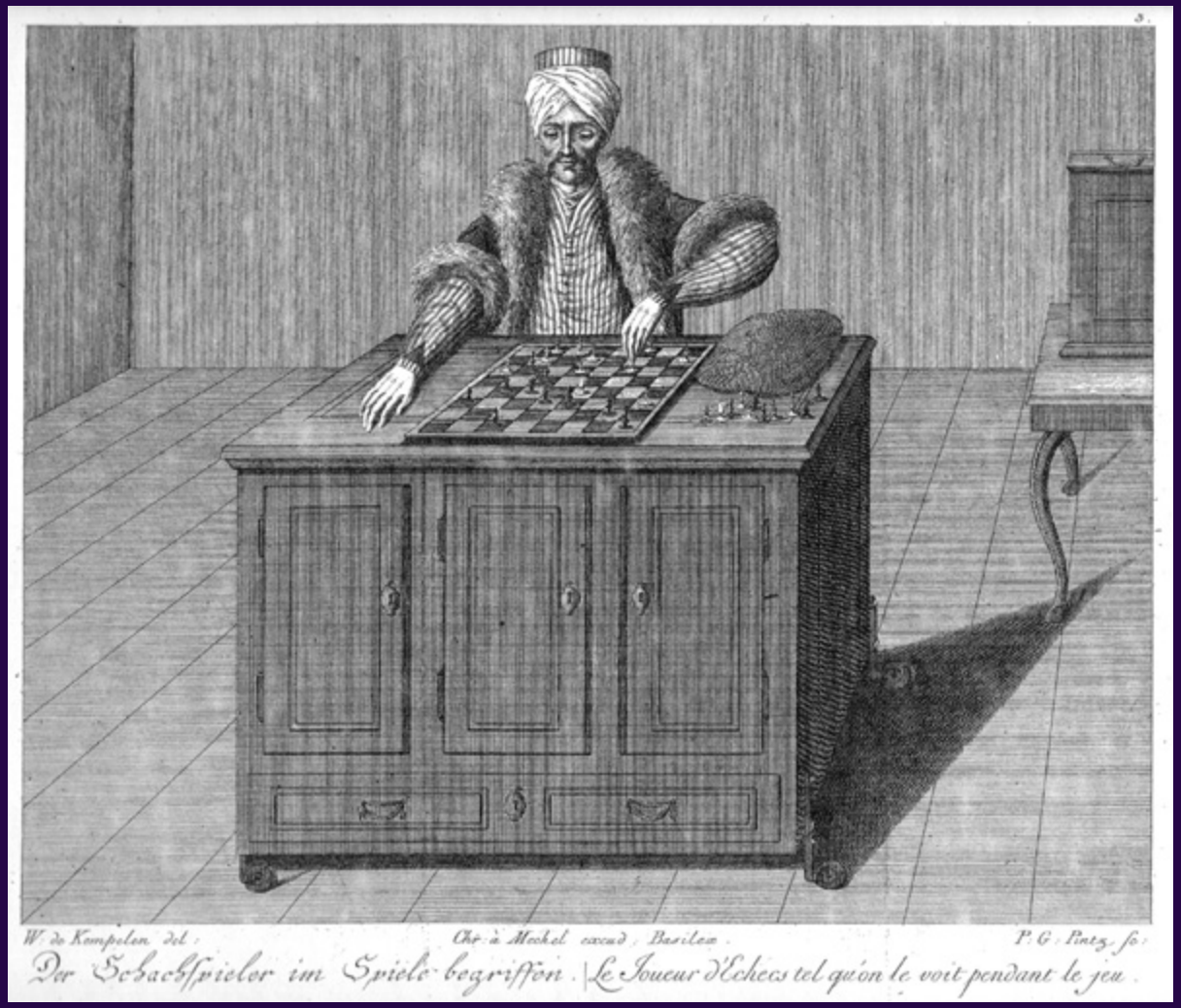
When is Algorithmic Secrecy Justified?

Ignacio N. Cofone (McGill University)
Katherine J. Strandburg (New York University)

Montreal-Seoul, 29 September 2020

1. Context



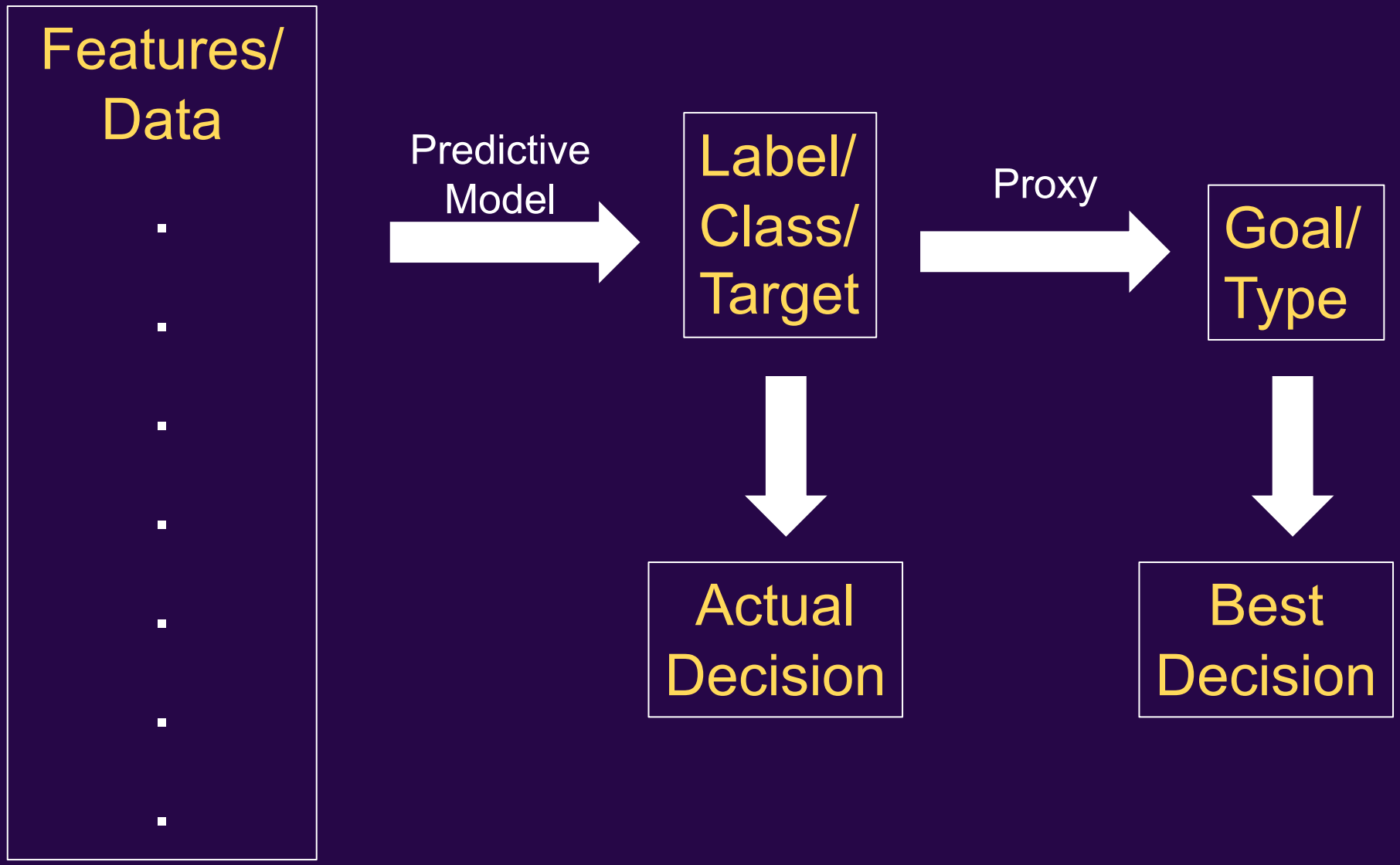
W. de Kempelen del. Chr. à Mechel excud. Basileæ. P. G. Pintz. sc.

Der Schachspieler im Spiele begriffen. | Le Joueur d'Echecs tel qu'on le voit pendant le jeu.

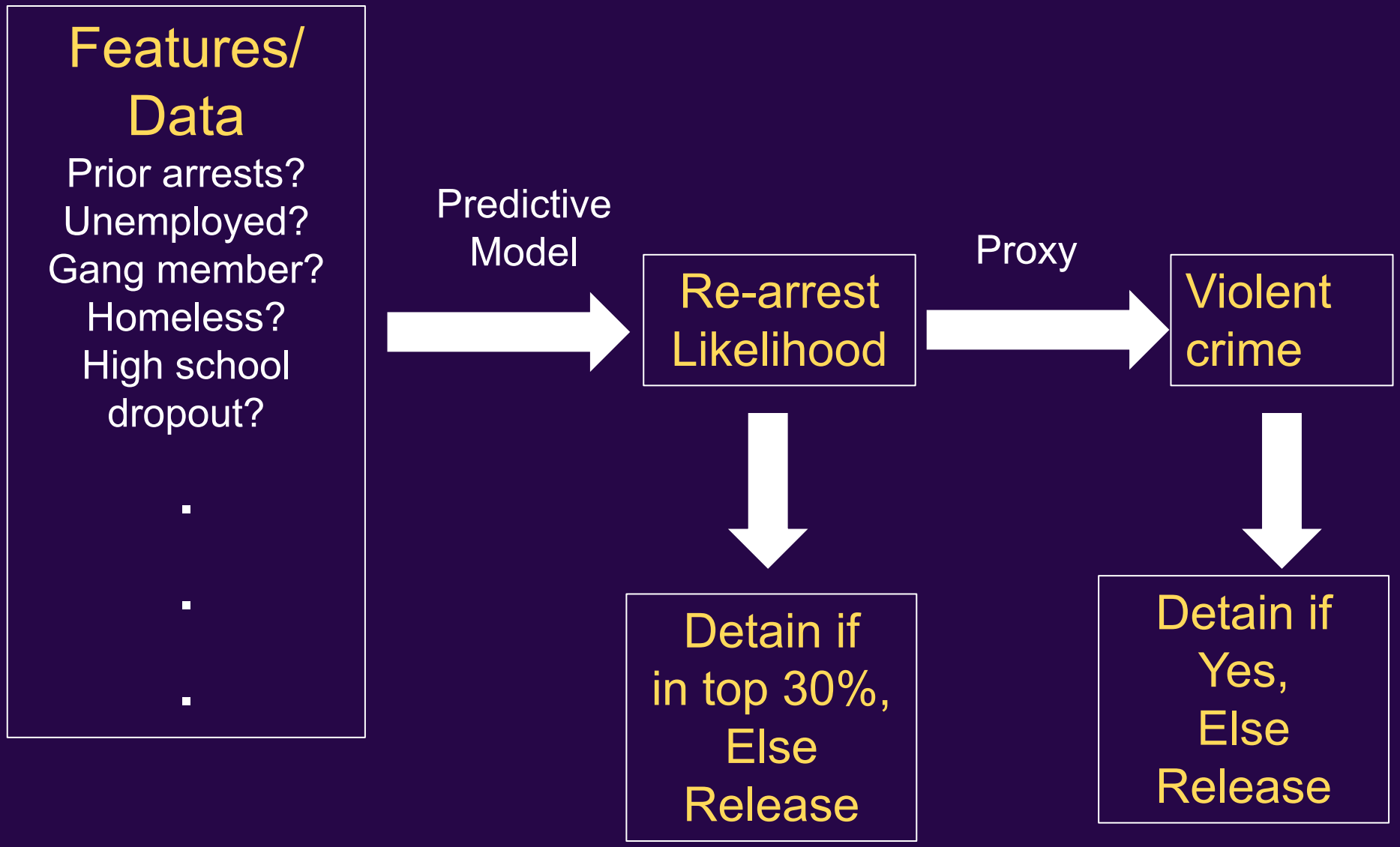
What are algorithmic decisions

- Hiring, finance, criminal procedure, health, online speech.

Anatomy of a predictive algorithm



Anatomy of a predictive algorithm



2. The tradeoff

The importance of transparency

- Compliance
- Error and bias correction
- Procedural rights / accountability

The secrecy arguments

- “If we don’t keep this algorithm secret, people will game the system”
- “If we don’t keep this algorithm secret, our competitors will gain advantage”
- A wide range of decision-making arenas
  - Industry
  - Government
  - Academia

Disclosure v. Nondisclosure Trade-off

- Social costs of disclosure
  - Can/will decision-subjects game the system if disclosed?
  - Can/will competitors gain advantage?

- Social benefits of disclosure
  - Accountability
    - Shirking, Bias, Private interests
  - Error correction
    - "Gaming" a noisy/biased proxy
  - Compliance
    - "Gaming" that improves eligibility for beneficial decisions

Types of disclosure

1. Training data
2. Sources of training data
3. Code
4. Model
5. Features/labels
6. Feature/label weights
7. Output variable
8. Ultimate goal

Should Disclosure Be Mandated?

- Why not leave it to the market?
  - Should we trust private entities to choose what to disclose?
- Do they have the right incentives to make the trade-off?
  - Designers don't fully account for the social value of disclosure-triggered compliance
  - They may not account for the social costs of inaccuracy/bias
  - They may gave self-serving incentives to hide details of the process
    - Regulated aspects, e.g. discrimination in hiring
- How should judges and policymakers decide?

3. What trade-offs exist when secrecy concerns are real?

A. Costs: whether the concern is warranted

- Examine whether disclosure would produce
  - Socially undesirable gaming
  - Risky disclosures in terms of trade secrecy
- When does disclosure lead to competitors free-riding?
  - Always?
  - For certain types of disclosure?
    - Features vs code
  - For certain modes of disclosure?
    - Under seal

When can people game the system?

1. Proxies are not tightly tied to decision-making criteria
2. Disclosure pertains to features that are modifiable by decision-subjects at an appropriate time
3. Modifying those features is cost-effective
4. Modifying those features improves the proxy without improving decision-subject's eligibility
   - Compliance is not "gaming
5. Proxy correctly labels decision subject as "bad"

- Conditions are cumulative

B. Costs: loss in system accuracy

- Examine the accuracy of the proxy being discussed
  - If accurate, good reason to keep secret
    - Algorithmic equivalent of 4A particularity
  - If inaccurate, proxy is of low social value
    - Losing it may be a small social loss
    - Disclosure may lead to error correction
    - Distributional concerns

- The costs are different for gaming and TS, but the benefits are the same

C. Incentive alignment

- False positives and false negatives as error categories

False positives and negatives

Proxy\Type	"Bad"	"Good"
Detrimental	TP = detain violent recidivists TN = do not hire bad employees	FP = detain non-recidivists FN = do not hire good employees
Beneficial	FN = release recidivists FP = hire bad employees	TN = release non-recidivists TP = hire good employees

C. Incentive alignment

- Sometimes, utility aligns
  - But e.g. child services
  - Designer: minimize FN
  - Society: minimize FP

- Incentive alignment matters
  - If aligned, good reason to keep secret
  - If misaligned, good reason to disclose

Principal-Agent Problem: Recidivism

Social perspective:

- Presumption of innocence, Racially biased proxy

FP \ FN	High	Low
High	1 (Useless)	2 (Get a better proxy)
Low	3 (Err on the side of justice)	4 (Great)

Decisionmaker perspective:

- Reputation, Racially biased proxy

FP \ FN	High	Low
High	1 (Useless)	3 (Ruined a couple of lives but I can keep problems to myself)
Low	2 (Yikes, I let some recidivists out)	4 (Great)

Principal-Agent Problem: Recidivism

FP \ FN	High	Low
High	Society = NO Decisionmaker = NO If Errors are observable	Society = NO Decisionmaker = Yes
Low	Society = Yes Decisionmaker = No	Society = Yes Decisionmaker = Yes

- Externalizing error costs, DM preferences depend on social observability of error rate

Principal-Agent Problem: Employment

Social perspective:

- Concern about biased proxy

FP \ FN	High	Low
High	1 (Useless)	3 (Need to do more screening, but not too bad)
Low	2 (Only accurate for white guys?)	4 (Great)

Decisionmaker perspective:

- Only concerned about hiring good enough employees

FP \ FN	High	Low
High	1 (Useless)	2 (More work for me)
Low	3 (Everybody I hired is good)	4 (Great)

Principal-Agent Problem: Employment

FP \ FN	High	Low
High	Society = No Decisionmaker = No	Society = Yes Decisionmaker = NO
Low	Society = No Decisionmaker = Yes	Society = Yes Decisionmaker = Yes

- Externalizing error costs partially, decisionmaker preferences depend on false positives only

Conclusions

1. Disclosure is often of high social value
2. Gaming is harder than the rhetoric suggests
3. Principal-agent problems are common
4. Algorithm performance is determined by accuracy (noisiness of proxies), FP/FN trade-offs *and* gaming/TS.
5. Even when gaming is possible, it's sometimes less socially costly than algorithmic secrecy
6. Secrecy should not be the default policy choice

Conclusions

- Taxonomy

- Allow secrecy of any aspect of an algorithm if disclosure
1. leads to gaming or free-riding
2. of a valuable proxy,
3. of an algorithmic designer with aligned social incentives

- Mandate disclosure if not