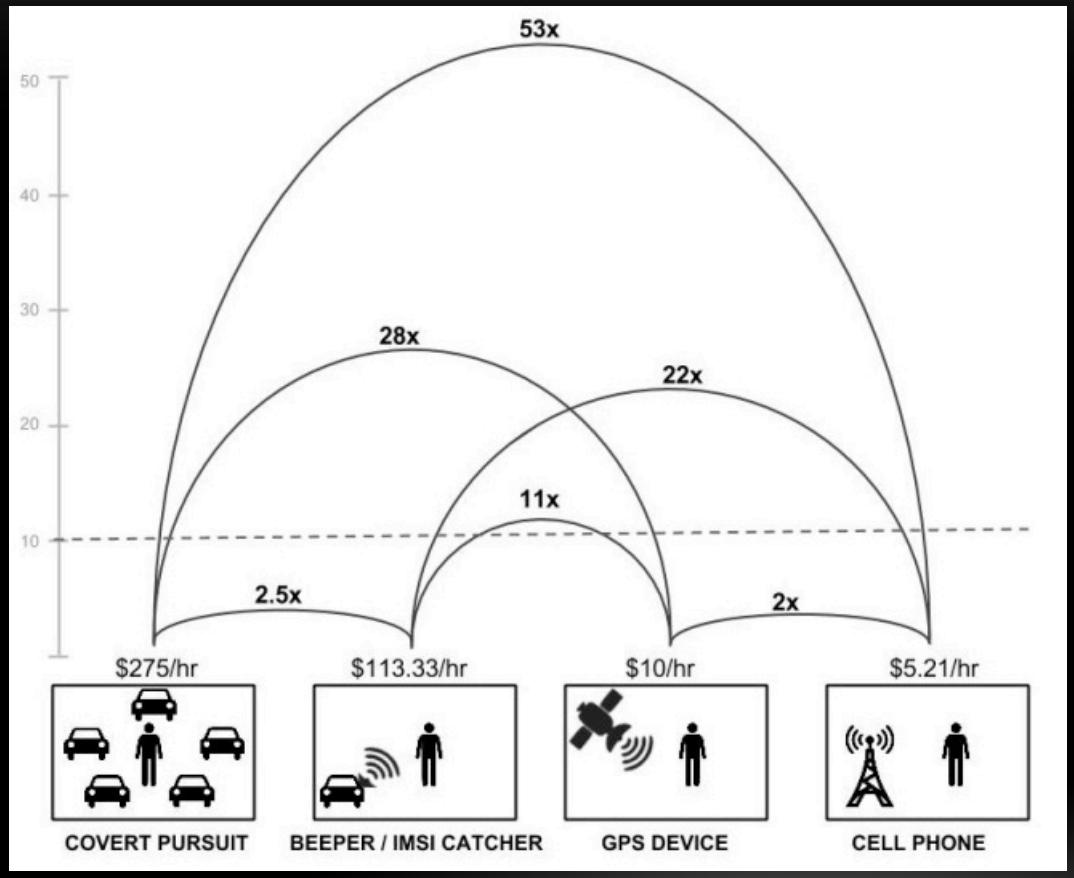
EQUILIBRIUM ADJUSTMENT (KERR)

"When changing technology or social practice makes evidence substantially harder for the government to obtain, the Supreme Court generally adopts lower Fourth Amendment protections for these new circumstances to help restore the status quo ante level of government power. On the other hand, when changing technology or social practice makes evidence substantially easier for the government to obtain, the Supreme Court often embraces higher protections to help restore the prior level of privacy protection."

JUSTICE ALITO IN *JONES*

"In the pre-computer age, the greatest protections of privacy were neither constitutional nor statutory, but practical. Traditional surveillance for any extended period of time was difficult and costly and therefore rarely undertaken. The surveillance at issue in this case—constant monitoring of the location of a vehicle for four weeks—would have required a large team of agents, multiple vehicles, and perhaps aerial assistance. Only an investigation of unusual importance could have justified such an expenditure of law enforcement resources. Devices like the one used in the present case, however, make long-term monitoring relatively easy and cheap."

THE COST OF TRACKING



From Kevin S. Bankston & Ashkan Soltani, *Tiny Constables and the Cost of Surveillance: Making Cents Out of* United States v. Jones, 123 Yale L.J. Online 335 (2014)

SHOULD POLICE NEED A WARRANT FOR GPSTRACKING?

- Yes: One check on police abuse of their power is economic: they can't afford to trail very many people for a very long time

  ➔ Modern tracking is *much* cheaper

- Yes: *Patterns* of movement are very revealing

MOSAIC THEORY

- *Mosaic theory*: a large-enough collection of data points is very, very revealing, and violates "reasonable expectation of privacy"
- It is the *total pattern* of movements that is revealing
  - Law enforcement cannot afford to track (most) people for a month
- But—where do you draw the line? What is "large enough"?

US V. JONES (2012)

- Police attached a GPS tracker to Jones' car for 28 days
- The warrant had expired
- The Supreme Court overturned the conviction 9-0, but on classical Fourth Amendment grounds: a physical intrusion on his car

SOME JUDICIAL SUPPORT FOR MOSAIC THEORY

"Disclosed in [GPS] data . . . will be trips the indisputably private nature of which takes little imagination to conjure: trips to the psychiatrist, the plastic surgeon, the abortion clinic, the AIDS treatment center, the strip club, the criminal defense attorney, the by-the-hour motel, the union meeting, the mosque, synagogue or church, the gay bar and on and on."

Justice Sotomayor's concurrence in *Jones*

MORE SUPPORT

"We need not identify with precision the point at which the tracking of this vehicle became a search, for the line was surely crossed before the 4-week mark."

Justice Alito's concurrence in *Jones*,

joined by three other justices

BUT…

"[I]t remains unexplained why a 4-week investigation is 'surely' too long"

Opinion of the Court (by Justice Scalia) in *Jones*

ASSERTION

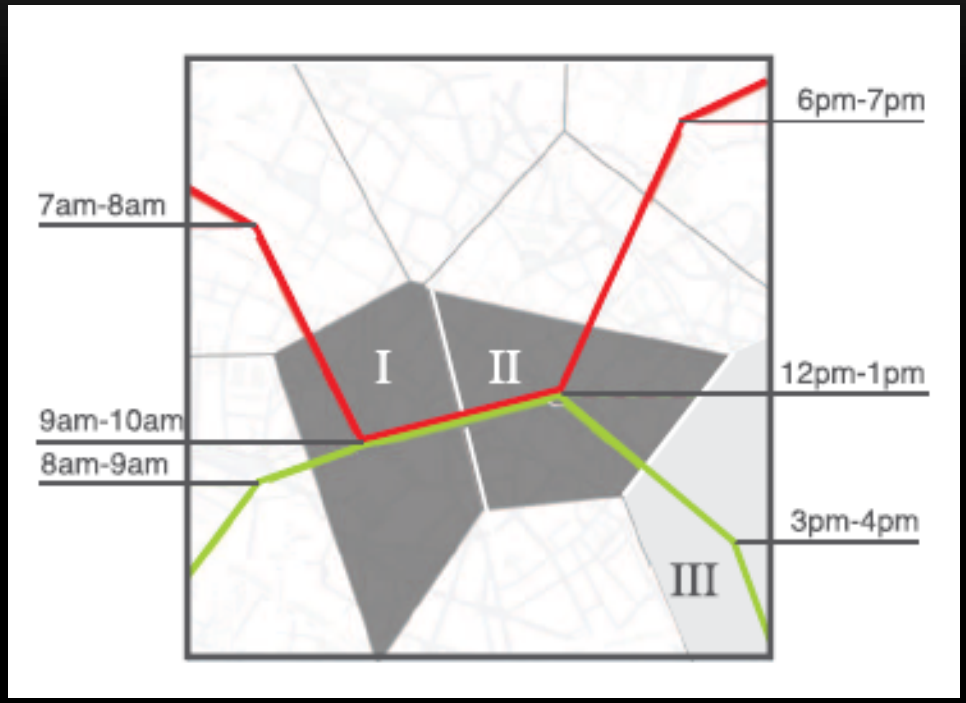
- *We have a mosaic when a suitable algorithm can make accurate enough predictions about a person, based on their location history*

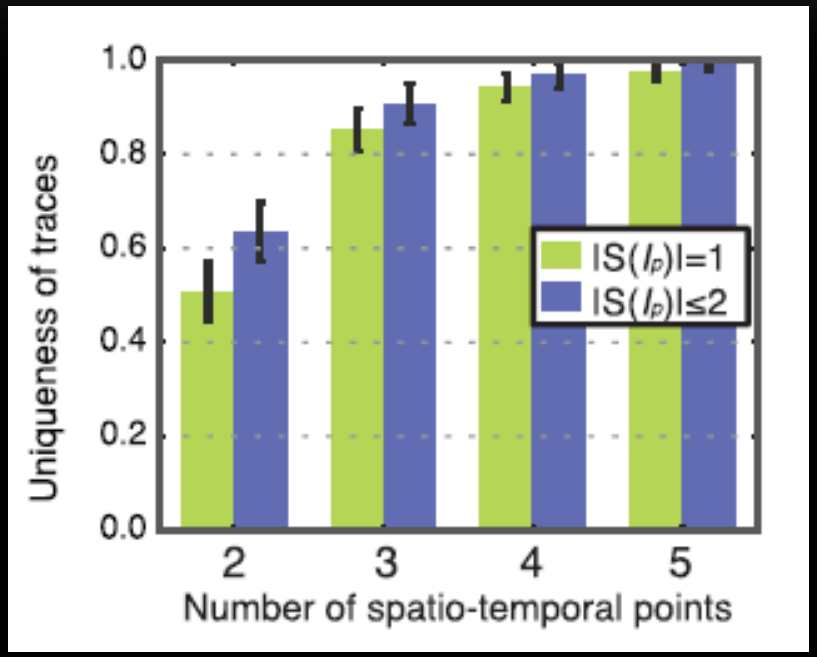
- Computer science questions
  - Do mosaics exist?
  - Can we draw a line?

MOSAIC THEORY AND MACHINE LEARNING: A HYPOTHESIS

- Use machine learning to make predictions based on location data
- When predictions are accurate enough, a mosaic exists
- In other words, use computer science to answer Justice Scalia's objection!

HUMAN MOBILITY PATTERNS





de Montjoye et al, Unique in the Crowd: The privacy bounds of human mobility. Nature srep. 3 (2013)

CREATION OF A MOSAIC

- Graph accuracy against time
- Intuitively, where the slope is increasing we can learn proportionally more from later observations than from earlier ones, that is, our prediction accuracy increases steeply
- *Where the slope has the highest increase, a transformation in the accuracy of factual predictions occurs and a mosaic is created*

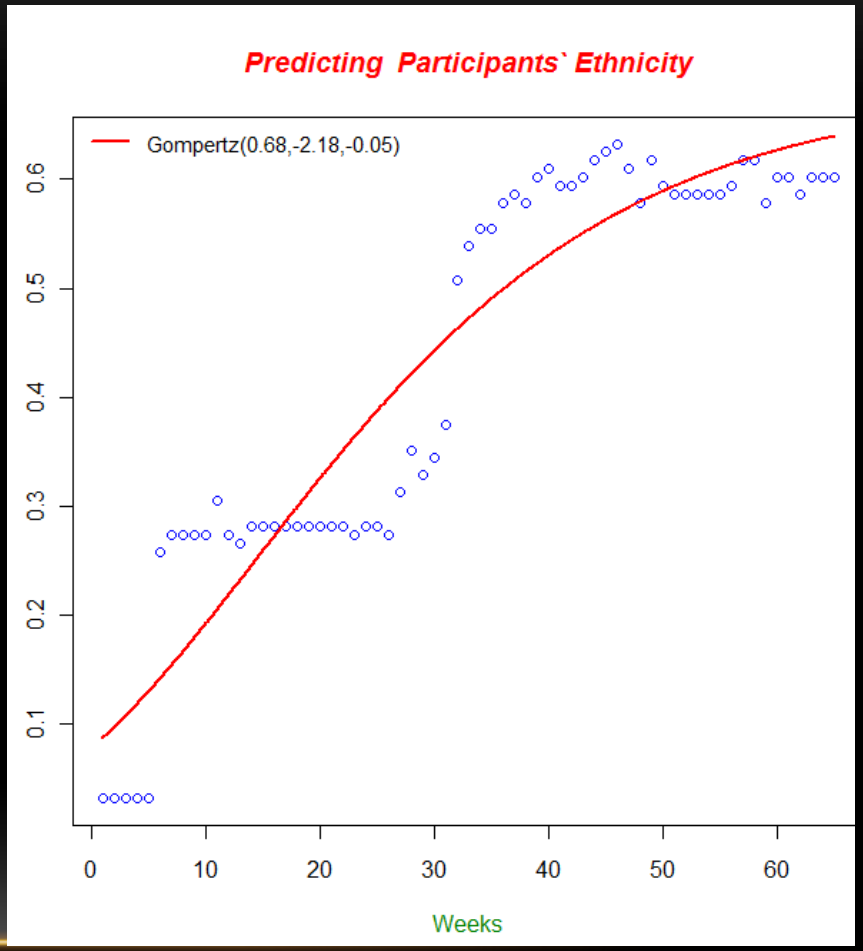
THE SECOND DERIVATIVE

- The **Second Derivative** indicates the Rate of Change in the Slope

- At a certain point, law enforcement can learn disproportionately more relative to the effort they have expended

MACHINE LEARNING AND MOSAIC THEORY

- The technical literature supports the basic premise: with enough points, the whole *is* greater than the sum of its parts
- Note the jump in accuracy at 5 weeks and 28 weeks



(Graph from Altshuler et al.)

ONE WEEK IS THE LIMIT

- Experiments show that week-to-week movements are very predictable (Sadilek & Krumm)
- Weekend movements are *more* predictable, though of course different than weekday movement
- With seven days of observation, you have a very good picture of someone's life

JUSTICE HARLAN IN *KATZ V. US* (1967)

"[T]here is a twofold requirement, first that a person have exhibited an actual (subjective) expectation of privacy and, second, that the expectation be one that society is prepared to recognize as 'reasonable'."

THE FOURTH AMENDMENT

- Does mosaic theory make tracking "unreasonable"?
- Do people have a "reasonable expectation of privacy" in their location *and the inferences that can be made from it*?
- Is it "one that society is prepared to recognize as 'reasonable'"?

WE DON'T KNOW

- Very few court rulings have addressed location privacy head-on
- Most rulings rejecting the claim have relied on other legal principles
- Some day, it will reach the Supreme Court

CURRENT LEGAL STATUS

- The DC Circuit Court has adopted the mosaic theory
- The Massachusetts Supreme Court has, too, and set a limit of two weeks (though without giving a reason for that limit)
- The 11th Circuit originally ruled for it, but that was overturned *en banc (US v. Davis*, 573 Fed. Appx. 925 (2014)); the Supreme Court has declined to hear the case
- Note: the *en banc* ruling in *Davis* was based on historical records, not real-time GPS tracking, and on the "third party doctrine" applied to phone company business records
  - Mosaic theory wasn't rejected by this opinion

THE THIRD PARTY DOCTRINE

- You no longer have a privacy interest in information you voluntarily share with a third party
- Example: the phone number you dial isn't protected because you "gave" it to the phone company (*Smith v. Maryland*, 1979)
- Have you "given" your location to your cell phone company?

WHERE ARE WE?

- From a technical perspective, mosaic theory is correct: you really can build a very full picture of someone with enough data points
- The limit should be about one week
- But—movements are still in public
- But—there are other legal issues that might arise in specific cases, such as the third party doctrine

RESULTS

- The science alone isn't enough
- Fundamentally, this is a legal question, not a technical one. We can supply facts but the courts determine the law. Getting the right answer requires both kinds of input, legal and technical.
- Paper: http://lawandlibertyblog.com/s/Hutchins.pdf

PERSONNEL

- Steven M. Bellovin: computer science, especially security and privacy
- Renée Hutchins: law, especially Fourth Amendment
- Tony Jebara: computer science, especially machine learning
- Sebastian Zimmeck: computer science PhD student (privacy and machine learning)—but he's also a lawyer