OpenBanking, Data Protection, Machine Learning and Credit Risk Models
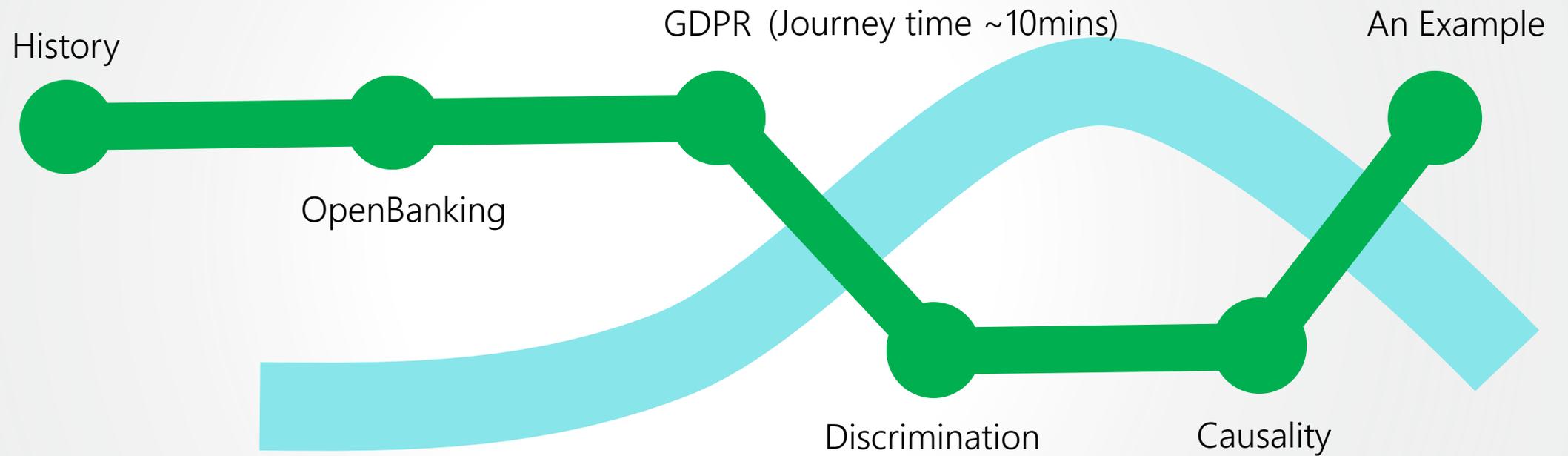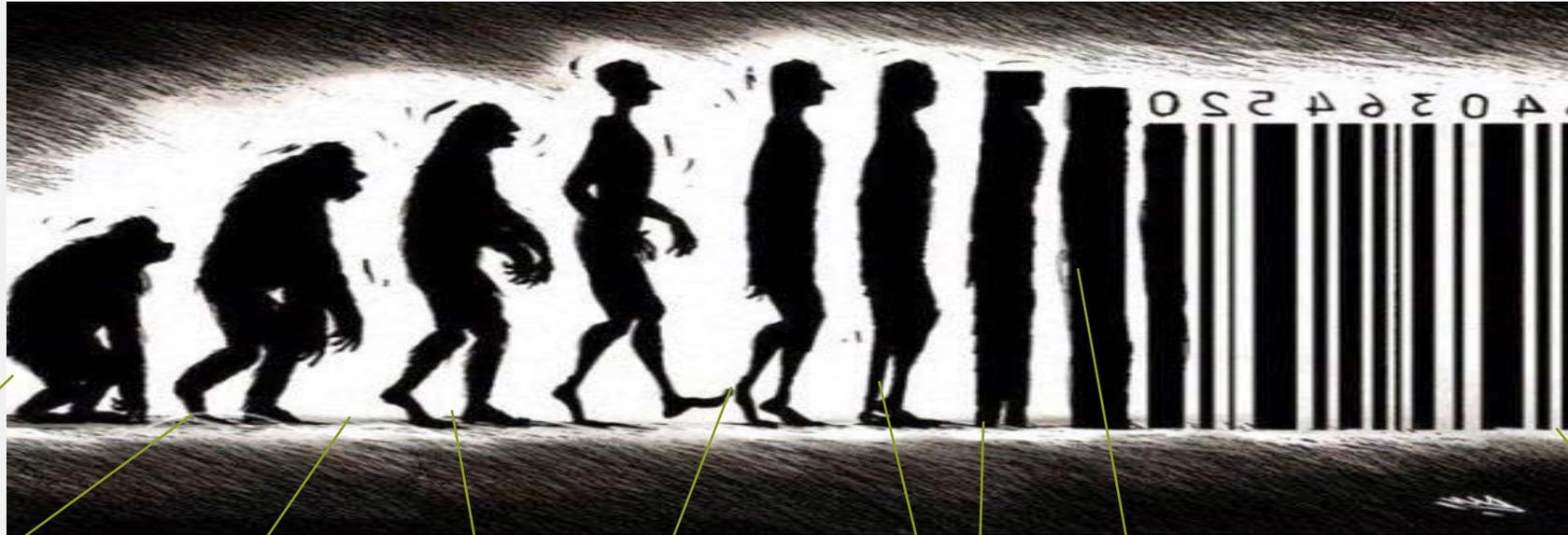
Is my automated decision fair?

Mark Somers

A Map of this talk

History

OpenBanking

GDPR (Journey time ~10mins)

Discrimination

Causality

An Example

Credit Regulation – an historical perspective

Innovate

Regulate

Avoid

700 BC Jews banned usury to other Jews. 325 AD 1st Council of Nicaea - Christians

1290 Edward I expels Jews from England because he owes them a fortune

1640 Charles I seizes the gold held by individuals at Mint and replaces with debt

1970 Fair Credit Reporting Act (US) requires explanation of reasons for "adverse decisions" and accurate information

2018 (EU) GDPR & Open Banking

2000BC Babylonians create harvest day lending

1345 Edward III defaults to Venetians on debts used to fund 100 years war

1899 Retail Credit Company (now Equifax) set up to build credit files in Georgia USA

1958 FICO builds first credit score

Renaissance in Banking?

- Venetian banking combined lending, futures market, weather insurance, trade credit, shipping insurance. Avoided ban on usury by using "bills of exchange" which used differentials in exchange rates

- These bankers rapidly developed a complex product set – they were facilitated by the invention of double entry book-keeping in early 1300's - improving data quality

Could Open Banking lead to a new renaissance in lending?

- Facilitate trusted 3rd parties to use transactional information. End of information advantage of current account service providers (mainly big banks).

- True single view of customer even for multiply banked customers should be feasible.

- Enable new services to move money around for consumers to optimise rates

- Reduce costs of transactions by reducing layers in payment processing

- **Opportunities for machine learning / AI**

Programmatic Banking (AKA Banking as a Service)

- What is it? Enables users to encode complex payment instructions to create bespoke financial products.

- Relatively Simple Credit Examples:
  - Collections tools to support fair collections for customers in financial hardship. Identifying when a customer has disposable income – potentially releasing money back in an emergency.
  - Facilitating payment sharing
  - Debt allocation optimisation across loans

- More sophisticated applications:
  - Electronic contracts – "signing" a contract would set-up a future payment from client to service provider. Bank has a key role:
    - Provide credit to service provider backed by the security of the contract,
    - Provide guarantee to client of delivery
    - Provide funding to client (where appropriate) and evidence of ability to pay

To use these services consumers will have to provide more information (and give more control) over their finances than before.

GDPR: Key Requirements for Credit Risk

GDPR comes into force in May 2018

- GDPR is much wider than credit risk (covering anyone who stores data and processes it). It will however have an impact on Credit Risk practices:
  - *The right to be informed* (*what data is being captured, stored, processed and for what purpose*)
  - The right of access.
  - The right to rectification.
  - The right to erasure.
  - The right to restrict processing.
  - The right to data portability.
  - The right to object *(in general but particularly to marketing activity)*
  - *Rights in relation to automated decision making and profiling* (*to understand how automated decisions are used*)

GDPR Profiling Requirements (ICO)

Organisations must in addition to current Data Protection Act (DPA) requirements:

- Ensure processing is fair and transparent by providing meaningful information about the logic involved, as well as the significance and the envisaged consequences.

- Use appropriate mathematical or statistical procedures for the profiling.

- GDPR requires you to demonstrate **how** you comply with the principles (DPA just required that you complied)

ICO Paper: "Big data, artificial intelligence, machine learning and data protection"

Selected quotes from the ICO paper:

- The autonomous and opaque nature of machine learning algorithms can mean that decisions based on their output may only be identified as having been discriminatory afterwards – when the effects have already been felt by the people discriminated against. For instance, ProPublica analysed 7,000 'risk scores' produced by a machine learning tool used in some US states to predict the future criminal behaviour of defendants. The findings revealed discrimination based on race, with black defendants falsely classified as future criminals on nearly twice as many occasions as white defendants. Detecting discriminatory decisions in hindsight will not be sufficient to comply with the accountability provisions of the GDPR. Big data **analysts will need to find ways to build (unfair) discrimination detection into their automated systems** to prevent such decisions being made in the first place.

- The distinction between correlation and causation is very important

- Big data organisations therefore need to exercise caution before relying on machine learning decisions that cannot be rationalised in human understandable terms. If an insurance company cannot work out the intricate nuances that cause their online application system to turn some people away but accept others (however reasonable those underlying reasons may be), how can it hope to explain this to the individuals affected?

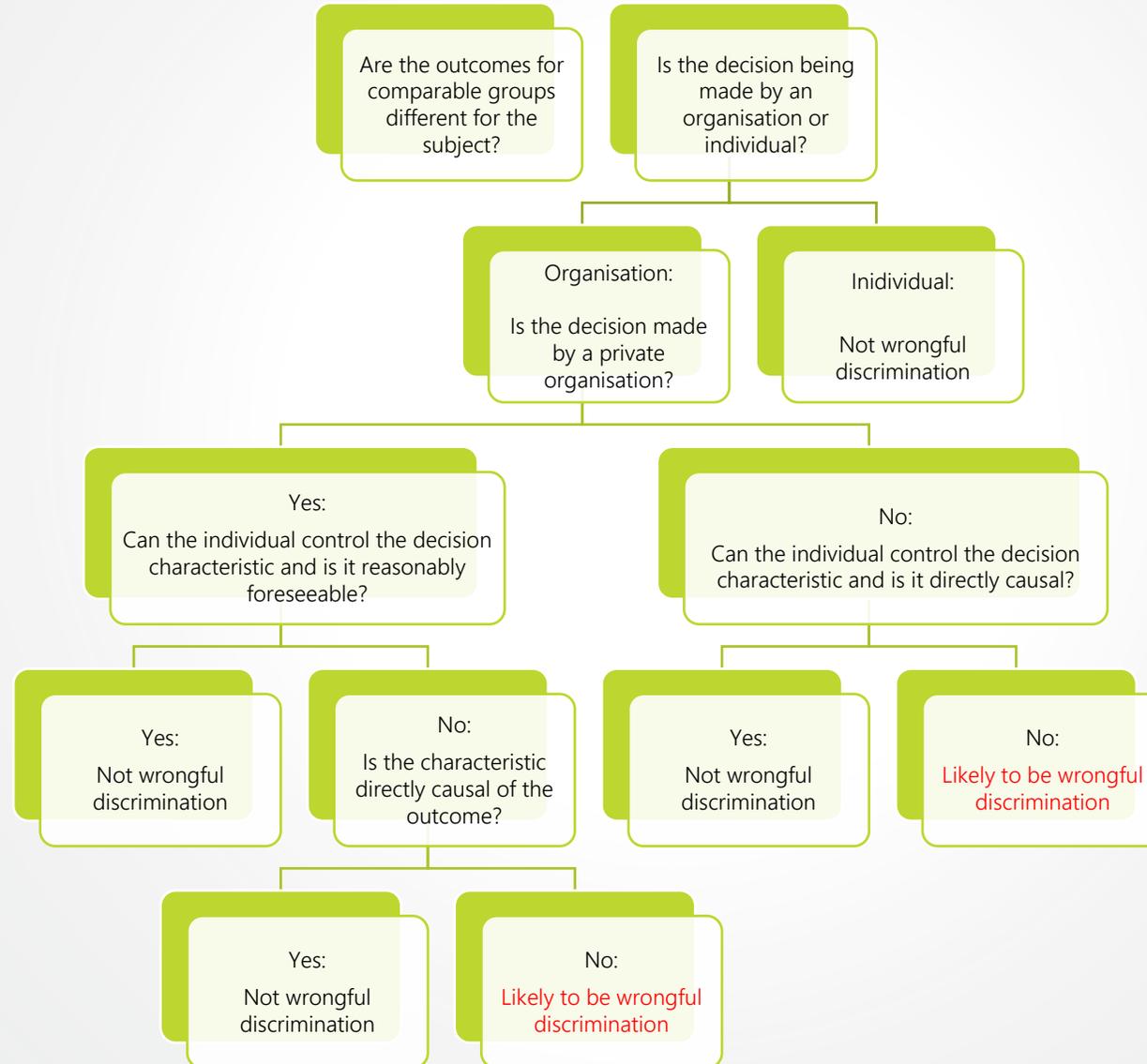# How do you build discrimination detection into algorithms?

To automate discrimination detection you need to define what discrimination is (it is not just limited to using predefined high sensitivity fields such as gender or race, but needs to be defined more broadly and fundamentally):
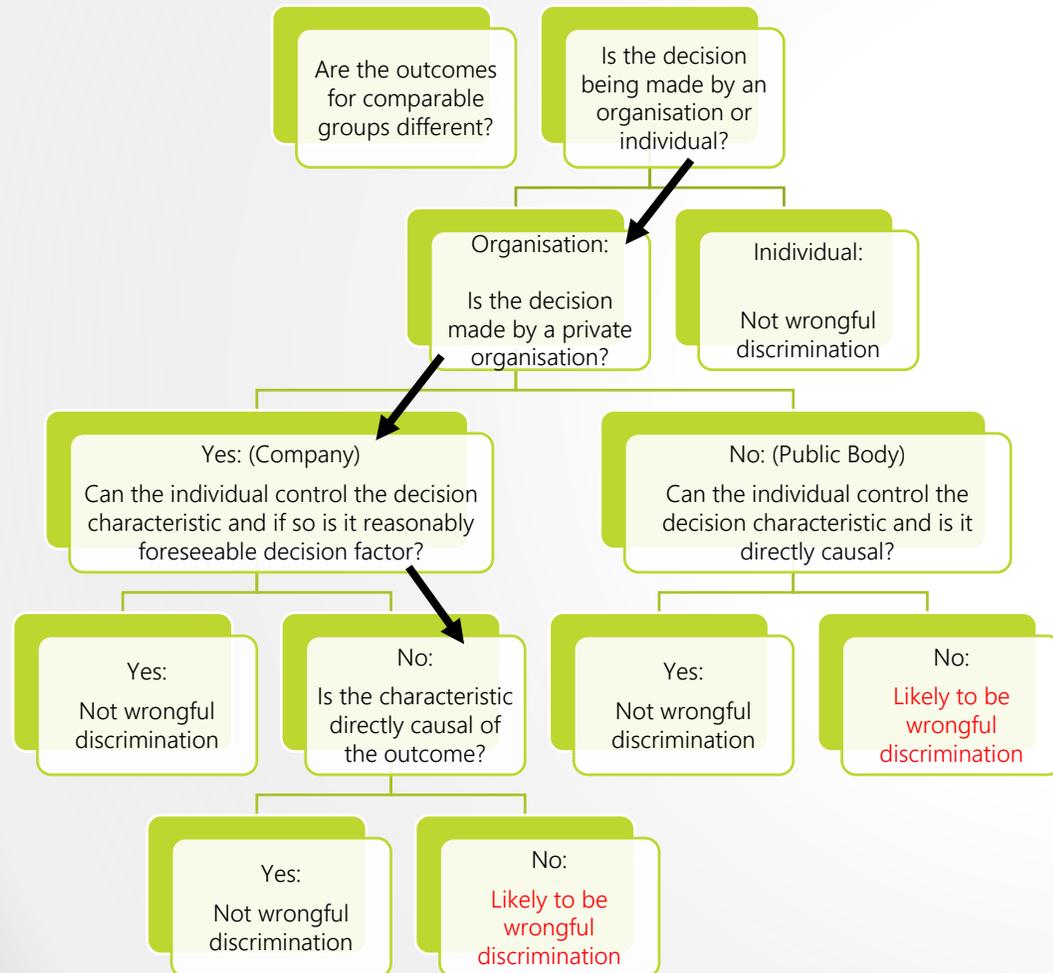
Propositions:

- Wrongful Discrimination is by its nature comparing the outcomes of two or more groups. To be meaningful these groups must be comparable and the outcome must matter to the individual. (GDPR)

- Wrongful discrimination applies to corporations or public bodies but not individuals. (GDPR)

- Wrongful for a public body to discriminate due to some aspect that is not directly causally associated with the objective of the selection rule (ICO paper)

- It is unfair for corporations to discriminate due to some aspect that defines the group that they cannot reasonably control or that is not directly causal of the outcome selected for (ICO paper)

- Any discrimination on features that are within an individual's control should be transparent (GDPR)

- It is not wrongful to discriminate on any basis if the individual has freely given permission and the consent is not dependent on service (GDPR)

These requirements are largely restatement of GDPR requirements but seem plausible statements of good practice in any case.

A proposed decision tree to identify discriminatory decisions:

Are the outcomes for comparable groups different for the subject?

Is the decision being made by an organisation or individual?

Organisation:

Is the decision made by a private organisation?

Inidividual:

Not wrongful discrimination

Yes:

Can the individual control the decision characteristic and is it reasonably foreseeable?

No:

Can the individual control the decision characteristic and is it directly causal?

Yes:

Not wrongful discrimination

No:

Is the characteristic directly causal of the outcome?

Yes:

Not wrongful discrimination

No:

Likely to be wrongful discrimination

Yes:

Not wrongful discrimination

No:

Likely to be wrongful discrimination

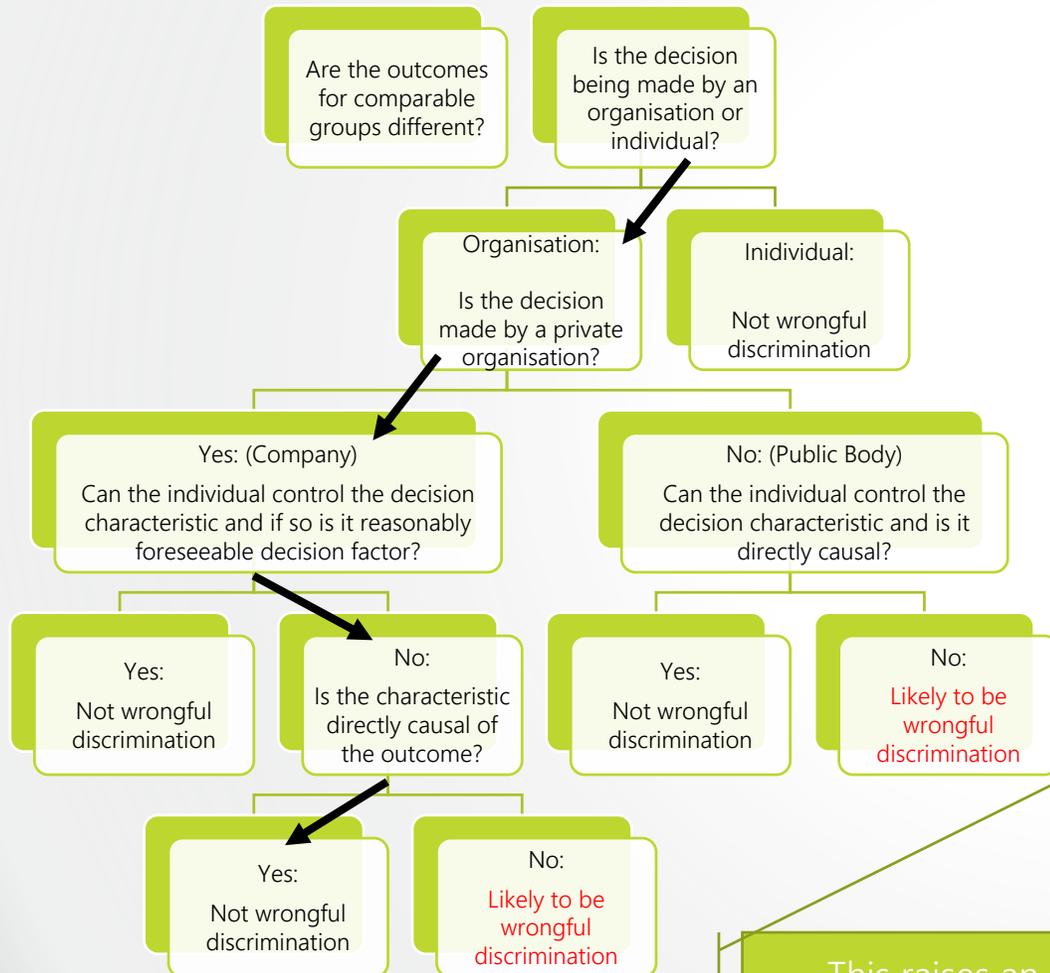# Example 1: Age for credit decisions

- The key choice: is the characteristic a cause of default?
  - For mortgage lending arguably lending to people into retirement is causal of higher risk and is therefore a legitimate decision characteristic
  - Most scoring algorithms associate young people with higher risk – being young may not be a direct cause, it may simply be a correlation which young people cannot control. That would imply it is wrongful discrimination.

Example 2: Bureau Searches for credit decisions
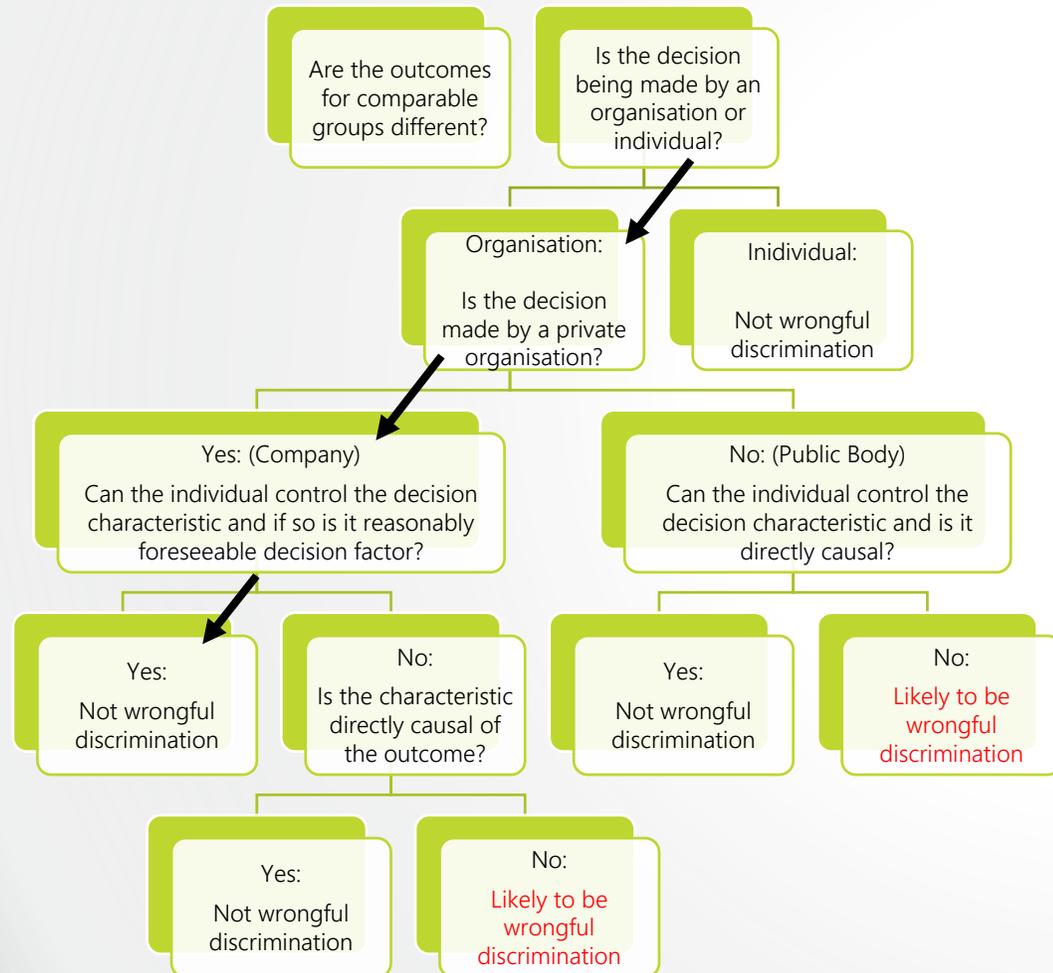
- The key choice: is the characteristic under the individuals control and is it a foreseeable decision factor?
  - In this case consumers can control how many credit searches they make
  - It is also reasonably well known that many searches would reduce the chance that a bank would accept a customer

Example 3: Employment Status for credit decisions

4most

Are the outcomes for comparable groups different?

Is the decision being made by an organisation or individual?

Organisation:

Is the decision made by a private organisation?

Inidividual:

Not wrongful discrimination

Yes: (Company)

Can the individual control the decision characteristic and if so is it reasonably foreseeable decision factor?

No: (Public Body)

Can the individual control the decision characteristic and is it directly causal?

Yes:

Not wrongful discrimination

No:

Is the characteristic directly causal of the outcome?

Yes:

Not wrongful discrimination

No:

Likely to be wrongful discrimination

Yes:

Not wrongful discrimination

No:

Likely to be wrongful discrimination

- The key choice: is the characteristic under the individuals control and is it causal of credit default?
  - In this case consumers cannot easily directly control their employment status
  - Considering causality it is the case that being unemployed or self-employed might be directly causally related to higher credit risk
  - On this basis it should be fair for banks to use in credit decisions

This raises an important refinement. In practice we rarely know the causal relationships – therefore it should be the decision systems reasonably held belief about causality that is relevant

Can we apply this unfair discrimination approach to complex algorithms?

- Required inputs to the determinations are:
  - Is the decision being made by a public or private organisation?
  - Can the subject reasonably control the decision characteristic? ("do" operator)
  - Is discrimination on this characteristic a reasonably foreseeable consequence of the subjects behaviour with regards the decision being made?

- Required to be extracted from the decision algorithm:
  - For non-controllable characteristics is the characteristic directly causal of the outcome?

- Most machine learning (or statistical) approaches do not lend themselves to providing humanly intelligible "reasons" for their decisions.

- This is because a "reason" implies causality – and most statistical/machine learning tools model conditional probabilities (ie association) only.

- It is not the model type that determines this limitation but the way the model is built.

What do I mean by "Directly Causal"?

- Specifically I mean the "natural direct effect" as defined by J. Pearl (2009) but applying the do() only where the consumer can control the outcomes.
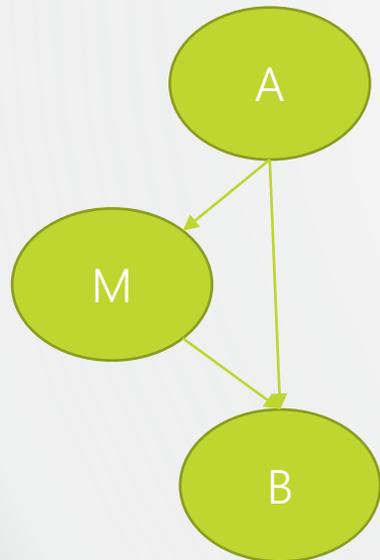
Calculate $p(B \mid do(M_{A=0}), A=1)$ for all M where controllable by the consumer

Calculate $p(B \mid A=1)$ for all M where not-controllable by the consumer

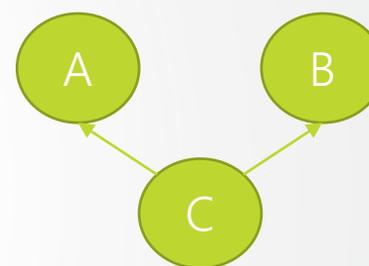If p(B) depends on A under this operation then it is a direct cause as far as the consumer is concerned.
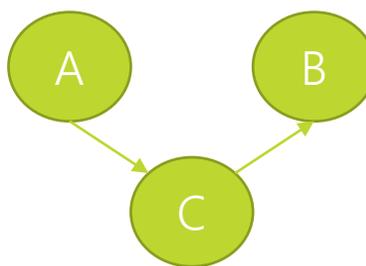
Algorithms that learn causal relationships

- Early work by Verma & Pearl with IC (Inductive Causation) Algorithms. These rely on the simple observation that conditional independence implies some evidence of causal mechanism:

Vs
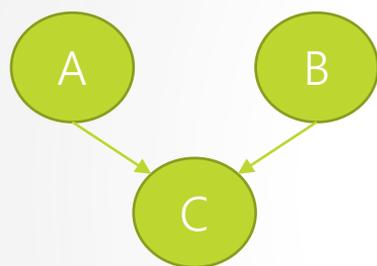
(A independent B)

(A dependent B)

- This approach may not be robust with limited real data where measuring independence is noisy and thus structure recovery is not consistent
- Extended to cover the possibility of "unobserved" or latent characteristics. For example A->B could equally be A<-L->B (ie L, a latent variable, is a common cause of A & B). IC* is an algorithm that can identify direct causes in some circumstances, they are denoted: A ->* B
- More recently algorithms (ANM, IGCI, D2C) using approaches that use other measures (beyond conditional correlation) and limited to pairwise comparison for computational speed in large problems. These do not typically handle latent inference.
- These approaches may be enhanced by using domain knowledge of event timing – for example a default event cannot cause the characteristics as observed at loan origination (but a latent variable could be a common cause of both)
- A simplification: To detect unfair discrimination we do not need to determine the full causal structure of the domain. We need to identify only the causality (or otherwise) of the characteristics over which the subject does not have control and their relationship to the outcome of interest.

Extension of IC* algorithm to detect unfair discrimination

- In our situation a better measure of independence (following Bontempi & Flauer) is based on conditional mutual information between characteristics x & y conditioned on z:

$$I(x,y|z) = \sum \sum \sum p(x,y|z).\ln\left(\frac{(p(x,y|z))}{p(x|z)p(y|z)}\right)$$

Higher mutual information implies stronger relationship

Limit evaluation to single conditioning characteristic

Information theoretic approach matches decision tree/forest approaches and common IV screening approach used in many scorecard builds

4most

The problem of indistinguishability

- An example of an indistinguishable 3 variable problem is where all 3 variables are pairwise correlated even when conditioned on the 3rd.

Or   Or   ...
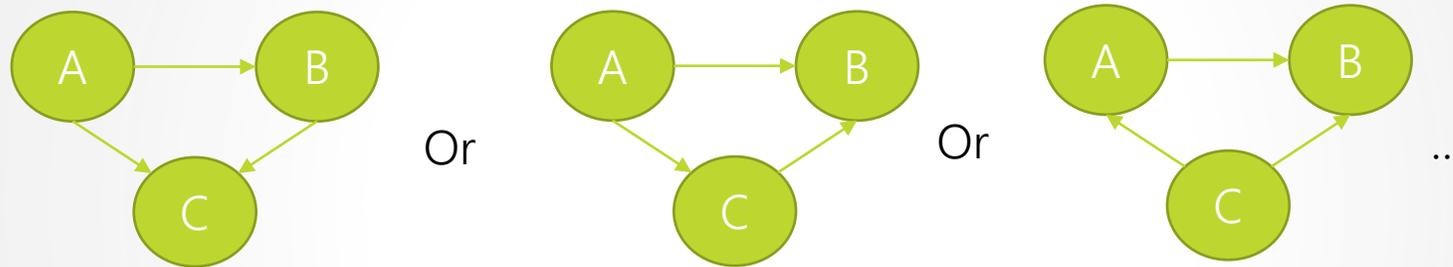
- What is the implication for unfair discrimination for cases where the causal relationship is not distinguishable?
  - A cautious approach would be to treat these relationships as non-causal for the purposes of unfair discrimination detection?

A simplified approach to determine unfair discrimination

- Step 1: Estimate probability of default conditional on other characteristics, $x_1,...,x_N$ (this is the model to be used to drive decisions – there is no direct restriction on model specification or type)

- Step 2: Estimate pair-wise joint probabilities between $\{x_1,...,x_N\}$ vs default conditional on all combinations of other x's.

- Step 3: Calculate conditional mutual information for each pair, $I(i,default \mid j)$– create unstarred directed network between nodes and default if conditioning does not remove mutual information ($I>0$). (Markov blanket) – based on prior knowledge that Default cannot cause predictors

- Step 4: For each x_i , x_j in the Markov Blanket for default $I(i,j)$. Where greater than 0 then add an undirected node.

- Step 5: For each x_p , x_q not in the Markov Blanket for default with mutual information to x_m in the markov blanket determine $I(p,q)$

- Step 6: Add a direct causal link (starred arrow) from x_m to Default where conditional independence is found between x_i and x_j but x_k not required to condition (This is based on Verma and Pearl's IC* algorithm)

Any characteristic with a direct causal link (->*) are legitimate decision characteristics even if they are not controllable by the subject.

Any other characteristic in the model of Step 1 must be controllable by the subject.

Estimating Step 1 and 2 may be performed using raw frequency data or in environments with more characteristics can be modelled using regressions, random forests etc. etc.
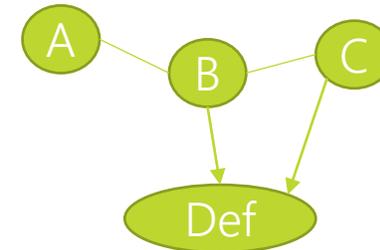
A simple example

I(x,Default|y)	Default	A	B	C	y's
A	0.1	0	0.001	0.05	
B	0.2	0.15	0	0.1	
C	0.1	0.1	0.04	0	
x's					
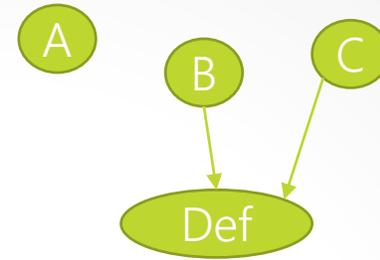
I(x,y)	A	B	C
A		0	0.08	0.001
B		0.07	0	0.08
C		0.001	0.07	0

Step 1,2,3

Step 4
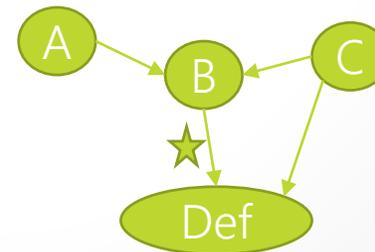
Step 5 & 6

Evaluation on Credit Data.

- Data extracted from US public website on individual loan performance: https://www.lendingclub.com/

- Used data from Q12016. Contains 133k records with a "bad rate" of circa 9% (write-off, in default or >30 DPD)

- Only modelled the funded population (ignored problems of reject inference)

- Removed information at outcome, bureau score, and internal grading system as predictors

- Screened on IV to identify most predictive 20 Characteristics

- 9 Characteristics enter our simple mode. This model is relatively weak (33% Gini) – their internal grading system has Gini c42%

Partial Structural Model recovered
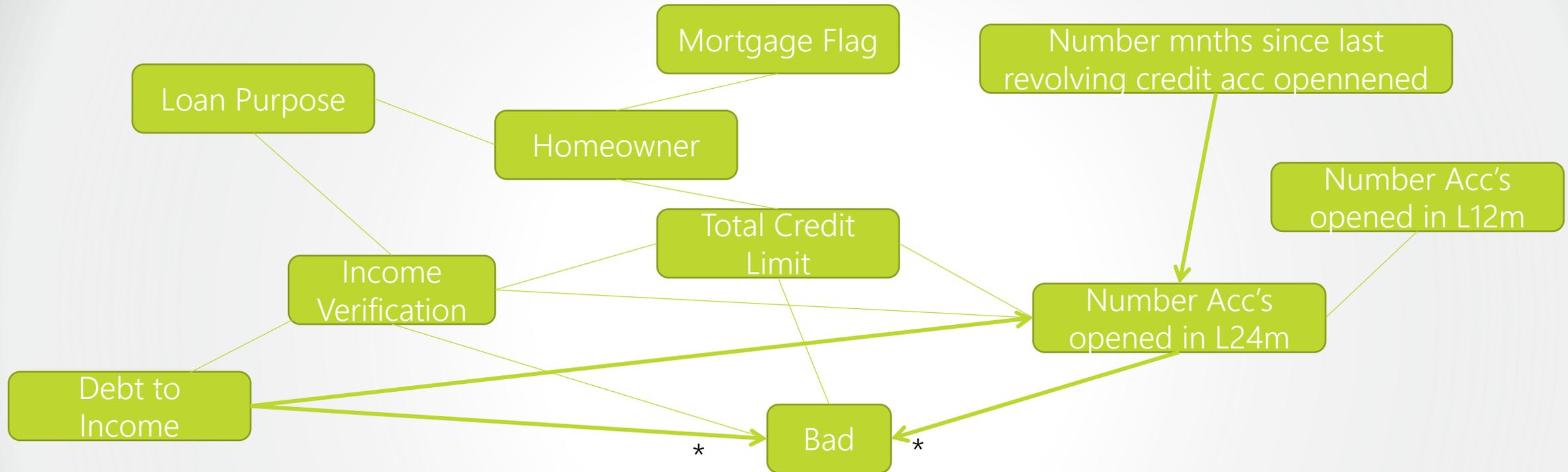
- Direct causal relationships of default were identified from the data available between Debt to Income and Default & Number of Accounts opened in L24m and Default. Signified by →* on diagram.
  - Inference based on observed independence of DTI and Number of months since last Revolving credit acc openned
- Characteristics that are significant in the scoring model are not necessarily in the Markov Blanket. Removing these characteristics only reduces Gini from (33% to 31%)

Results

	In Subjects Control?	Transparent?	Directly Causal?	Unfair Discrimination?
Debt To Income	N	Y	Y	N
Income Verification	N	Y	N	Y
Number of Account Openned in L24m	Y	Y	Y	N
Total Credit Limit	Y	?	N	N
Loan Purpose	Y	?	N	N
Home Owner	N	Y	N	Y
Mortgage	N	N	N	Y
Number of Months since Lst Rev Acc Open	Y	Y	N	N
Number of Account Openned in L12m	Y	Y	N	N

- Income Verification is potentially a characteristic that would amount to unfair discrimination. A self employed person cannot easily verify their income but this is not a cause of them being higher risk.
  - Homeownership and Mortgage are also potentially an unfair basis on which to discriminate (can't control and not causal) however this is not material because they are absorbed in other aspects of the model.
- Instead could include Loan Purpose back in the model. This gives a model with Gini 30% - so not a massive effect (without replacing Loan Purpose Gini goes to 26% so Causal structure helps to identify how to mitigate loss of characteristic.

A few concluding thoughts...

- The proposed algorithm isn't complete in the sense that creating the full causal network might uncover causal relations that would be missed by focussing on the default relationships only

- It is perfectly possible to disagree with the fair decisioning principles

- There are other significant challenges in agreeing whether a consumer can control a given characteristic or indeed understand their likely use in decisioning.