Discussion of “Non-rivalry and the Economics of Data” by Christopher Tonetti and Chad Jones

John M. Abowd
Chief Scientist and Associate Director for Research and Methodology
U.S. Census Bureau
Federal Reserve Bank of Atlanta Conference 2019 Financial Markets Conference
Amelia Island, FL, May 20, 2019

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Main Discussion Points

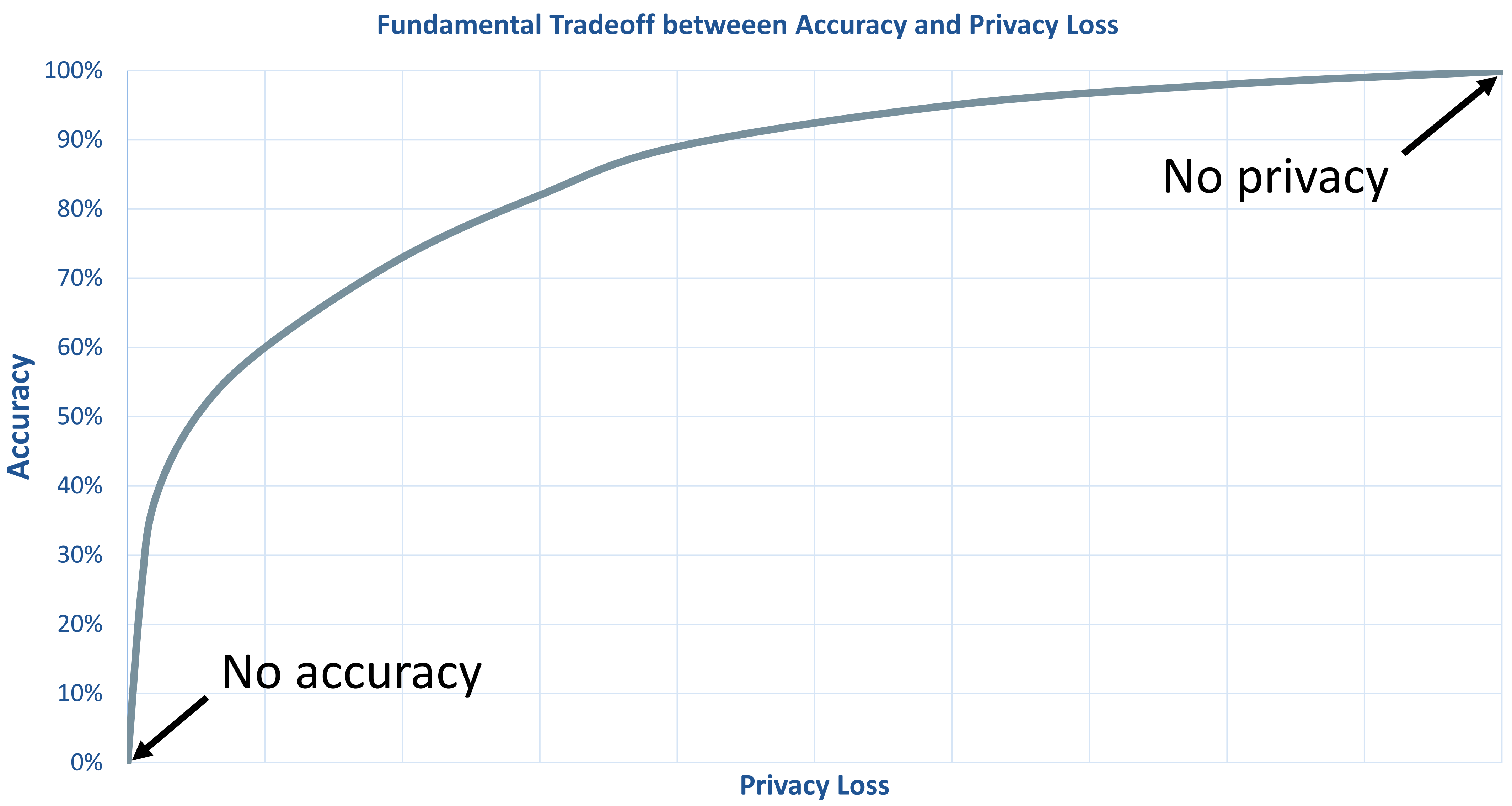
- Key implications
- Privacy loss doesn't have to be binary
- In examining market structure non-binary privacy loss is important
  - Data in the wild (local privacy protection)
  - Trusted curator (central privacy protection)
- Non-binary privacy loss supports non-rivalry on three dimensions
  - Data as an input
  - Output information goods (as distinct from physical consumer goods)
  - Privacy protection itself

Key Implications

- With non-rival data as an input, data should be shared until the marginal social benefit (extra consumer good variety) equals the marginal social cost (extra privacy loss to the consumers)
- Assigning the property rights to consumers comes closer to this outcome because they internalize the privacy loss and allow nearly-optimal data sharing
- Assigning the property rights to firms is more sub-optimal because they share less data out of fear of creative destruction
- Outlawing data sharing is a disaster because it severely limits the gains from non-rivalry

Privacy Loss Need not Be Binary

- In the Jones-Tonetti model, once the consumer surrenders her bits, all privacy over those bits is lost forever
- Privacy-preserving data-use models, based on generalizations of cryptographic semantic security, relax this assumption
  - Full privacy on the input bits = secure storage via encryption
  - Full privacy on the message (output bits) = full encryption = worthless message
  - Relaxation delivers a model where the permitted privacy loss allows the message to be fit for its intended use (here, product development)


Fundamental Tradeoff betweeen Accuracy and Privacy Loss
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Accuracy
Privacy Loss
No privacy
No accuracy

Untrusted Data Recipients (Firms)

- Assumed to receive the data with full precision
- But the internal uses in F(D,L) do not require full precision
- Market could be structured with competition over the precision of harvested data, but might still fail
- Once harvested, the data can be shared just as in the current model without additional privacy loss
- Called “local privacy-enhancing” technology
  - Google RAPPOR
  - Apple iOS 10+
  - Microsoft Windows 10

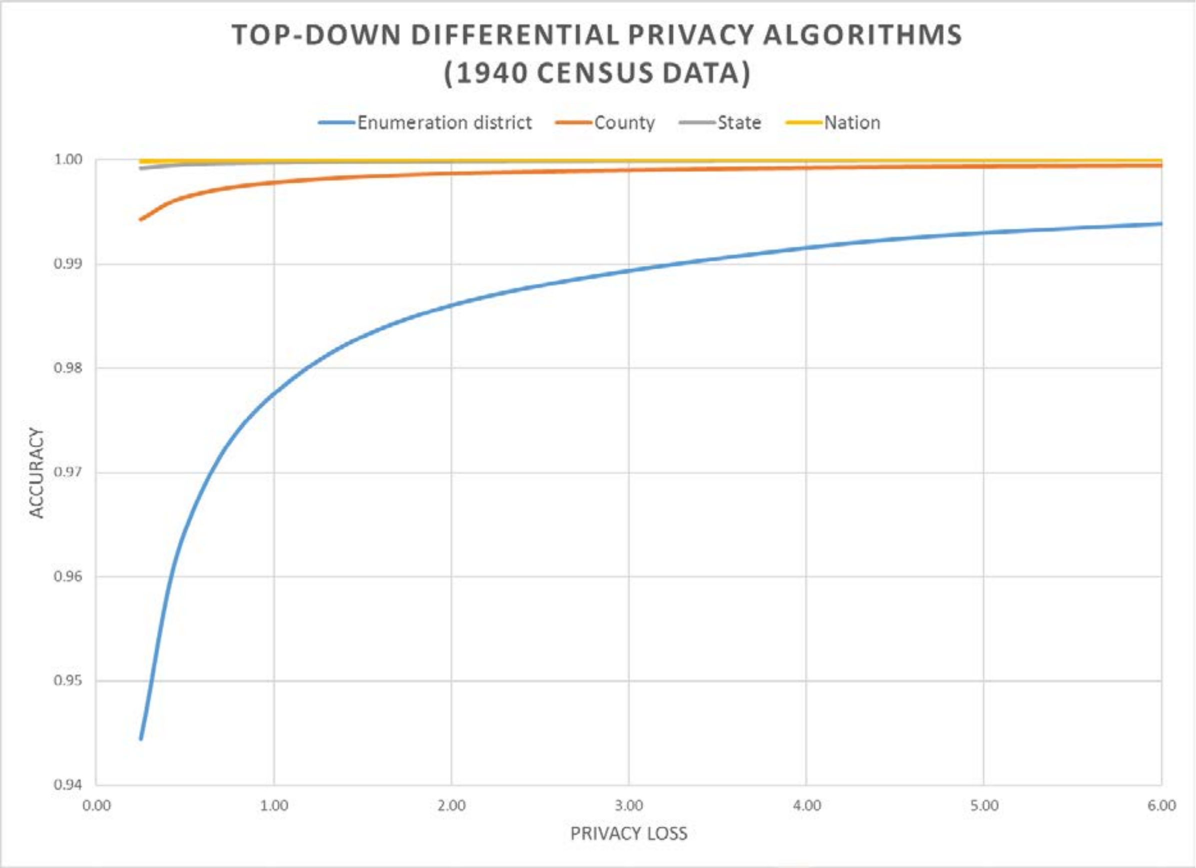
Trusted Data Recipients (Intermediaries)

- Also assumed to receive the data with full precision as custodian
- Data owner holds the private encryption key
- Information products are released to customers with required precision
- Market can handle the supplier (consumer) side
- Market will fail on the product (firm) side because the information product is also non-rival and more like an idea than an input

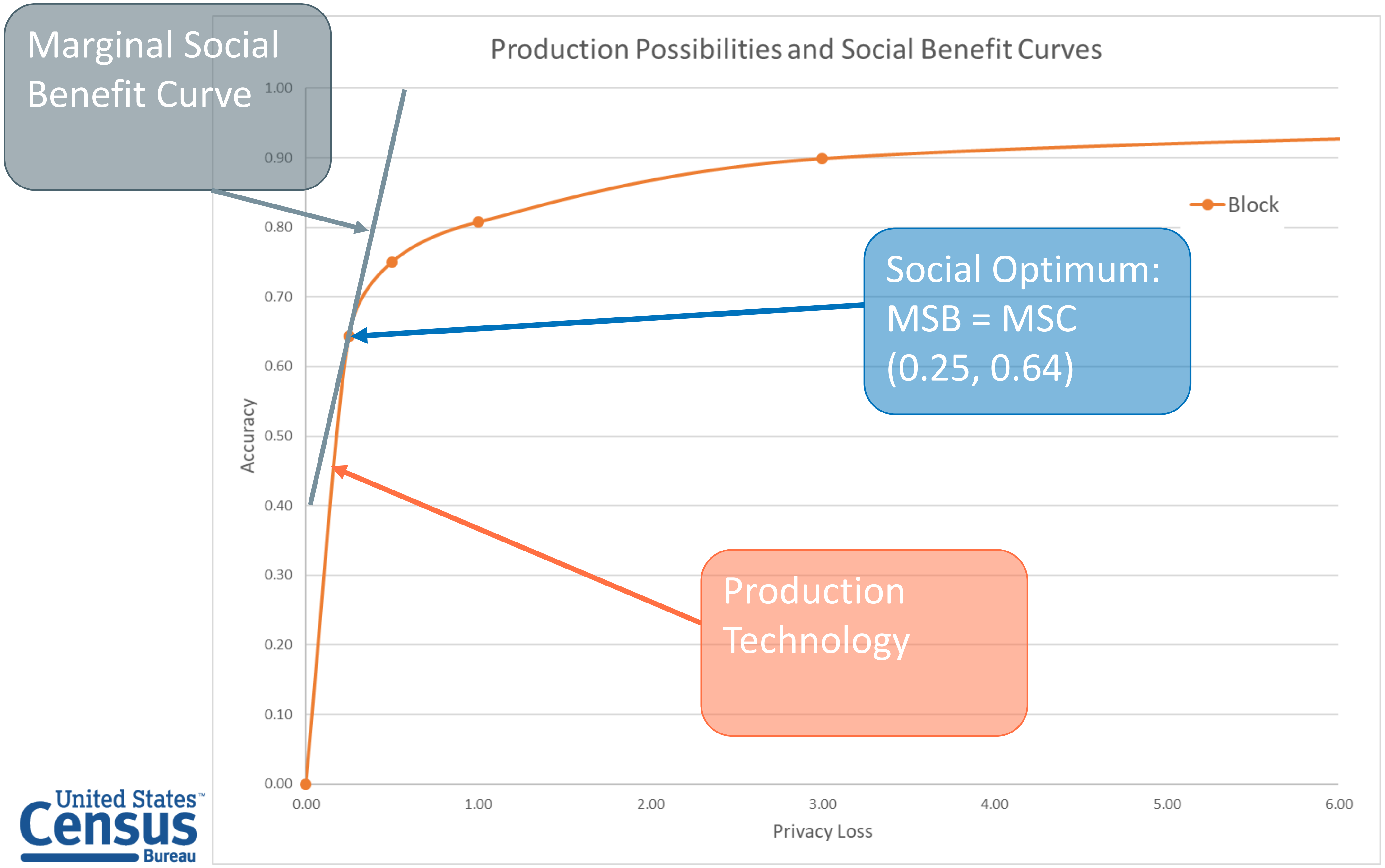
Examples from the 1940 and 2020 Censuses

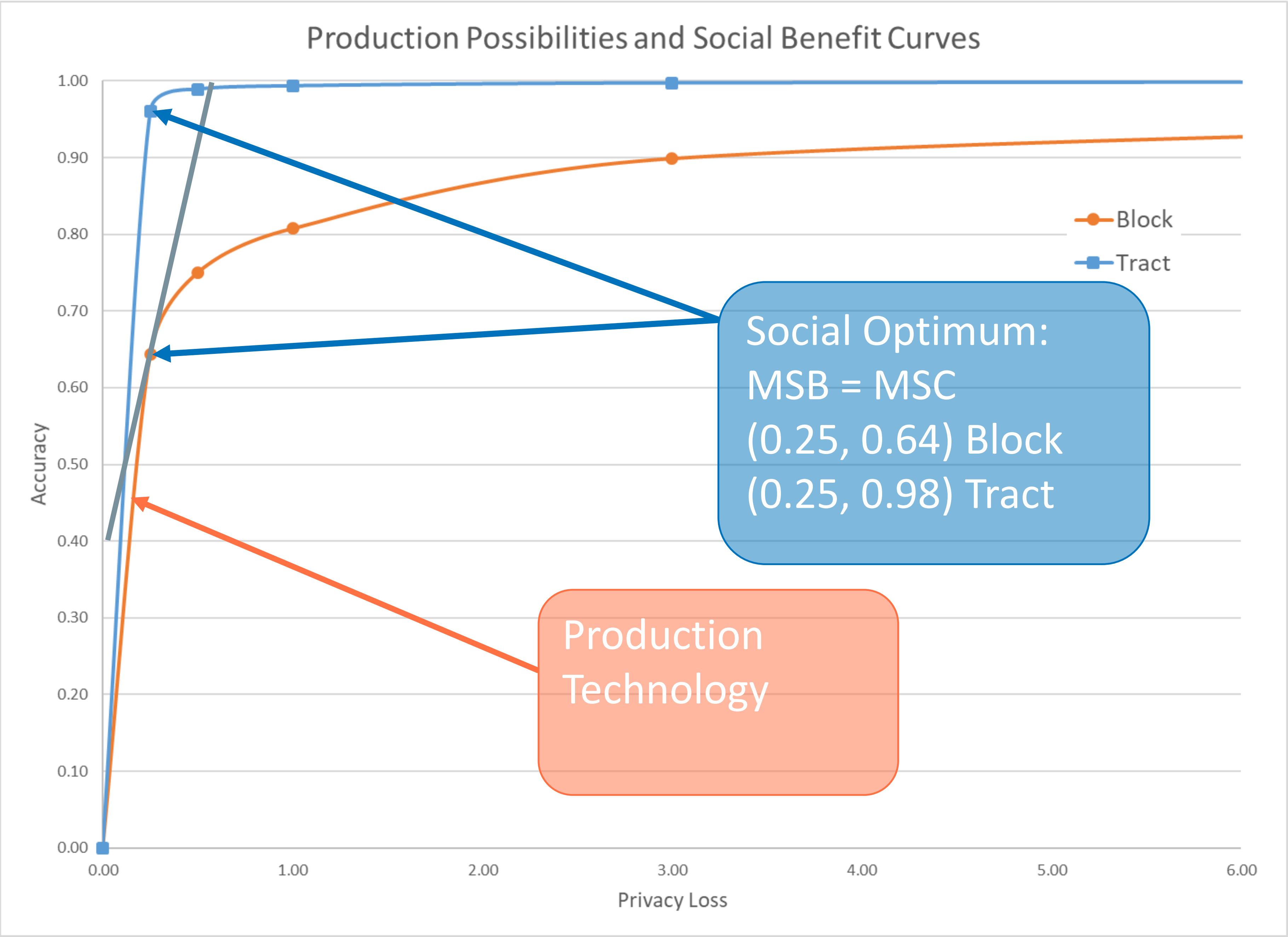
Two Candidate Algorithms

- Local model
  - Privacy protection applied to tables at the most detailed geographic level
  - All aggregations built from those tables
- Central model
  - Privacy preserving measurements at all levels of the geographic hierarchy
  - All aggregations get tuned accuracy


TOP-DOWN DIFFERENTIAL PRIVACY ALGORITHMS
(1940 CENSUS DATA)
Enumeration district
County
State
Nation
ACCURACY
1.00
0.99
0.98
0.97
0.96
0.95
0.94
0.00
1.00
2.00
3.00
4.00
5.00
6.00
PRIVACY LOSS


DISTRICT-BY-DISTRICT DIFFERENTIAL PRIVACY ALGORITHMS
(1940 CENSUS DATA)
Enumeration district
County
State
Nation
ACCURACY
1.00
0.99
0.98
0.97
0.96
0.95
0.94
0.00
1.00
2.00
3.00
4.00
5.00
6.00
PRIVACY LOSS


Production Possibilities and Social Benefit Curves
Marginal Social Benefit Curve
Social Optimum: MSB = MSC (0.25, 0.64)
Production Technology
Block
Accuracy
Privacy Loss
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.00
1.00
2.00
3.00
4.00
5.00
6.00



United States Census Bureau

Non-rivalry with Non-binary Privacy Loss

- Data as an input
  - Supported by current local implementations (RAPPOR, iOS, Windows 10)
- Output information goods (as distinct from physical consumer goods)
  - Supported by statistical agency implementations
  - Supported by newer open source PROCLOH, Google Privacy Amplification ML
- Privacy protection itself
  - VCG auctions (Ghosh and Roth, 2015)
  - Other mechanisms (Arrieta-Ibarra et al. 2018)

Thank you

John.Maron.Abowd@census.gov