Michael Veale
University College London // the Alan Turing Institute

Knowing without Seeing

Informational Power, Cryptosystems and the Law

@mikarv

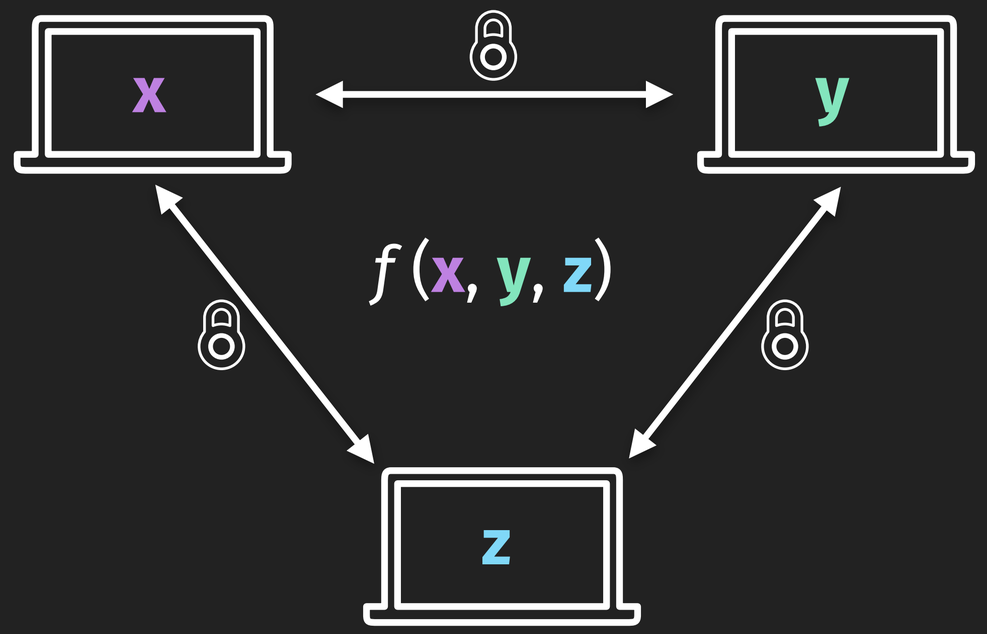
Centralised Data, Broadcast Data... is a third way possible?

References:

Secure multi-party computation (MPC)

A **secure multi-party computation** protocol allows many individuals to collectively compute an aggregate function over data they all hold pieces of, without revealing what they hold to any other player.

For example, they might train a machine learning classifier, or discover which of them has the most money.



See generally P Laud and L Kamm (eds), Applications of Secure Multiparty Computation (Cryptology and information security series volume 13, IOS Press 2015).

Homomorphic Encryption

Homomorphically encrypted data retains its structure when transformed: it can be operated on, such as added or multiplied, and the result later decrypted.

- I give you my encrypted data
- You manipulate it, return the result
- I decrypt the result
- I never learned the algorithm; you never saw my data.

multiply(encrypt(**x**), encrypt(**y**))

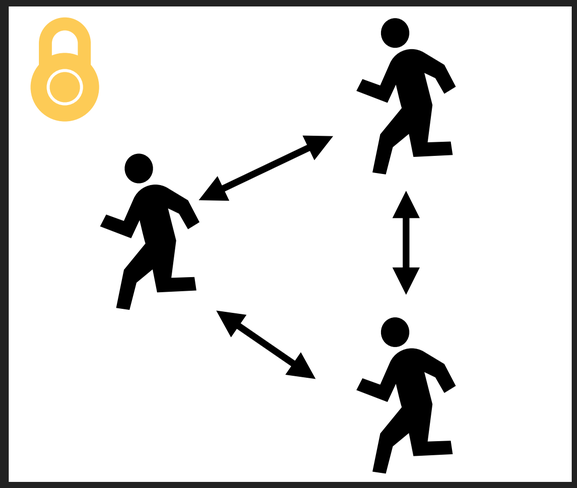
=

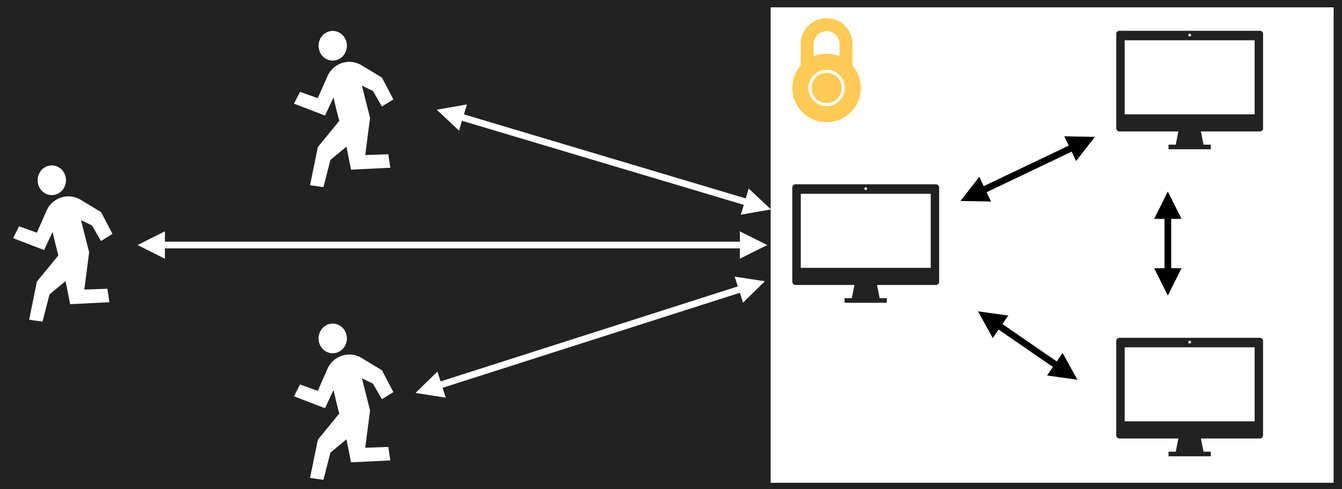
encrypt(multiply(**x**, **y**))

First proposed by Rivest RL, Adleman L, Dertouzos ML, and others, 'On Data Banks and Privacy Homomorphisms' in Foundations of Secure Computation (Academic Press 1978; the first 'fully' homomorphic system realised by Gentry C, 'Fully Homomorphic Encryption Using Ideal Lattices' in Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing (STOC '09, New York, NY, USA, ACM 2009).

Two common structures for using these cryptosystems

- **Edge**
  - Users calculate amongst themselves
- **Distributed, non-colluding**
  - Users trust an arrangement of servers (which could also be run by other users, or themselves) to calculate and not to collude such that privacy guarantees break.
- **Will return to legal distinctions later...**





Agricultural Auctioneering: A Beet Less Public

- WTO challenges to the CAP led the EU to cut sugar beet subsidies. Denmark 🇩🇰, with a monopoly buyer, had to restructure market so only the most efficient survived.
- Nationwide double auction to find the supply and demand curves, and the market clearing price.
  - *the sugar beet farmers submitted how much they would sell for each prices*
  - *the monopoly buyer submitted how much they would buy, and what for.*
- 80% of farmers surveyed were concerned about secrecy of submissions, both with regards to other farmers, and the monopoly buyer, who might use it to extort them.
- Alongside Aarhus University, secure multi-party computation was used to calculate the clearing price.

Bogetoft P, Christensen DL, Damgård I, Geisler M, Jakobsen T, Krøigaard M, Nielsen JD, and others, 'Secure Multiparty Computation Goes Live' in R Dingledine and P Golle (eds), Financial Cryptography and Data Security (Lecture Notes in Computer Science, Springer 2009).

Incentives Matter

- **User incentives ≠ system designer incentives**
  - Hard to know ‘what users want’ — what they do, what they say, or some more paternalistic ‘good life’? [Lyn18]
  - User perspectives from within dysfunctional sociotechnical systems can be stunted: what are the alternatives? How to escape network effects? [Slo18]
- **Societal desires are a bit clearer: we can look at law, politics, as guides**
  - Diverse media consumption as an societal good? [Hel18]
  - Avoiding reinforcing discrimination and prejudice, such as racism or discrimination against disabled individuals, in eg dating apps
  - Price discrimination [Bor17].
  - Policy interventions in areas of vulnerability: eg advertising high-interest loans, or promoting gambling or alcohol to addicted individuals

[Lyn18] Lyngs U and others, ‘So, Tell Me What Users Want, What They Really, Really Want!’, *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (ACM 2018); [Slo18] Slovák P, Frauenberger C and Fitzpatrick G, ‘Reflective Practicum: A Framework of Sensitising Concepts to Design for Transformative Reflection’, Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (ACM 2017); Helberger N, Karppinen K and D’Acunto L, ‘Exposure Diversity as a Design Principle for Recommender Systems’ (2018) 21 Information, Communication & Society 191.; [Bor17] Zuiderveen Borgesius F and Poort J, ‘Online Price Discrimination and EU Data Privacy Law’ (2017) 40 Journal of Consumer Policy 347..

Incentives at Tension

- **System designers routinely ignore environments, ‘low value’ users or non-users [Ove18].**
- **System designers also can try to shape populations to make them more legible or monetisable:**
  - ban jay-walking to make automated cars possible
  - migrate users away from news sites to central platforms
  - lock users into hardware ecosystems
  - change user registration behaviour (eg single sign-ins)
  - A/B test ‘addictive’ or ‘share’-inducing interfaces
- **Are these privacy problems? No — or not always.**
  - Consumer, competition, environment, employment [...]

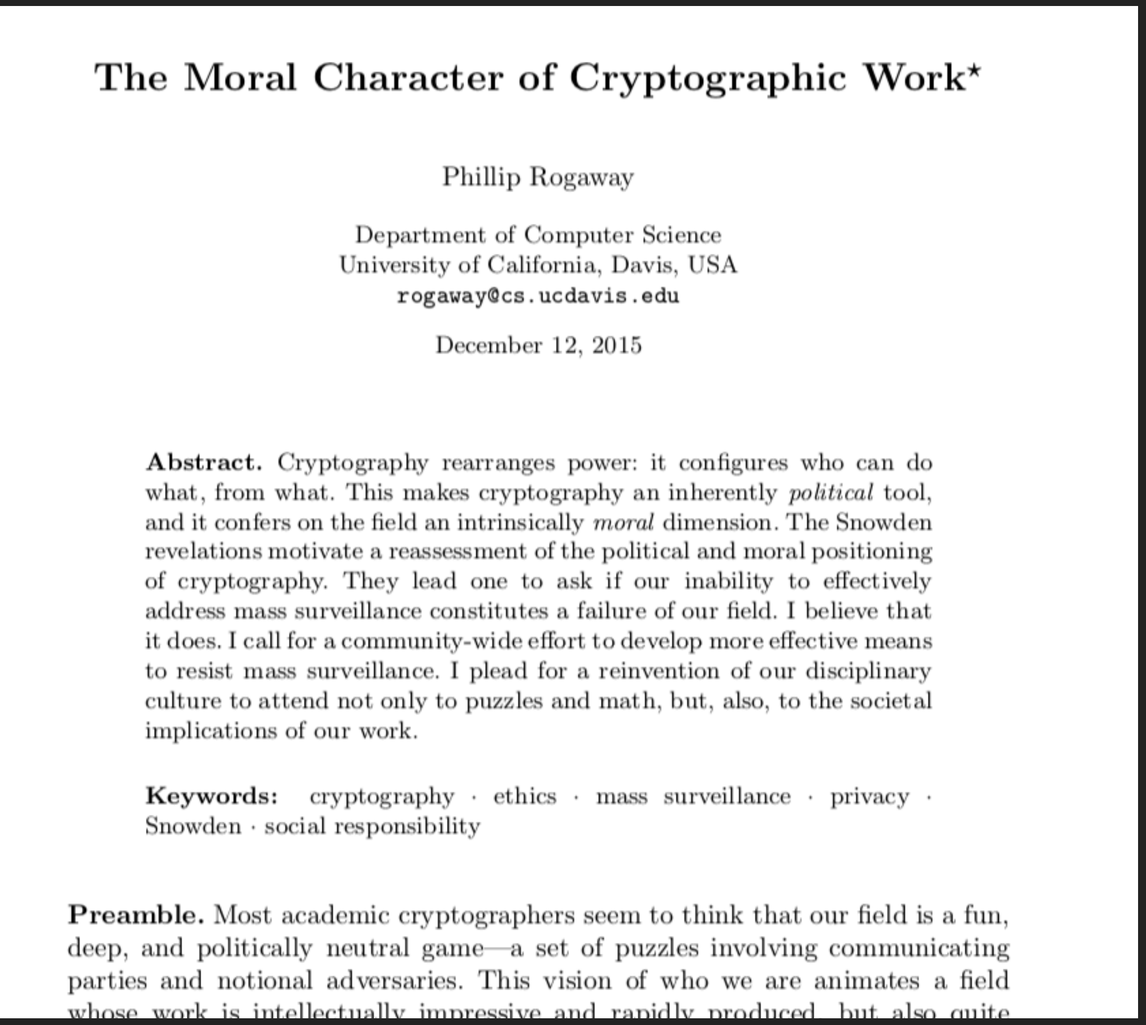
[Ove18] Overdorf R and others, ‘POTs: Protective Optimization Technologies’ [2018] arXiv:1806.02711 [cs] <http://arxiv.org/abs/1806.02711>.

Committing and Binding

- **Two technologies taking prominence in creating binding commitments in private situations**
  - Zero-knowledge proofs
  - Trusted execution environments

References:

Power and Cryptography



The Moral Character of Cryptographic Work*

Phillip Rogaway

Department of Computer Science
University of California, Davis, USA
rogaway@cs.ucdavis.edu

December 12, 2015

Abstract. Cryptography rearranges power: it configures who can do what, from what. This makes cryptography an inherently *political* tool, and it confers on the field an intrinsically *moral* dimension. The Snowden revelations motivate a reassessment of the political and moral positioning of cryptography. They lead one to ask if our inability to effectively address mass surveillance constitutes a failure of our field. I believe that it does. I call for a community-wide effort to develop more effective means to resist mass surveillance. I plead for a reinvention of our disciplinary culture to attend not only to puzzles and math, but, also, to the societal implications of our work.

Keywords: cryptography · ethics · mass surveillance · privacy · Snowden · social responsibility

Preamble. Most academic cryptographers seem to think that our field is a fun, deep, and politically neutral game—a set of puzzles involving communicating parties and notional adversaries. This vision of who we are animates a field whose work is intellectually impressive and rapidly produced, but also quite

Cypherpunk discourse seems sometimes to assume that cryptography will benefit ordinary people. [...] Cryptography can be developed in directions that tend to benefit the weak or the powerful. [...] One reason people might assume cryptography to benefit the weak is that they're thinking of cryptography as conventional encryption. Individuals with minimal resources can encrypt plaintexts in a manner that even a state-level adversary, lacking the key, won't be able to decrypt. But does it necessarily come out that way? To work, cryptographic primitives must be embedded into systems, and **those systems can realize arrangements of power that don't trivially flow from the nature of the tool**.

Rogaway P, 'The Moral Character of Cryptographic Work' [2015] Essay accompanying IACR Distinguished Lecture, AsiaCrypt 2015, Auckland, NZ..

Decentralised Dating: *“Lets do it at my place instead?”*

- **Users can encrypt their profile and secret share it among many servers. They can submit a similarly distributed query, with weights on numeric characteristics and a distance function [in [1], (a-a’)^2], a threshold, and retrieve user profiles that match these thresholds.**
- Who controls eg the distance function?

[1] Yi X and others, ‘Privacy-Preserving User Profile Matching in Social Networks’ [2019] IEEE Transactions on Knowledge and Data Engineering 1

Local Personalisation

- Can users be targeted in the same way as currently, but without data leaving their devices?
- Using MPC and/or homomorphic encryption, train shared models based on tracking data that never leaves an individual's phone.

for an early proposal, see Hamed Haddadi and others, 'MobiAd: Private and Scalable Mobile Advertising' in Proceedings of the Fifth ACM International Workshop on Mobility in the Evolving Internet Architecture (MobiArch '10, New York, NY, USA, ACM 2010).

Escaping from Client Side Profiling?

- Use cryptographic methods, like TEE and ZKPs, to check users are profiling themselves in the way a firm wants them to.
- Network effects as a further limit in some regards.
- Exacerbated by walled gardens, like iOS, and practical inability of users to check what code is running on their systems.

George Danezis and others, 'Private Client-Side Profiling with Random Forests and Hidden Markov Models' in Privacy Enhancing Technologies (Lecture Notes in Computer Science, Simone Fischer-Hübner and Matthew Wright eds, Springer Berlin Heidelberg 2012);

Moral stake in information generation?

- **Is the generation of aggregate information a free-for all?**
- **Do individuals deserve in the way insights that derive from their data are used, *even if*, as with good generalisable analysis, the analysis does not hinge on any single record alone? [Vea18]**
  - Connects to notions of *group privacy*
- **Would it be acceptable for individuals' sensitive data:**
  - medical records
  - phone usage
  - facial or biometric data
  - payment data

  to be mined in an encrypted manner without permission, even if the result was eg differentially private or aggregated?

[Vea18] Michael Veale, Reuben Binns and Lilian Edwards, 'Algorithms That Remember: Model Inversion Attacks and Data Protection Law' (2018) 376 Phil. Trans. R. Soc. A 20180083.

Ad conversion data: Google and Mastercard

- Sometimes hard for advertisers to know ‘what works’, eg when promoting brand awareness.
- Google knows when you saw/clicked. MasterCard knows when you spent. What if you could join the two?
- Google and MasterCard pair up using a cryptosystem (private set intersection) based in part upon homomorphic encryption.
- Input: two parties w/ personal data & shared identifiers. Output, aggregate, non-personal data on spend of those who saw ads.

said there is no revenue sharing agreement with its partners.

A Google spokeswoman declined to comment on the partnership with Mastercard, but addressed the ads tool. "Before we launched this beta product last year, we built a new, double-blind encryption technology that prevents both Google and our partners from viewing our respective users' personally identifiable information," the company said in a statement. "We do not have access to any personal information from our partners' credit and debit cards, nor do we share any personal information with our partners." The company said people can opt out of ad tracking using Google's "Web and App Activity" online console. Inside Google, multiple people raised objections that the service did not have a more obvious way for cardholders to opt out of the tracking, one of the people said.

Ion M, Kreuter B, Nergiz E, Patel S, Saxena S, Seth K, Shanahan D, and Yung M, ‘Private Intersection-Sum Protocol with Applications to Attributing Aggregate Ad Conversions’ [2017] Google Inc.

the law

note: English and Welsh courts do not use gavels

The General Data Protection Regulation [single slide warning ⚠]

- You are a data controller if you determine the means and purposes of processing of personal data — data which relates to an identifiable data subject.
- To process it, you must have a legal ground, which can be consent, but also proportionate legitimate interest, contract, Member State law, emergency, etc.
- The purpose for processing must be limited and well defined. (purpose limitation)
- Data processed should only be necessary for the purpose. (data minimisation)
- Data subjects have (qualified) rights including access, erasure, rectification.
- There are notification requirements if you have a data breach.
- You must carry out a Data Protection Impact Assessment if there are high risks.
- The burden of proof falls primarily on the data controller.
- You must build data protection principles into system design (DP by Design).
- In the very worst cases, fines can reach a max of 2–4% of global turnover.
- *Plus more, which the margins of this slide are too narrow to contain…*

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the pro- tection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1 (GDPR) .

Is personal data being processed?

- **Personal data is**

  any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

- **To determine whether someone is identifiable, must consider what is *reasonably likely* (recital 26).**
- **Court has taken a wide view on this (see *Breyer*):**
  - not necessary that that 'all the information enabling the identification [..] must be in the hands of one person'
  - as soon as there is a third party, irrespective of who it may be, capable of using those dynamic IP addresses to identify network user

References:

Most cryptosystems are likely to be processing personal data

- Where a set of servers assumed not to collude are used, the fact they are in an arrangement — potentially contractual — with a limited number of actors who know of each other, presents a similar case to the issues in *Breyer*.
- Where edge computing is used, situation may arguably be less clear-cut. However, consider the role of a service provider: a mobile operating system, or software provider such as *Signal*. Insofar as they are a coordinating actor (eg though software updates), such users are clearly linked.
- Lastly, there is the possibility that **users are data controller of their own personal data**, which the French regulator, the CNIL, has said is possible. In this case, as the users have access to identifiable data, personal data is being processed in the system.
- **Saying cryptosystems such as these do not legally process personal data will usually be a shaky argument.**

Who is/are the controller(s)?

- **Who determines the means and purposes of processing?**
- **Controllers do not need to see personal data themselves (***Wirtschaftsakademie***)**
  - A Facebook fan page was a joint controller of Facebook data because it benefitted from aggregate insights and had some say on attracting users and setting parameters for audience metrics (*Wirtschaftsakademie*).
  - It is likely that a website owner will be joint controller with Facebook if it installs a Facebook Pixel (*Fashion ID AG Opinion*, *Bobek*).
- **Co-ordinating actors who do not see data can be controllers** (*Jehovan T*)
- **Joint controllership will be common in complex systems, cannot be contracted out of. Yet joint controllers are not equally responsible** (*W'mie*, *Fashion ID AG*).
- **Conclusion: Many organisations may be responsible, but each may not be responsible for everything.**

What of data subject rights: access, objection, erasure, etc.

- Data controllers will often not be in a position to access, object, erase data themselves, where it is held locally or in an obscured manner.
- We see this already with Apple and Siri data, for example, where they refuse to provide access for design reasons [Vea18b].
- Rights users usually trigger against data controllers, they must now trigger against their own devices.
- Importance of *data protection by design*: it's not the same as privacy by design!
- Furthermore, what lawful basis for processing data: either consent or legitimate interest, both of which require withdrawal and/or objection ability to be possible.
- Potentially cutting consequences for walled gardens and designers, but unclear how data protection will intervene in these processes.

[Vea18b] Michael Veale, Reuben Binns and Jef Ausloos, 'When Data Protection by Design and Data Subject Rights Clash' (2018) 8 International Data Privacy Law 19.

What of automated decision-making safeguards?

- Automated, significant ‘decisions’ or measures are clearly possible within cryptosystems.
- The remedy within data protection law for significant, solely automated decision is a human-in-the-loop [Edw18], but this makes little sense here.
- What safeguards for, say, a decentralised financial system cryptographically re-creating the notion of credit scoring? Or decentralised social networks automatically taking down content deemed offensive or illegal?

[Edw18] Lilian Edwards and Michael Veale, ‘Slave to the Algorithm? Why a “Right to an Explanation” Is Probably Not the Remedy You Are Looking For’ (2017) 16 Duke Law & Technology Review 18 <https://osf.io/97upg> accessed 6 August 2018.

Concluding thoughts

- Privacy protective systems can still be manipulative, present users with take-it-or-leave it offers, or propagate unfair outcomes.
- The technologies promoted by PETs researchers to bind the hands of data controllers so that they only carry out intended, privacy-preserving protocols might also be used to bind the hands of users into unfair optimisation systems.
- These systems do not fall outside of data protection law (contrary to much being said in the preamble of papers), but it might not serve as effective protection.
- Important fundamental rights to access data, move it, determine how it is used are not easily reconcilable. In some cases, they might be able to be designed in, but the law is not clear.
- To do: look at other relevant areas of law that would govern entire systems and identify points of intervention and leverage.