Why formalizing fairness won't fix (algorithmic) discrimination *(reloaded)*

Bettina Berendt
KU Leuven

Reloaded version presented at the Dagstuhl Seminar "10 Years of Web Science – Closing the Loop", 28 June 2018
An earlier version was presented at Hannover's L3S, 24 November 2017

Publications and materials on privacy, privacy education, discrimination, and ethics at https://people.cs.kuleuven.be/~bettina.berendt/
See end of this slide set for references and URLs

Who's being discriminated against here?



Ad related to **latanya sweeney** ⓘ

Latanya Sweeney Truth
www.instantcheckmate.com/
Looking for **Latanya Sweeney**? Check **Latanya Sweeney's** Arrests.

Ads by Google

Latanya Sweeney, Arrested?
1) Enter Name and State. 2) Access Full Background Checks Instantly.
www.instantcheckmate.com/

Latanya Sweeney
Public Records Found For: **Latanya Sweeney**. View Now.
www.publicrecords.com/

La Tanya
Search for La Tanya Look Up Fast Results now!
www.ask.com/La+Tanya

(Sweeney, 2013)

Algorithms don't discriminate – people do (well, and algorithms help)

Misunderstandings about the roles of algorithms and people at the …

- Conceptual level
- Level of causing discrimination
- Level of addressing discrimination
  (with a bridge to accountability)

Algorithms don't discriminate – people do (well, and algorithms help)

Misunderstandings about the roles of algorithms and people at the …

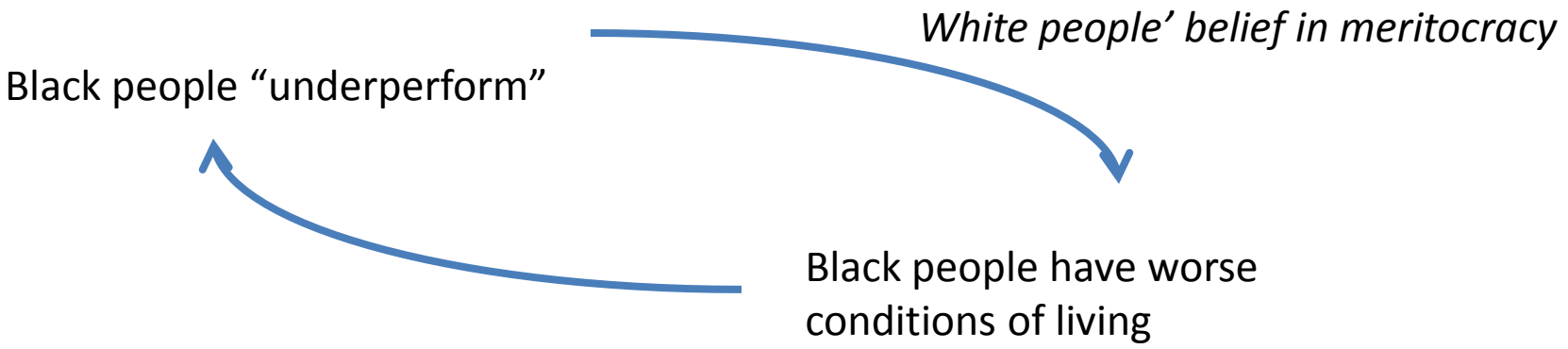
- Conceptual level
- Level of causing discrimination
- Level of addressing discrimination
  (with a bridge to accountability)
- My **key claims are formatted in bold purple**.
- Disclaimer: 30 minutes restriction requires *stark* simplifications.

Inspiration: “Feedback loops” are a well-known and well-studied problem

- Myrdal, G. (1944). An American Dilemma:
- Cumulative causation / vicious+virtuous cycles

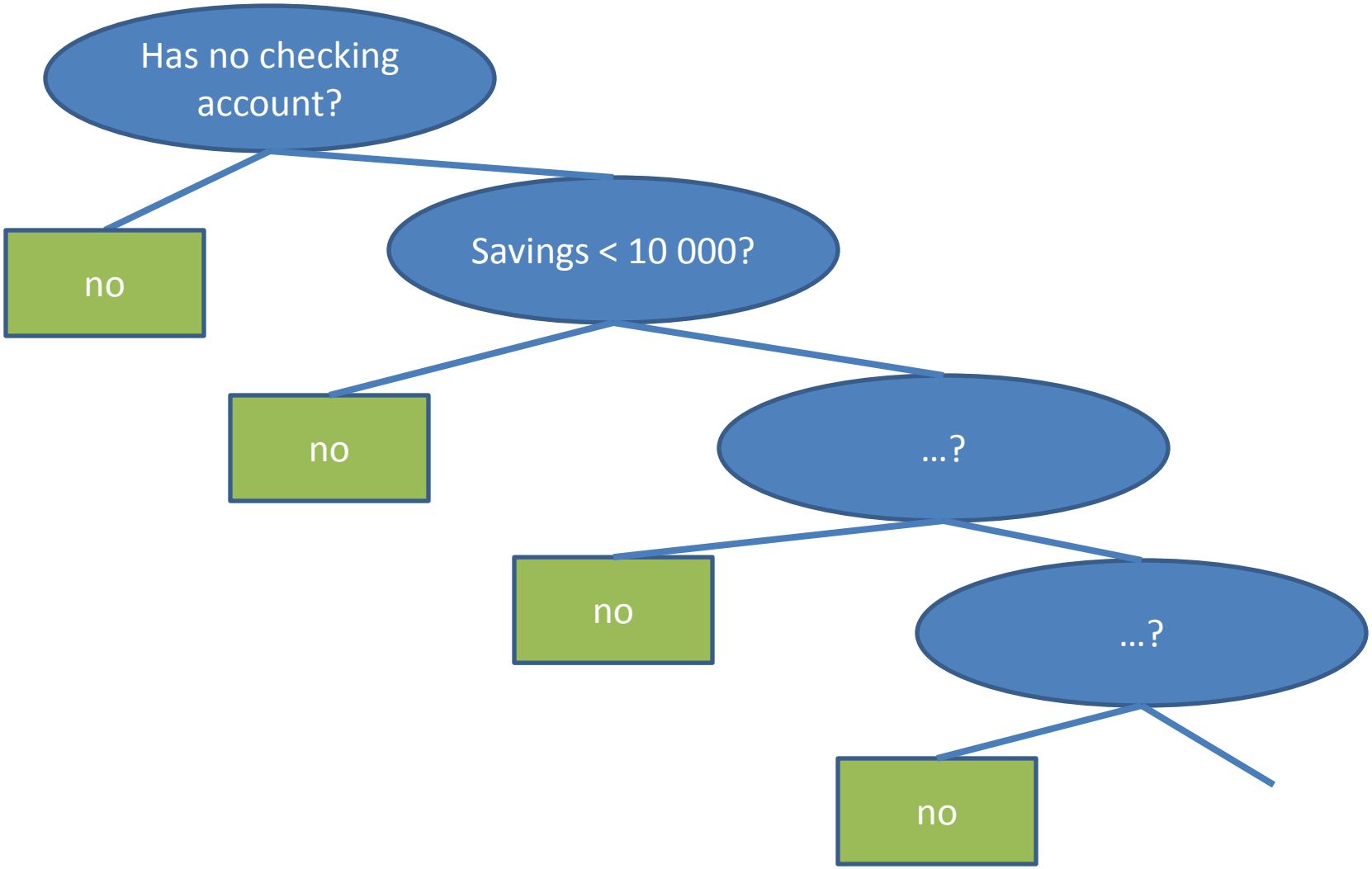
Inspiration: “Feedback loops” are a well-known and well-studied problem

- Myrdal, G. (1944). An American Dilemma:
- Cumulative causation / vicious+virtuous cycles

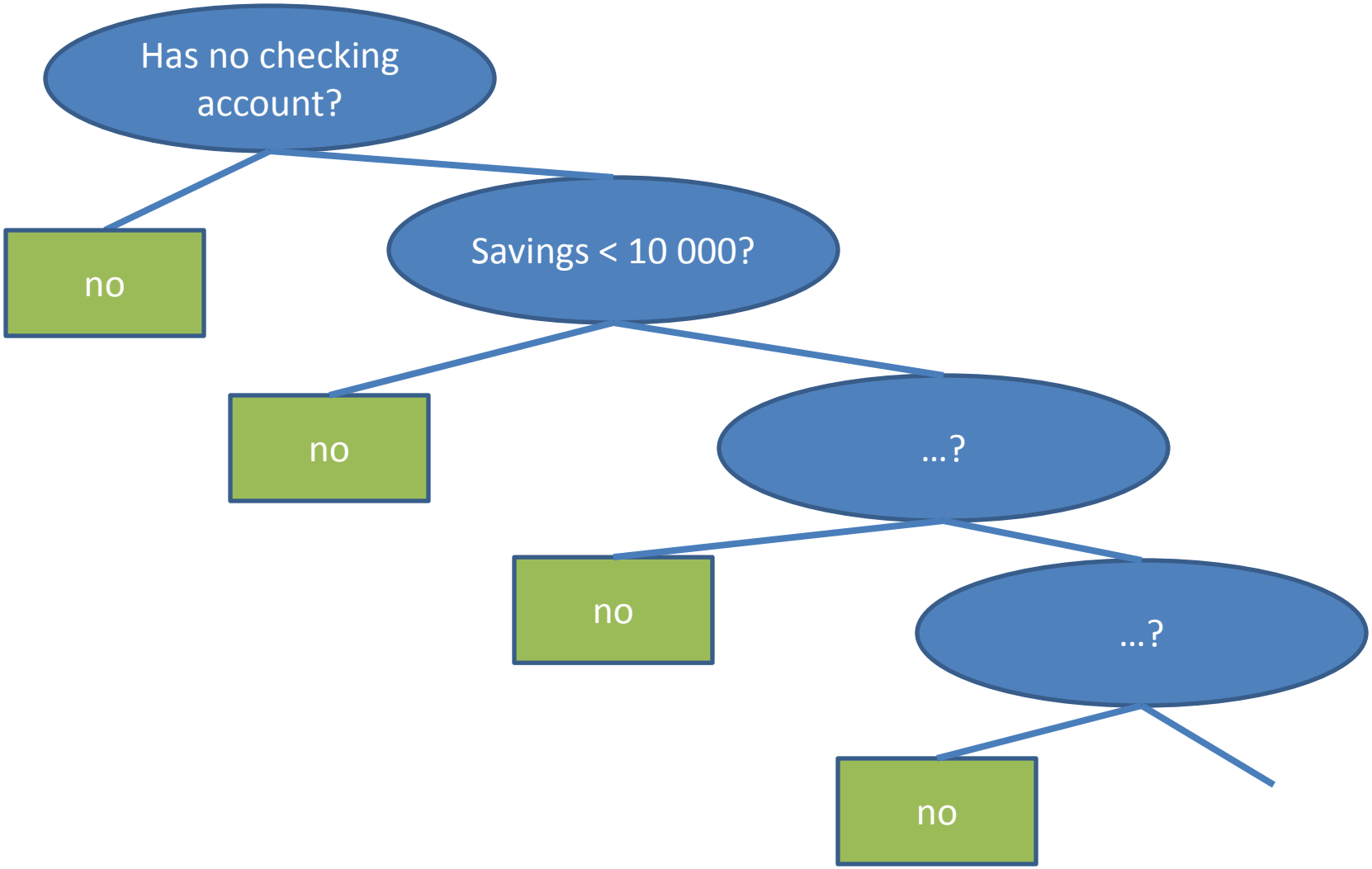


Background:
Data mining
Discrimination

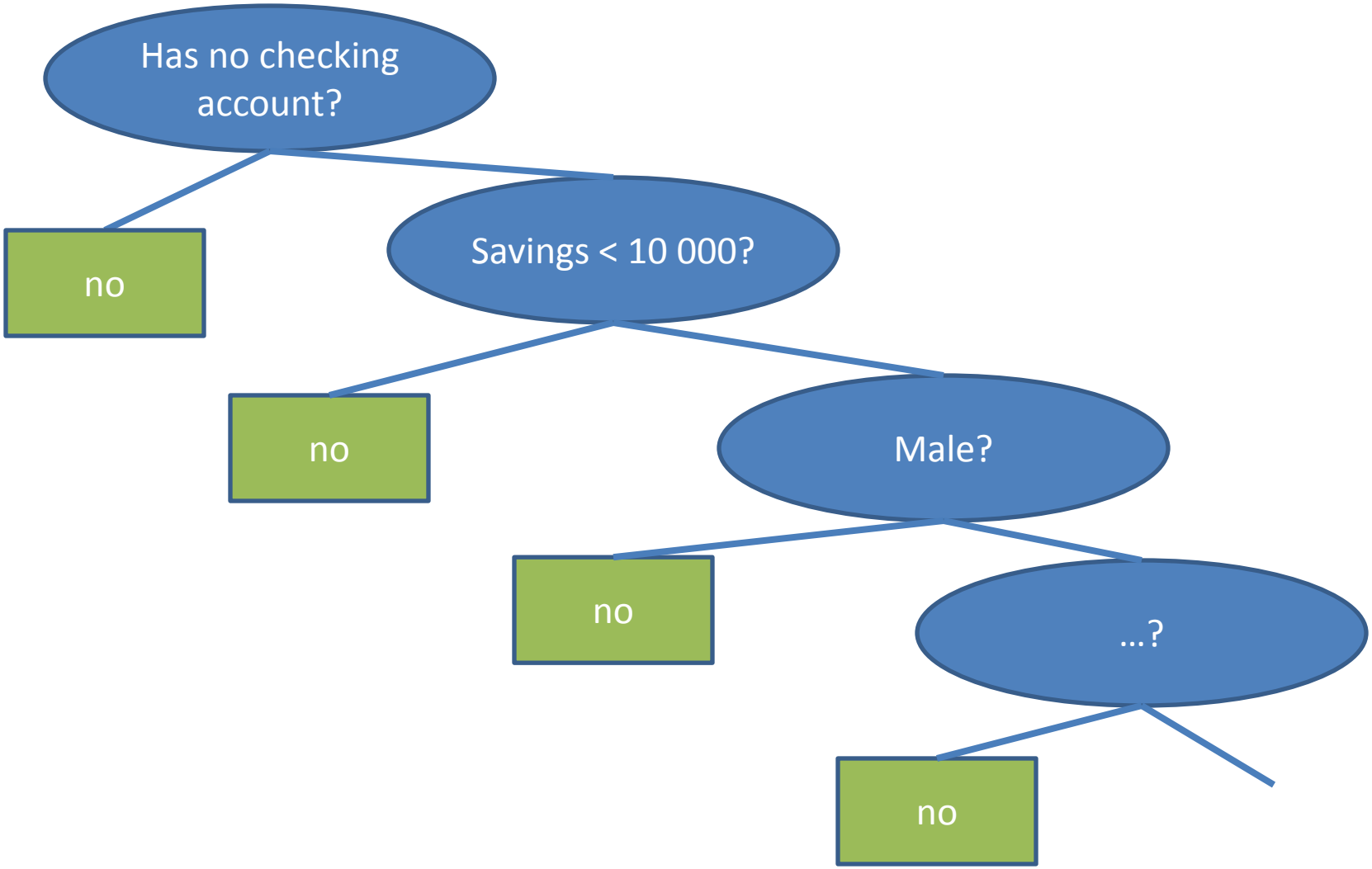
Data mining *is* discrimination!



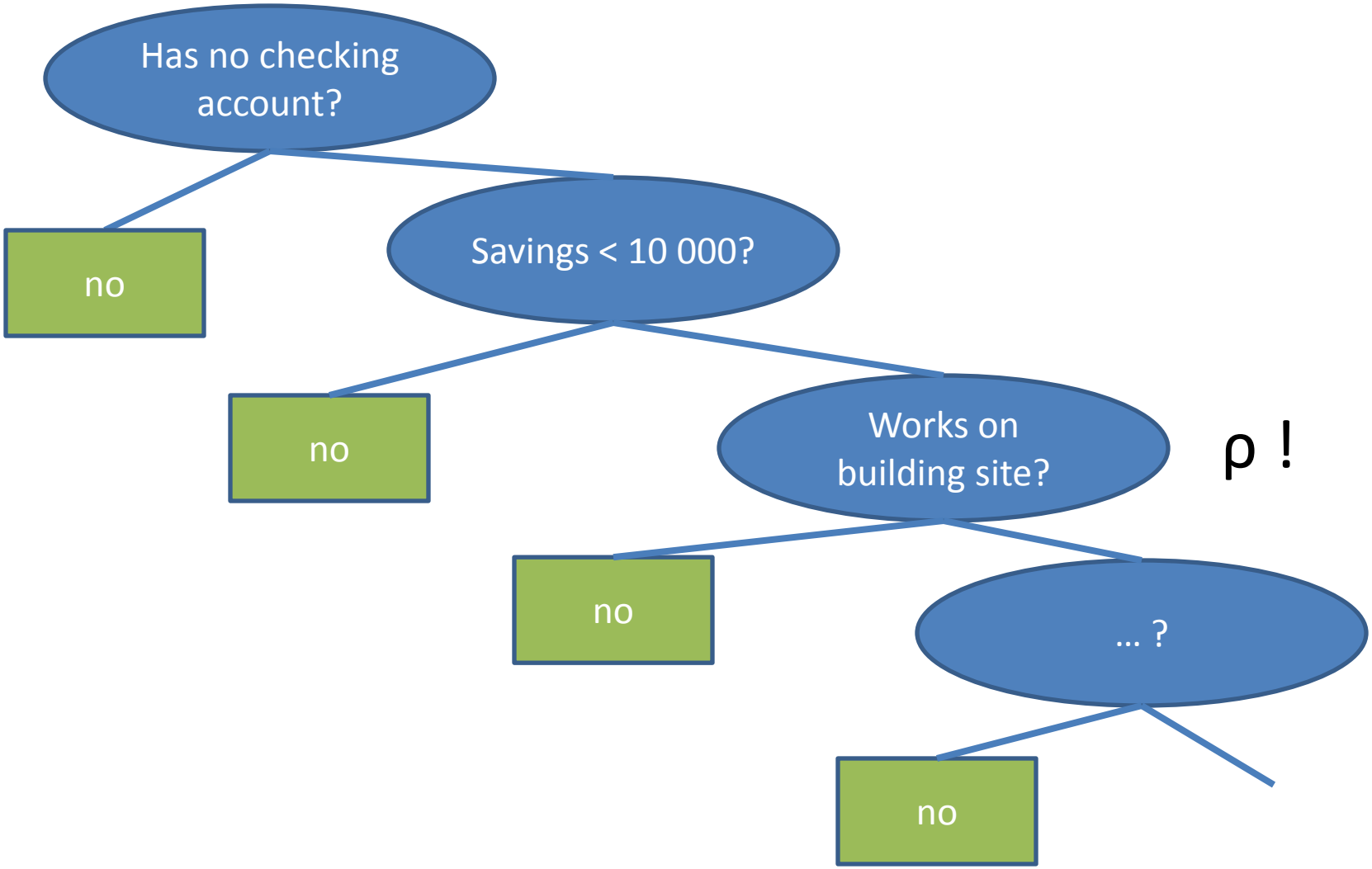
Legal view: Differentiation



Unlawful direct discrimination



Unlawful indirect discrimination



„You may not ...“ vs. “you may ...“

- In many areas, including
  - Labour
  - Loans
  - Insurance
- The protected-by-law grounds differ by area, but usually include gender, disability, age and sexual orientation, cultural, religious and linguistic beliefs/affiliation

“You may no longer ...“

- European Court of Justice (2011) Case C-236/09, Association Belge des Consommateurs Test-Achats ASBL and Others v Conseil des ministres:
  - … the use of sex as an actuarial factor should not result in differences in individuals’ premiums and benefits. …
- General Data Protection Regulation (EU law as of 2018)
  - Various recitals, among others:
    - 71: the data controller should take measures to “[prevent] discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation …”

- Historical examples: only { rich | white | male } people get to vote

Society *is* discrimination!

Society *is* discrimination

- This implies, among other things:
- **Mitigating discrimination is a moving target.**
- **We cannot “solve” “the discrimination problem” in AI / Web Science / …**

Approach:
Discrimination-aware data mining
(Pedreschi, Ruggieri, & Turini, 2008; & many others since)

for discrimination prevention and detection

Pedreschi, Ruggieri, & Turini (2008)

- PD and PND items: potentially (not) discriminatory

  goal: want to detect & block mined rules such as

  purpose=new_car & gender = female → credit=no

  measures of discriminatory power of a rule include

  elift (B&A → C) = conf (B&A → C) / conf (B → C) ,

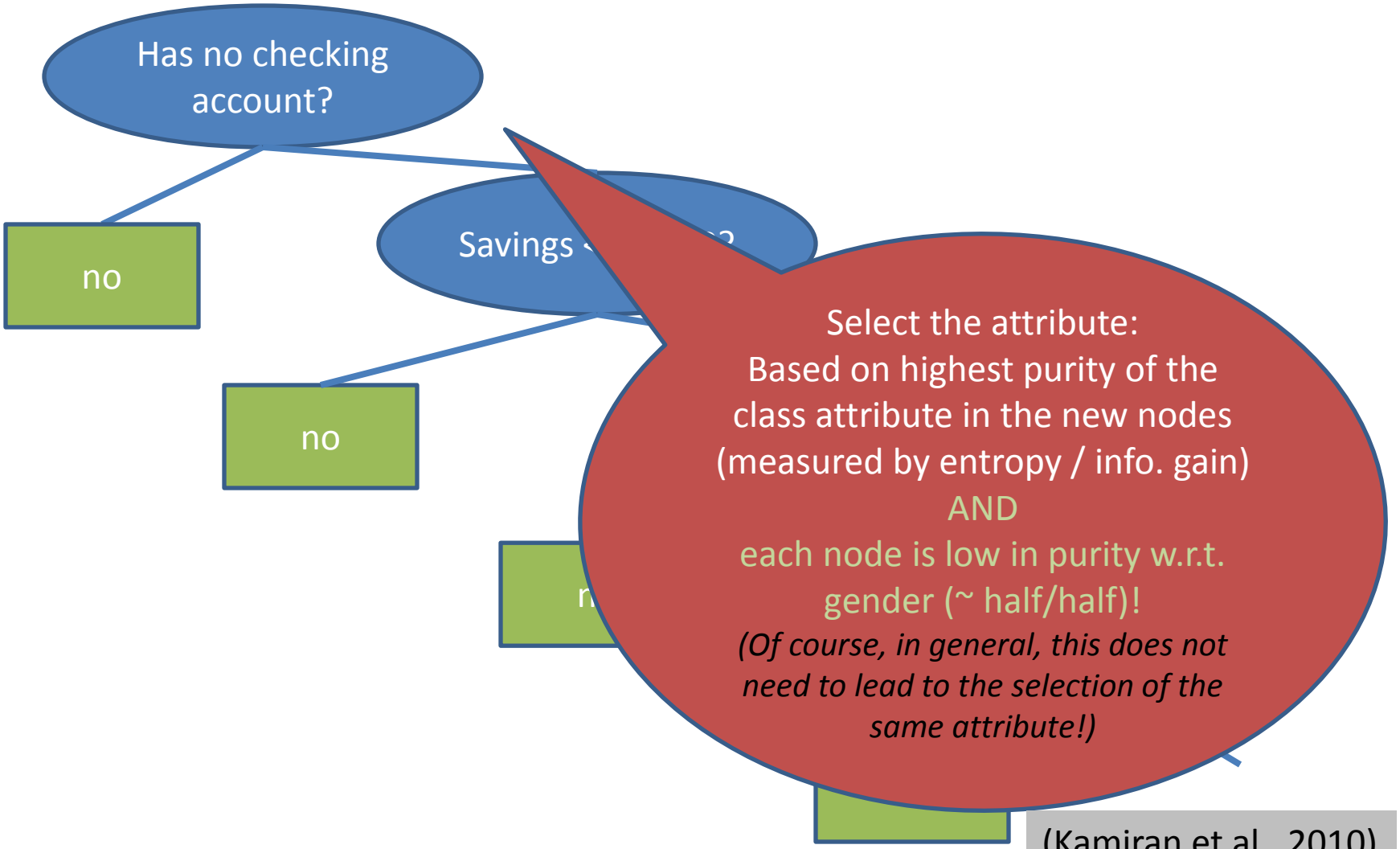
  where A is a PD item and B a PND item

Note: 2 uses/tasks of data mining here:

- Descriptive
  - "In the past, women who got a loan for a new car often defaulted on it."
- Prescriptive
  - (Therefore) "Women who want a new car should not get a loan."

(One way of)

Fixing data mining algorithms



(Kamiran et al., 2010)

Problem solved?

(Some) Conceptual challenges

The notion of discrimination

(or (un)fairness)

Discrimination =?= not distinguishing by a given attribute

(with some notes on possibly fundamental limits on formalization)

The Importance of Aristotelian Equality in Law

- Equality and discrimination are two sides of the same coin:

"The principle of equality requires that equal situations are treated equally and unequal situations differently. Failure to do so will amount to discrimination unless an objective and reasonable justification exists" – Explanatory memorandum protocol 12 to the ECHR

Westen: "Equality is an empty vessel, with no substantive moral content of its own"

1. When should individuals/situations be considered equals?
2. What constitutes equal treatment?
3. What should we be fundamentally concerned to equalize?

One would expect the law to provide answers to these questions

(Slide: thanks to Laurens Naudts, 2015)

Equal treatment = non-discrimination?

“I treat all my employees the same –
They enter the office by the stairs.”



A recent example:

Taddeucci and McCall vs. Italy

European Court of Human Rights judgment, 30/6/2016

The case concerned a refusal by the Italian authorities to grant a residence permit to a gay couple on family grounds.

The Court found in particular that the situation of Mr Taddeucci and Mr McCall, a gay couple, could not be understood as comparable to that of an unmarried heterosexual couple. As they could not marry or, at the relevant time, obtain any other form of legal recognition of their situation in Italy, they could not be classified as “spouses” under national law. ....

Thus the Court concluded … that there had been a violation of Article 14 (prohibition of discrimination) taken together with Article 8 (right to respect for private and family life) of the European Convention on Human Rights.

Summary: key problems of common computational notions of discrimination / (un)fairness

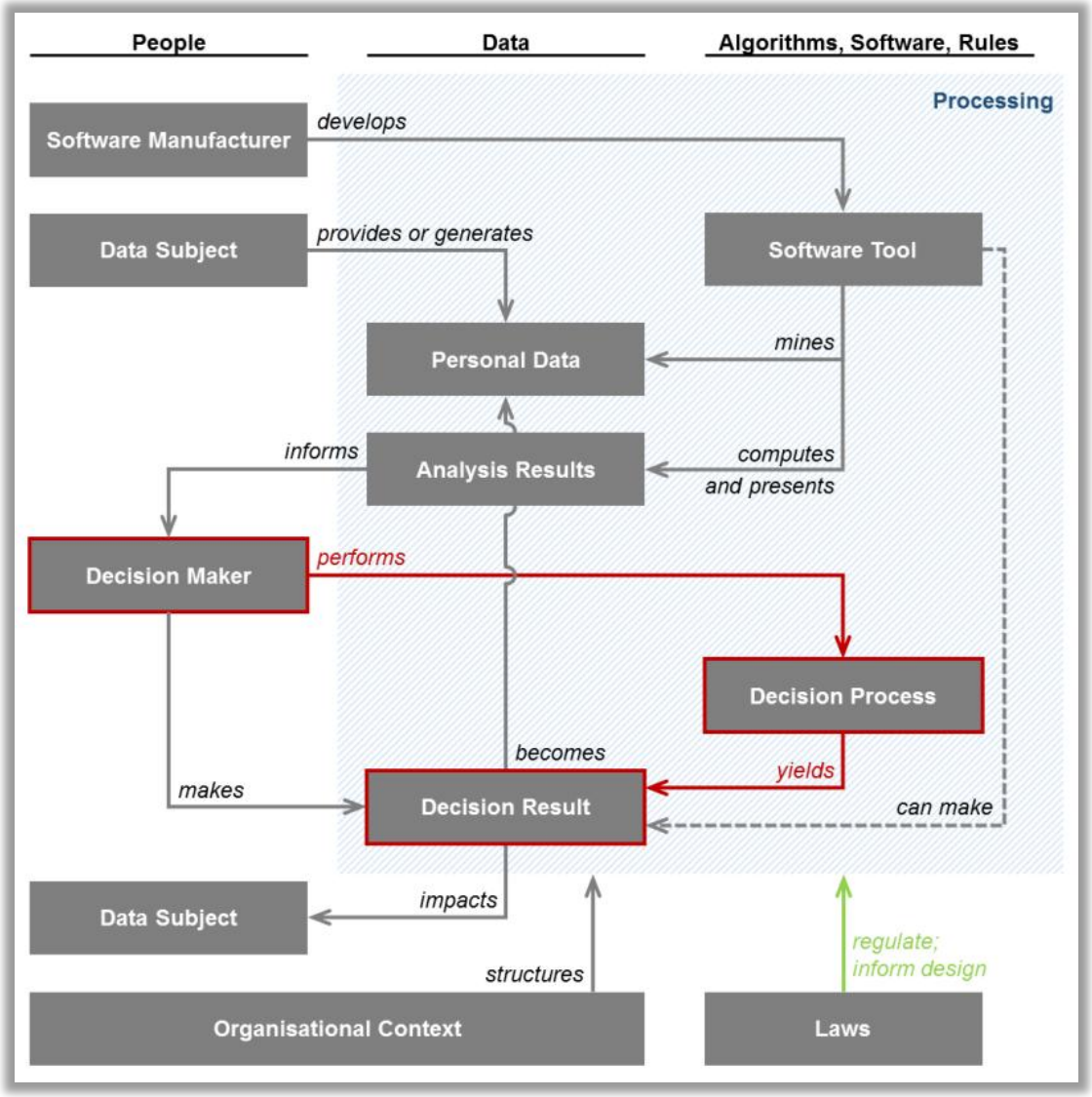
- Neglects the agency of those “treated” by data science / AI / etc.
- Including the goal: empowerment
- **Discrimination mitigation involves, crucially, on human agency and empowerment.**

People contribute to creating discrimination

… through their contribution to dataset creation

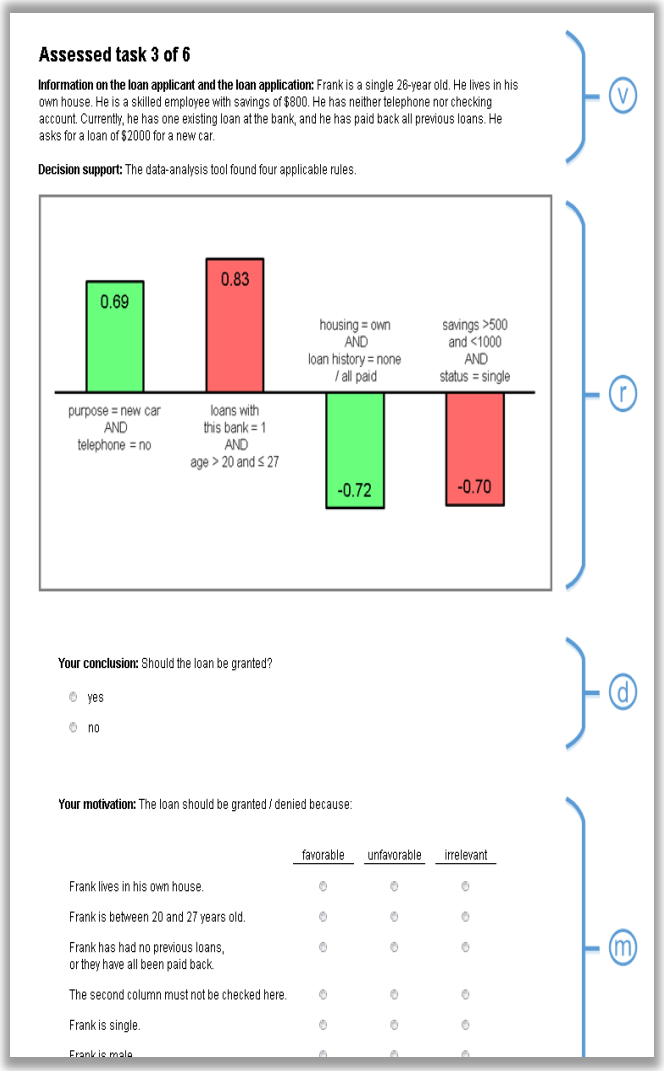
- (we've heard that in several talks, so only a reminder here)

… through their various structural involvements in decision making with data science



(Berendt & Preibusch, 2017) ; *SPION* and *Privacy in Social Networks* projects

Experiment with 215 mTurkers: Data mining assisted decision making and discrimination



Task structure

- Vignette, describing applicant and application
- Rules: positive/negative risks, flagged
- Decision and motivation, optional comment

Experimental variation

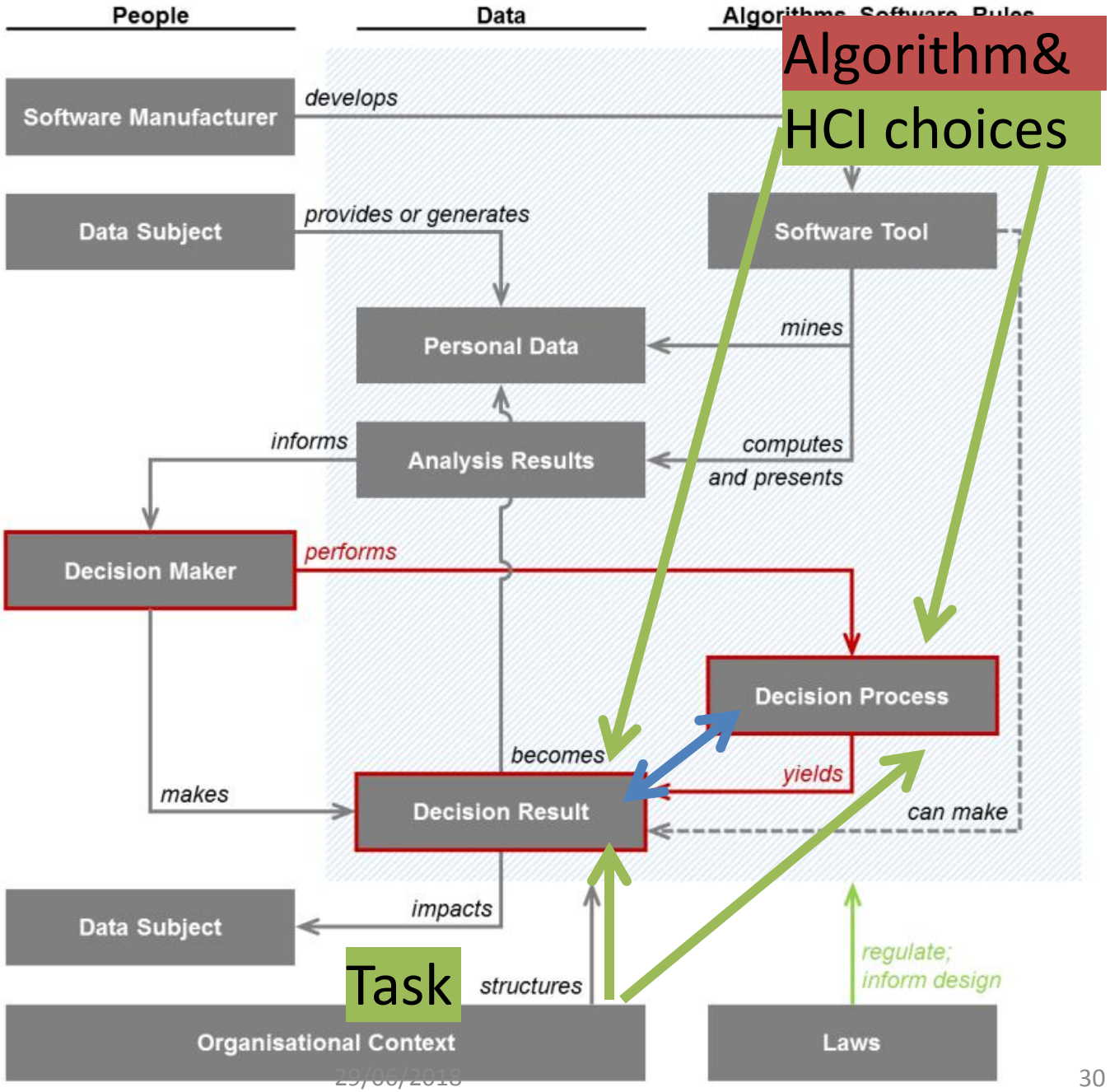
- Setting (ENFORCE: bank or DETECT: anti-discrimination agency)
- Data mining form / Interface

Results

- Mining form/interface: differential impact in settings
- Biases persist even if algorithms constrained

(Berendt & Preibusch, 2014, 2017)

Summary of findings



People’s contributions are needed to mitigate discrimination

Principal-agent relations: Two approaches

- Agent guarantees (properties of) behaviour to principal
- Principal can hold agent to account

Accountability: GDPR, Computer Science

Principles

Article 5

Principles relating to processing of personal data

1. Personal data shall be:

(a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');

2. The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('accountability').



Accountability: political science

(Bovens, 2007)

A relationship qualifies as a case of accountability when:

1. There is a relationship between an actor and a forum
2. in which the actor is obliged
3. to explain and justify
4. his conduct,
5. the forum can pose questions,
6. pass judgement,
7. and the actor may face consequences.

Privacy Eng./GDPR approach:
Accountability: a political science concept & the GDPR

A relationship qualifies as a case of accountability when:

1. There is a relationship between an actor and a forum
2. in which the actor is obliged
3. to explain and justify
4. his conduct,
5. the forum can pose questions,
6. pass judgement,
7. and the actor may face consequences.

(Bovens, 2007)

"Technical and organisational measures" relating to

processing of pers. data per se

GDPR obligations

Rec. 71, Art. 13-15, 20

Art. 5 (2)

Art. 13-15, 20, 22; 5(2)

) Art.

) 83, 58

Conclusion: This is another instance of

"**we need careful, interdisciplinarily grounded, use of terminology and concepts**"

Accountability

=/= Transparency

Accountability

=/= Responsibility

=/= Liability

And: each of these issues has a long history and even "solutions" – and **we should draw on the law and not only on ethics**

References

Article 29 Data Protection Working Party (2010). Opinion 3/2010 on the principle of accountability (00062/10/EN WP 173). http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp173_en.pdf

Berendt, B. & Preibusch, S. (2014). Better decision support through exploratory discrimination-aware data mining: foundations and empirical evidence. Artificial Intelligence and Law, 22 (2), 175-209. https://people.cs.kuleuven.be/~bettina.berendt/Papers/berendt_preibusch_2014.pdf

Berendt, B. & Preibusch, S. (2017). Toward accountable discrimination-aware data mining: The importance of keeping the human in the loop – and under the looking-glass. Big Data, 5(2). DOI: 10.1089/big.2016.0055. https://people.cs.kuleuven.be/~bettina.berendt/Papers/berendt_preibusch_2017_last_author_version.pdf

Bovens M. (2007). Analysing and assessing public accountability. A conceptual framework. European Law Journal, 13, 447–468.

European Court of Justice (2011) Case C-236/09, Association Belge des Consommateurs Test-Achats ASBL and Others v Conseil des ministres. http://curia.europa.eu/juris/liste.jsf?language=en&num=C-236/M099

Kamiran, F., Calders, T., & Pechenizkiy, M. (2010). Discrimination Aware Decision Tree Learning. ICDM 2010: 869-874. http://wwwis.win.tue.nl/~tcalders/pubs/TR10-13.pdf

Luong BT, Ruggieri S, Turini F (2011). k-nn as an implementation of situation testing for discrimination discovery and prevention. In: KDD, pp 502–510. ACM.

Naudts, L. (2015). Algorithms – Legal Framework. Presentation in the Privacy and Big Data course, KU Leuven.

Pedreschi D, Ruggieri S, Turini F (2008) Discrimination-aware data mining. In: Proceedings of KDD'08, pp 560–568. ACM. http://www.di.unipi.it/~ruggieri/Papers/kdd2008.pdff

Sweeney, L. (2013). Discrimination in Online Ad Delivery. Communications of the ACM, 56(5), 44-54. cited from earlier version available at SSRN:

http://ssrn.com/abstract=2208240 or http://dx.doi.org/10.2139/ssrn.2208240

General Data Protection Regulation: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG