

The “Artificial” Intelligence of Law

RPA research group on Segregation, Polarisation and Inequality
Amsterdam, 18 December, 2025



Giovanni Sileno
g.sileno@uva.nl

University of Amsterdam

Outline of the talk

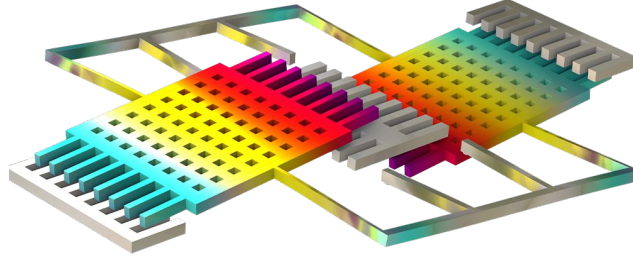
- who am I?
- historical outline of the research tracks in AI & Law
- some examples of
 - the struggle quantitative vs qualitative approaches to law,
 - on the challenges of norm modelling,
- on the fundamental issues with defining/measuring (un)fairness
[work in progress]

My positions in academia

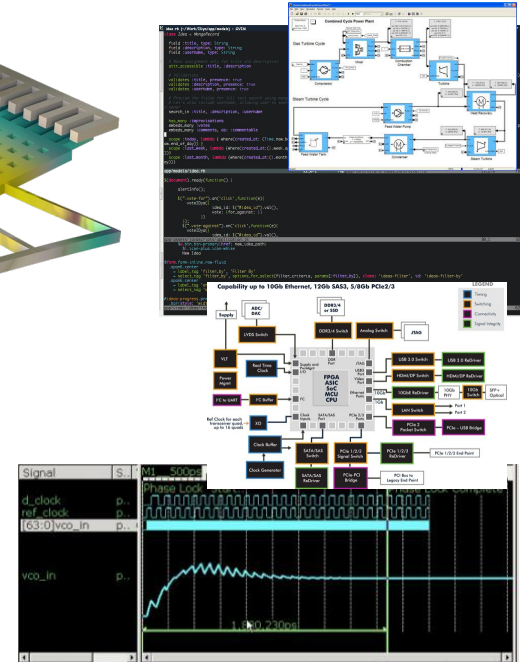
- M.Sc. Electronics Engineering, Politecnico of Torino and ENSIMAG

My positions in academia

- M.Sc. Electronics Engineering, Politecnico of Torino and ENSIMAG

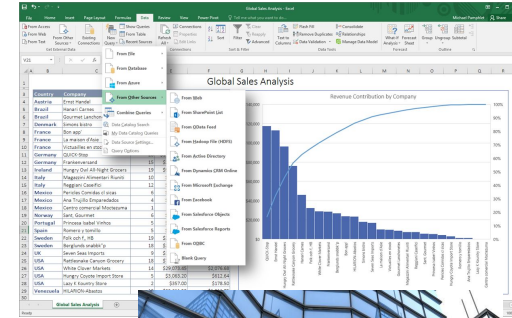


My thesis was about the conception/design of **autonomous embedded systems** through **modeling and simulation**, integrating: *hardware models, software models, environmental (probabilistic) models.*



Reality check

After my degree, I worked as **business analyst** (McKinsey & co.), dealing with several organizations.



Reality check

After my degree, I worked as *business analyst* (McKinsey & co.), dealing with several organizations.

I learned that:

- ❖ most organizational design decisions are based upon common-sense heuristics, implicit models/criteria

⇒ **no modeling and simulation!**

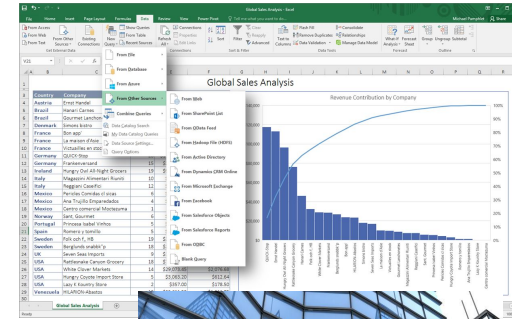


Reality check

After my degree, I worked as **business analyst** (McKinsey & co.), dealing with several organizations.

I learned that:

- ❖ most organizational design decisions are based upon common-sense heuristics, implicit models/criteria
⇒ **no modeling and simulation!**
- ❖ lack of transparency on design choices implied **opacity** and **resistance to contextualization** in operations.
⇒ **root cause of “bureaucratic alienation”**

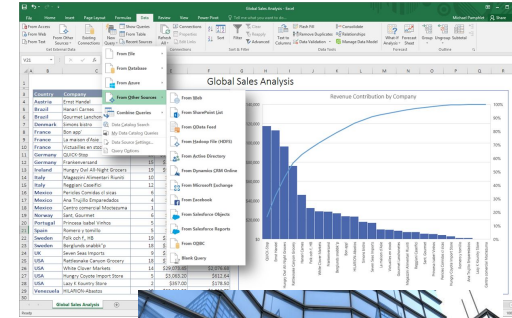


Reality check

After my degree, I worked as **business analyst** (McKinsey & co.), dealing with several organizations.

I learned that:

- ❖ most organizational design decisions are based upon common-sense heuristics, implicit models/criteria
⇒ no modeling and simulation!
- ❖ lack of transparency on design choices implied **opacity** and **resistance to contextualization** in operations.



*need for tools for **systematic** institutional design and analysis!*

My positions in academia

- M.Sc. Electronics Engineering, Politecnico of Torino and ENSIMAG
- PhD in AI & Law, Leibniz Center for Law (Law Faculty), UvA



Aligning Law and Action

Giovanni Sileno

My positions in academia

- M.Sc. Electronics Engineering, Politecnico of Torino and ENSIMAG
- PhD in AI & Law, Leibniz Center for Law (Law Faculty), UvA
- Postdoc in Cognitive Systems, Télécom Paris and Paris Dauphine University

My positions in academia

- M.Sc. Electronics Engineering, Politecnico of Torino and ENSIMAG
- PhD in AI & Law, Leibniz Center for Law (Law Faculty), UvA
- Postdoc in Cognitive Systems, Télécom Paris and Paris Dauphine University
- Postdoc in Regulation of Data-Sharing Infrastructures, CCI/SNE, Ivi, UvA

My positions in academia

- M.Sc. Electronics Engineering, Politecnico of Torino and ENSIMAG
- PhD in AI & Law, Leibniz Center for Law (Law Faculty), UvA
- Postdoc in Cognitive Systems, Télécom Paris and Paris Dauphine University
- Postdoc in Regulation of Data-Sharing Infrastructures, CCI/SNE, Ivi, UvA
- Assistant Professor in Socially Intelligent Artificial Systems, SIAS, Ivi, UvA

My positions in academia

- M.Sc. Electronics Engineering, Politecnico of Torino and ENSIMAG
- PhD in AI & Law, Leibniz Center for Law (Law Faculty), UvA
- Postdoc in Cognitive Systems, Télécom Paris and Paris Dauphine University
- Postdoc in Regulation of Data-Sharing Infrastructures, CCI/SNE, IvI, UvA
- Assistant Professor in Socially Intelligent Artificial Systems, SIAS, IvI, UvA

	M.Sc.	PhD	Postdoc 1	Postdoc 2	Assistant
COMPUTATION					
NORMATIVE SYSTEMS					
COGNITIVE MODELLING					
SOCIO-TECHNICAL SYSTEMS					
APPLIED MACHINE LEARNING					

my research efforts/interests

	DESCRIPTIVE	PRESCRIPTIVE
INDIVIDUAL AGENCY	cognitive modelling	agent-based programming
COLLECTIVE AGENCY	agent-based modelling (and simulation)	computational legal theory

my research efforts/interests

	DESCRIPTIVE	PRESCRIPTIVE
INDIVIDUAL AGENCY	cognitive modelling STS: narratives and values	agent-based programming
COLLECTIVE AGENCY	agent-based modelling (and simulation)	computational legal theory STS: co-design, co-creation methods

A brief outline on AI & Law research tracks

But first a crucial disclaimer:

AI for Law $\neq$ Law for AI

But first a crucial disclaimer:

AI for Law $\neq$ Law for AI

We devise and use computational methods for the study of and for applications in legal/normative systems at large

We do legal (normative/policy) scholarship about computational infrastructures

But first a crucial disclaimer:

AI for Law $\neq$ Law for AI

We devise and use computational methods for the study of and for applications in legal/normative systems at large

We do legal (normative/policy) scholarship about computational infrastructures

Here we **REGULATE**
(contemporary, mainstream)
technologies and their uses.

But first a crucial disclaimer:

AI for Law $\neq$ Law for AI

We devise and use computational methods for the study of and for applications in legal/normative systems at large

Here we **DISCOVER/DESIGN** (new) techniques/processes for analysing or implementing regulations

We do legal (normative/policy) scholarship about computational infrastructures

Here we **REGULATE** (contemporary, mainstream) technologies and their uses.

The beginning (?)

The beginning (?)

Leibniz (also a lawyer) strongly believed that calculation would be the key to settle all human conflicts and disagreements.



Leibniz: ***Calculemus!*** ~1686

The beginning (?)

Leibniz (also a lawyer) strongly believed that calculation would be the key to settle all human conflicts and disagreements.

**...using characteristics numbers
instead of concepts:**

animal = 2

rational = 3

man = rational animal = $2 \times 3 = 6$



Leibniz: ***Calculemus!*** ~1686

The beginning (?)

...building upon and further
strengthening the analytical tradition!

iustum, licitum	} est quicquid	{	possibile
injustum, illicitum			impossibile
æquum, debitum			necessarium
indebitum			omissibile

indifferens est simul iustum et omissibile



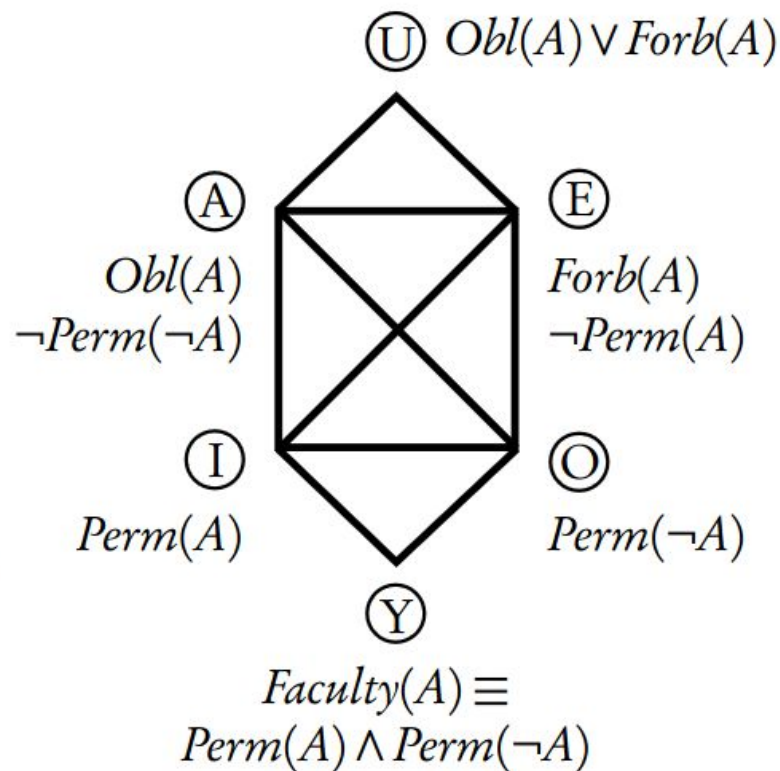
Leibniz: ***Calculus!*** ~1686

The beginning (?)

...building upon and further
strengthening the analytical tradition!

$\left. \begin{array}{l} \text{iustum, licitum} \\ \text{iniustum, illicitum} \\ \text{æquum, debitum} \\ \text{indebitum} \end{array} \right\}$	$\left. \begin{array}{l} \text{est quicquid} \\ \text{possibile} \\ \text{impossibile} \\ \text{necessarium} \\ \text{omissibile} \end{array} \right\}$
--	---

indifferens est simul iustum et omissibile



deontic logic, core element of theories of normative systems

A "fresh" a-new restart

L. Thorne McCarty

Reflections on TAXMAN: An Experiment in Artificial
Intelligence and Legal Reasoning*

Table of Contents

I. Introduction	A. States and Events
II. Some Fundamentals of Corporate Re- organization Law	B. Description Mechanisms
III. Some Fundamentals of Semantic in- formation Processing	C. Analysis Mechanisms
IV. Implementations	V. Results
	VI. Feasible Extensions
	VII. Conclusion

I. Introduction

This article addresses – in a highly unorthodox way – a series of questions which are surely among the most orthodox that can be found anywhere in the legal and jurisprudential literature. The questions are concerned primarily with legal concepts: What is the structure of a legal concept? What is the process by which legal concepts are transformed and modified? What are the arguments that lawyers characteristically make with respect to legal concepts? These are the kinds of questions that the Analytical Jurists once treated as essential, and the Realists dismissed as irrelevant; the kinds of questions that Cardozo ascribed to the Method of Philosophy in the judicial process, as opposed to the Method of Sociology¹. In short, I will be raising here some of the classical problems of legal reasoning, in their most classical form.

McCarty, **TAXMAN** – 1972-77

A “fresh” a-new restart

***Generic logic solvers
as Prolog ⇒***

L. Thorne McCarty

Reflections on TAXMAN: An Experiment in Artificial
Intelligence and Legal Reasoning*

Table of Contents

I. Introduction	A. States and Events
II. Some Fundamentals of Corporate Re- organization Law	B. Description Mechanisms
III. Some Fundamentals of Semantic In- formation Processing	C. Analysis Mechanisms
IV. Implementations	V. Results
	VI. Feasible Extensions
	VII. Conclusion

I. Introduction

This article addresses – in a highly unorthodox way – a series of questions which are surely among the most orthodox that can be found anywhere in the legal and jurisprudential literature. The questions are concerned primarily with legal concepts: What is the structure of a legal concept? What is the process by which legal concepts are transformed and modified? What are the arguments that lawyers characteristically make with respect to legal concepts? These are the kinds of questions that the Analytical Jurists once treated as essential, and the Realists dismissed as irrelevant; the kinds of questions that Cardozo ascribed to the Method of Philosophy in the judicial process, as opposed to the Method of Sociology¹. In short, I will be raising here some of the classical problems of legal reasoning, in their most classical form.

McCarty, TAXMAN – 1972-77

A “fresh” a-new restart

***Generic logic solvers
as Prolog ⇒***

L. Thorne McCarthy

Reflections on TAXMAN: An Experiment in Artificial
Intelligence and Legal Reasoning*

Table of Contents

I. Introduction
II. Some Fundamentals of Corporate Re-
organization Law
III. Some Fundamentals of Semantic In-
formation Processing
IV. Implementations

A. States and Events
B. Description Mechanisms
C. Analysis Mechanisms
V. Results
VI. Feasible Extensions
VII. Conclusion

I. Introduction

This article addresses – in a highly unorthodox way – a series of questions which are surely among the most orthodox that can be found anywhere in the legal and jurisprudential literature. The questions are concerned primarily with legal concepts: What is the structure of a legal concept? What is the process by which legal concepts are transformed and modified? What are the arguments that lawyers characteristically make with respect to legal concepts? These are the kinds of questions that the Analytical Jurists once treated as essential, and the Realists dismissed as irrelevant; the kinds of questions that Cardozo ascribed to the Method of Philosophy in the judicial process, as opposed to the Method of Sociology¹. In short, I will be raising here some of the classical problems of legal reasoning, in their most classical form.

McCarthy, TAXMAN – 1972-77

ARTICLES

THE BRITISH NATIONALITY ACT AS A LOGIC PROGRAM

The formalization of legislation and the development of computer systems to assist with legal problem solving provide a rich domain for developing and testing artificial-intelligence technology.

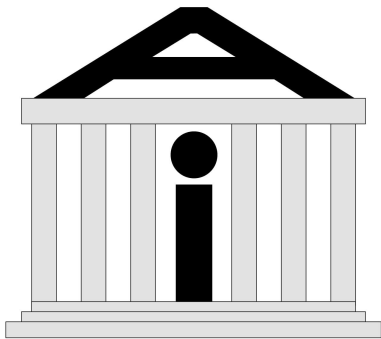
M. J. SERGOT, F. SADRI, R. A. KOWALSKI, F. KRIWACZEK, P. HAMMOND, and H. T. CORY

There are essentially two kinds of law, *case law*, determined by earlier court decisions, and *statutes*, determined by legislation. Substantial amounts of statutory law are basically definitional in nature and

the accuracy of the implementation and makes the implementation easier to maintain as the legislation changes and as case law evolves to augment the legislation.

Sergot, Sadri, Kowalski et al. – 1986

In the heyday of AI & Law: 80-90s



International Association for Artificial
Intelligence and Law (IAAIL)
since 1987 => ICAIL conference

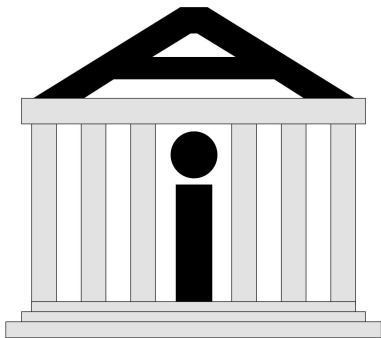


JURIX

The Foundation for Legal Knowledge Based Systems

Foundation for Legal Knowledge
Based Systems (JURIX)
since 1988 => JURIX conference

In the heyday of AI & Law: 80-90s



International Association for Artificial
Intelligence and Law (IAAIL)

since 1987 => ICAIL conference

ICAIL2026 @ Singapore

<https://site.smu.edu.sg/icail-2026>

submission deadline 16 January 2026,
plausibly postponed end of January



JURIX

The Foundation for Legal Knowledge Based Systems

Foundation for Legal Knowledge
Based Systems (JURIX)

since 1988 => JURIX conference

JURIX2026 @ Toulouse

call for papers in May/June 2026,
submission in September 2026
conference in December 2026

In the heyday of AI & Law: 80-90s



International Association for Artificial
Intelligence and Law (IAAIL)
since 1987 => ICAIL



JURIX

The Foundation for Legal Knowledge Based Systems

Foundation for Legal Knowledge
Based Systems (JURIX)
since 1988 => JURIX

UvA researchers had also a prominent role in this wave, resulting in the founding of the **Leibniz Center for Law** at the Law Faculty



Shallow-Knowledge vs Deep-Knowledge approaches

information
retrieval tools

natural languages
processing techniques

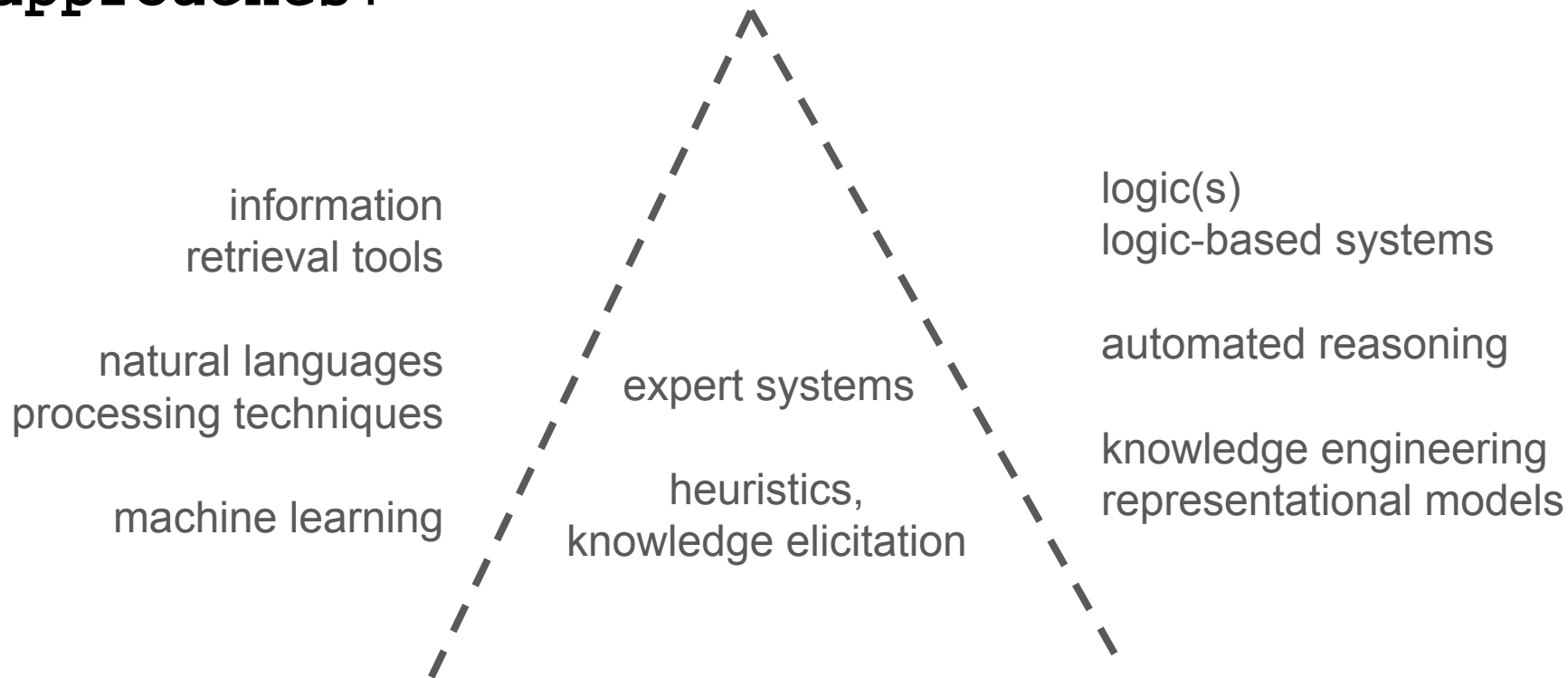
machine learning

logic(s)
logic-based systems

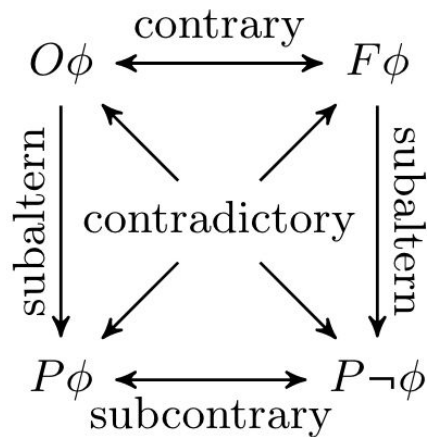
automated reasoning

knowledge engineering
representational models

Shallow-Knowledge vs Deep-Knowledge approaches+



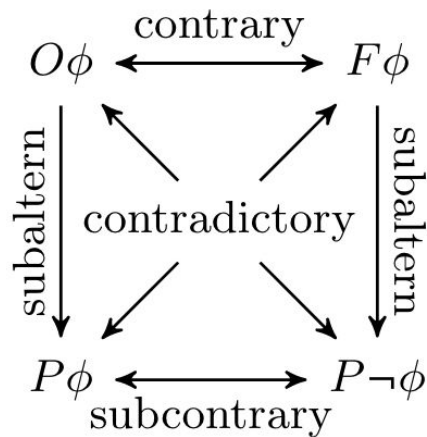
Deontic logics and Normative systems



Deontic logics

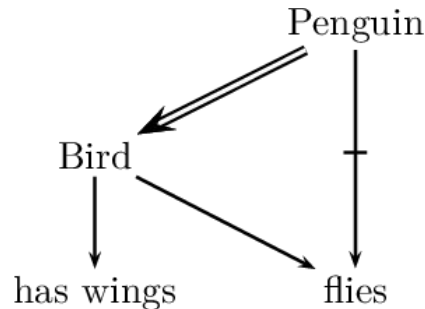
relevant concepts: **normative modalities, violations,**
...

Deontic logics and Normative systems



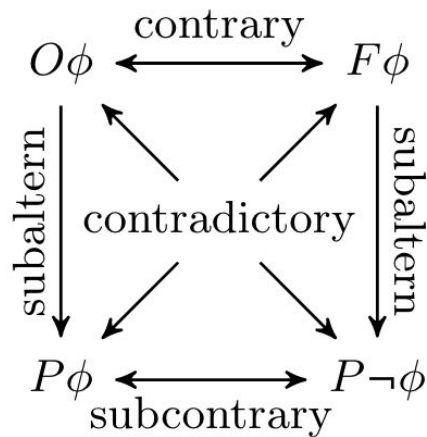
Deontic logics

relevant concepts: **normative modalities**, **violations**,
conflicts, **priorities**



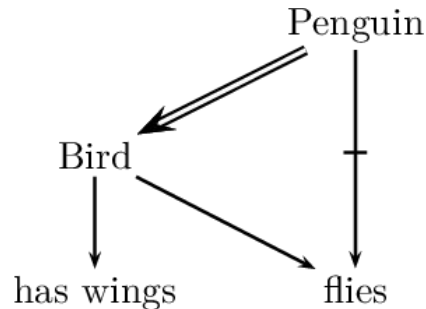
non-monotonic (defeasible) aspects

Deontic logics and Normative systems



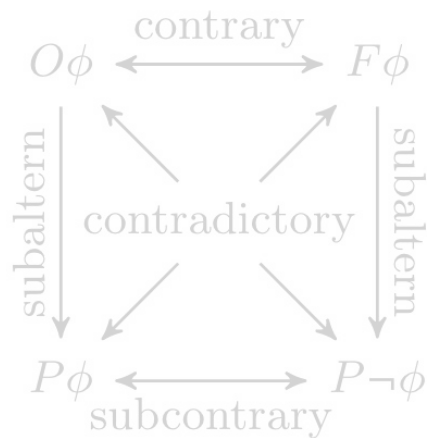
Deontic logics

relevant concepts: **normative modalities**,
violations, conflicts, priorities + **constitutive aspects?**

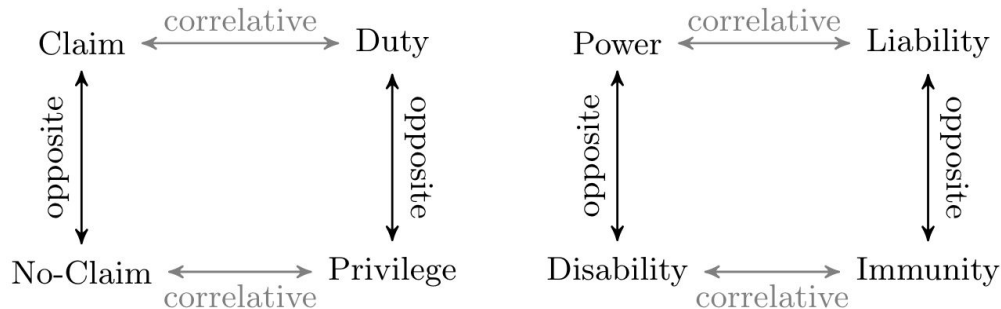


non-monotonic (defeasible) aspects

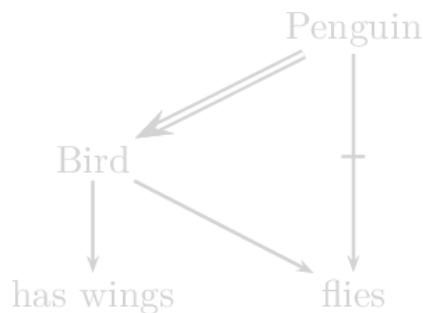
Deontic logics and Normative systems



Deontic logics



Hohfeld's framework of normative concepts



relevant concepts:
power, roles,
norm dynamics

non-monotonic (defeasible) aspects

Case-based reasoning

113

Artificial Intelligence and Law 1: 113–208, 1992.
© 1992 Kluwer Academic Publishers. Printed in the Netherlands.

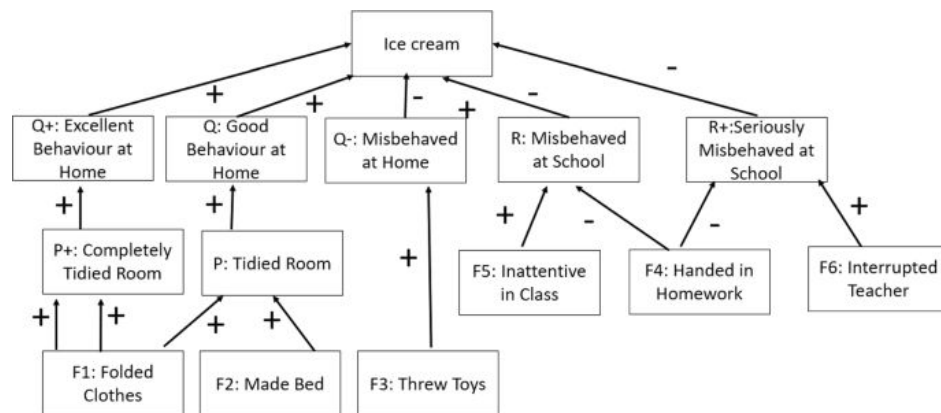
Case-Based Reasoning and its Implications for Legal Expert Systems

KEVIN D. ASHLEY

*Assistant Professor of Law and Intelligent Systems,
University of Pittsburgh,
School of Law
and
Learning Research and Development Center,
Pittsburgh, PA 15260, U.S.A.
e-mail: ashley@vms.cis.pitt.edu
tel: (412) 648-1495, 624-7496*

(Received 16 November 1991, accepted 8 October 1992)

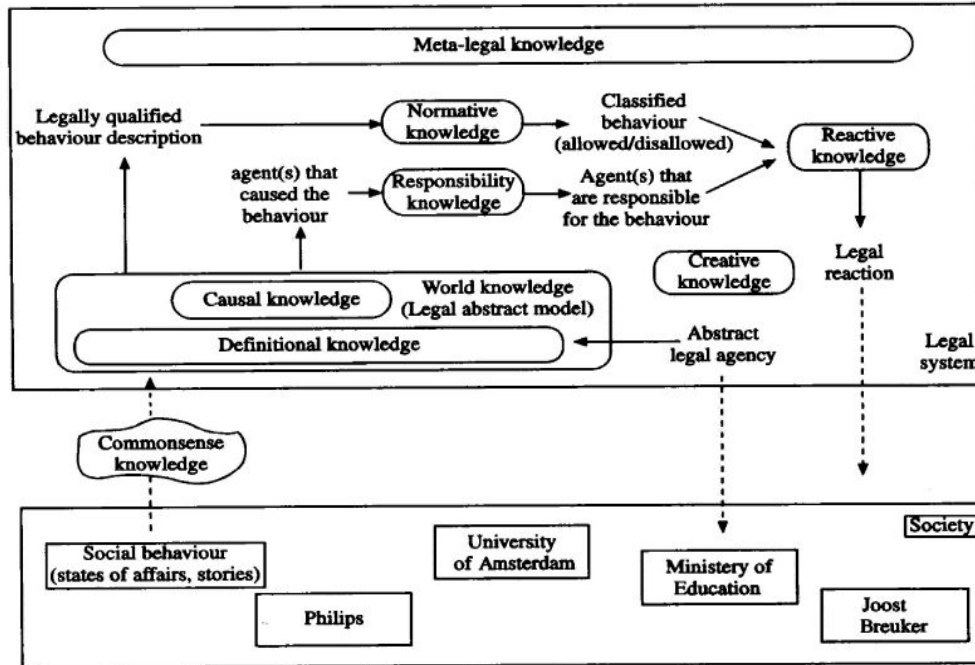
Abstract. Reasoners compare problems to prior cases to draw conclusions about a problem and guide decision making. All Case-Based Reasoning (CBR) employs some methods for generalizing from cases to support indexing and relevance assessment and evidences two basic inference methods: constraining search by tracing a solution from a past case or evaluating a case by comparing it to past cases. Across domains and tasks, however, humans reason with cases in subtly different ways evidencing different mixes of and mechanisms for these components.



relevant concepts: **factors** and **dimensions**

Ashley, Bench-Capon, Horty, ...

Expert Systems

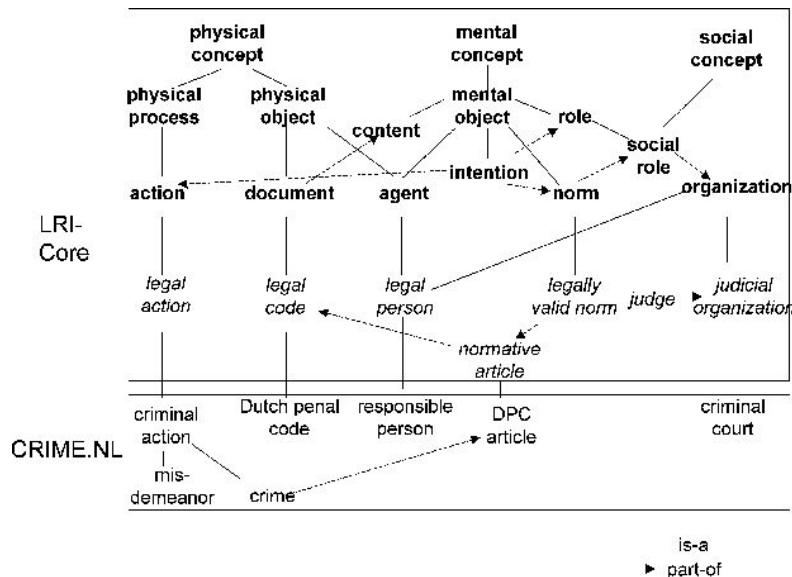


relevant aspects:

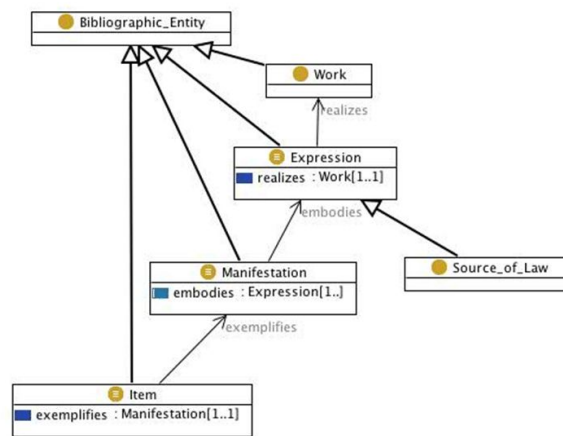
- knowledge domains, coupled with
- problem-solving methods and heuristics per type of task

Breuker, ...

Ontologies, and the Semantic Web



LRI-core (legal core ontology)
– applied eg. to improve search



CEN MetaLEX for legal documents

Breuker, Winkels, Van Engers, Hoekstra, Boer, ...
Palmirani, Casanovas, Francesconi...

Recommender Systems

Towards a Legal Recommender System

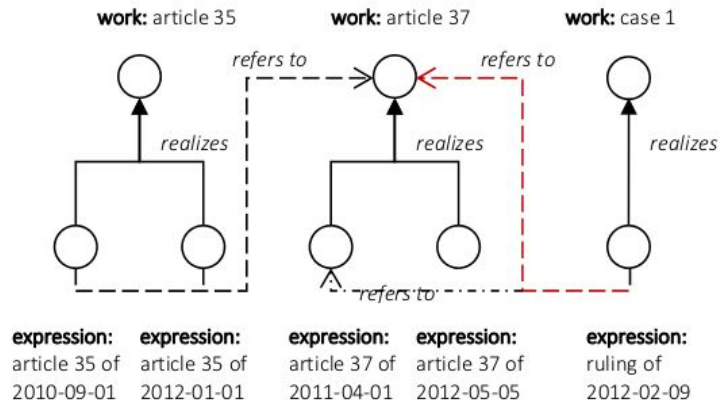
Radboud WINKELS¹, Alexander BOER,
Bart VREDEBREGT & Alexander van SOMEREN
Leibniz Center for Law, University of Amsterdam

Abstract. In this paper we present the results of ongoing research aimed at a legal recommender system where users of a legislative portal receive suggestions of other relevant sources of law, given a focus document. We describe how we make references in case law to legislation explicit and machine readable, and how we use this information to adapt the suggestions of other relevant sources of law. We also describe an experiment in categorizing the references in case law, both by human experts and unsupervised machine learning. Results are tested in a prototype for Immigration Law.

Keywords. Semantic Web, citations, MetaLex, network analysis, cluster analysis, reasons for citing

1. Introduction

More and more sources of law are freely available online in the Netherlands, but also in the rest of Europe and the world. Most of the time however, these are stand-alone databases, containing one type of documents, not linked to other sources. For instance the Dutch portal for case law – rechtspraak.nl – contains a (small) part of all judicial decisions in the Netherlands. Case citations in these decisions are sometimes explicitly linked, references to legislation are not.²



network analysis in law

Winkels, ...

Recommender Systems & NLP methods

Towards a Legal Recommender System

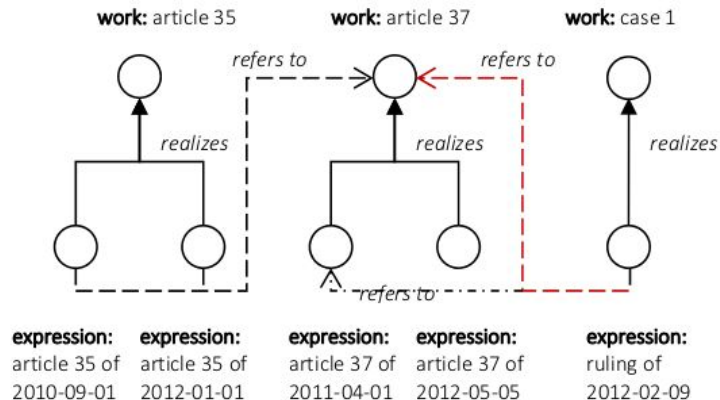
Radboud WINKELS¹, Alexander BOER,
Bart VREDEBREGT & Alexander van SOMEREN
Leibniz Center for Law, University of Amsterdam

Abstract. In this paper we present the results of ongoing research aimed at a legal recommender system where users of a legislative portal receive suggestions of other relevant sources of law, given a focus document. We describe how we make references in case law to legislation explicit and machine readable, and how we use this information to adapt the suggestions of other relevant sources of law. We also describe an experiment in categorizing the references in case law, both by human experts and unsupervised machine learning. Results are tested in a prototype for Immigration Law.

Keywords. Semantic Web, citations, MetaLex, network analysis, cluster analysis, reasons for citing

1. Introduction

More and more sources of law are freely available online in the Netherlands, but also in the rest of Europe and the world. Most of the time however, these are stand-alone databases, containing one type of documents, not linked to other sources. For instance the Dutch portal for case law – rechtspraak.nl – contains a (small) part of all judicial decisions in the Netherlands. Case citations in these decisions are sometimes explicitly linked, references to legislation are not.²



network analysis in law
topic modelling (possibly connecting to factors)
argument mining...

Winkels, Ashley, ...

Formal argumentation

Arg1 $\xrightarrow{\text{attacks}}$ Arg2

Dung's framework

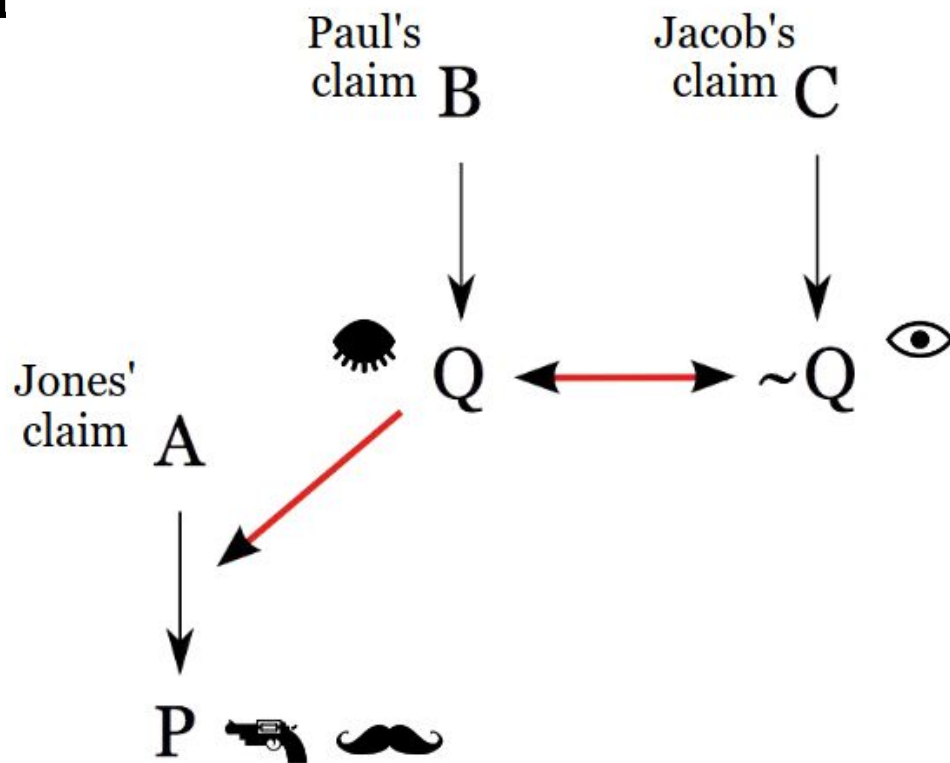
arguments and attack relationships

Formal argumentation

Arg1 $\xrightarrow{\text{attacks}}$ Arg2

Dung's framework

arguments and attack relationships

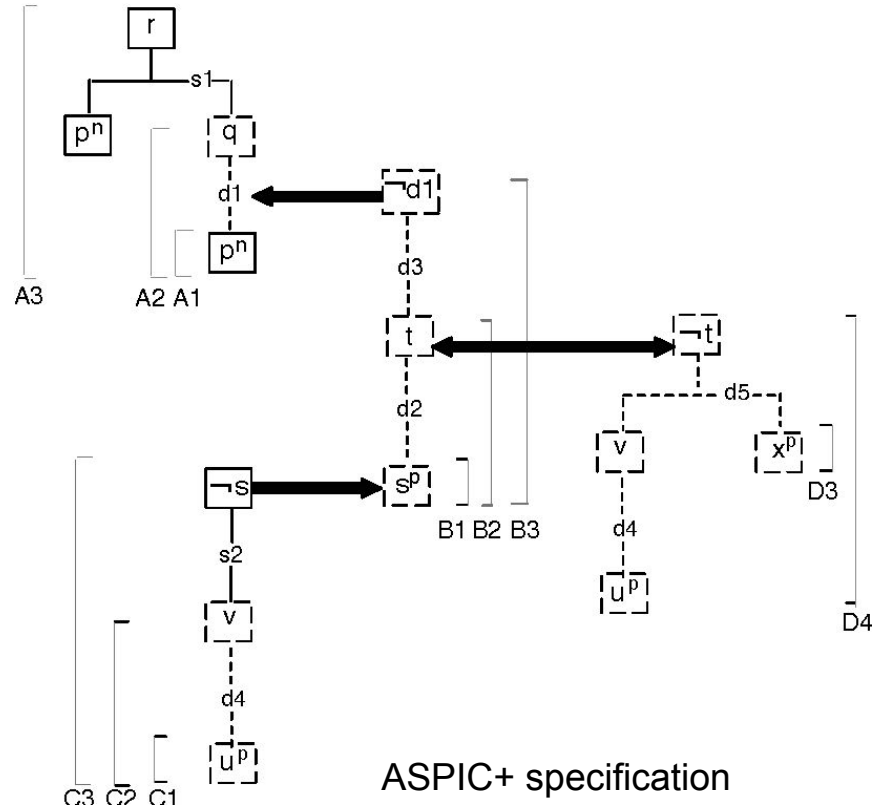


Formal argumentation

$\text{Arg1} \xrightarrow{\text{attacks}} \text{Arg2}$

Dung's framework

arguments and attack relationships



Now? Use of LLMs is exploding everywhere!

- The underlying assumption is essentially the same: reasoning as crunching numbers (although on latent spaces)...

Now? Use of LLMs is exploding everywhere!

- The underlying assumption is essentially the same: reasoning as crunching numbers (although on latent spaces)...
- A huge numbers of contributions concern LLM-centred pipelines:
 - for access to justice,
 - for human-right impact assessment (eg. using *personas*, *vignettes*),
 - for knowledge extraction
 - for text interpretation
 - for legal prediction...

Now? Use of LLMs is exploding everywhere!

- The underlying assumption is essentially the same: reasoning as crunching numbers (although on latent spaces)...
- A huge numbers of contributions concern LLM-centred pipelines:
 - for access to justice,
 - for human-right impact assessment (eg. using *personas*, *vignettes*),
 - for knowledge extraction
 - for text interpretation
 - for legal prediction...
- **Reasoning is still weak.** There is space for neuro-symbolic approaches: towards hybrid methods for **normative system specifications**, or **argumentation** or **case-based reasoning constructs**?

On the positive side...

Plenty of datasets available!

On the positive side...

Plenty of datasets available!

<https://github.com/openlegaldata/awesome-legal-data>

master ▾

1 Branch 0 Tags

Go to file

<> Code ▾

umarbutter feat: add MLEB (#7)

290fab3 · 2 months ago 24 Commits

README.md

feat: add MLEB (#7)

2 months ago

README

Awesome Legal Data

 awesome

A curated list of resources dedicated to legal data. The collection contains data sets, corpora, benchmarks, tools and other links related to the legal domain — grouped by region/country. Most resources are openly available.

Legend: (*Open*) = publicly accessible (may require free key); (*Commercial*) = paid/subscription; (*API*) = offers programmatic access/bulk data.

Global & Multi-Jurisdictional

- [WorldLII](#) — Federated gateway to free legal information worldwide (legislation, case law, treaties) across national LII's. (*Open*)
- [vLex](#) — Global legal research platform covering 100+ jurisdictions; case law, legislation, journals, and analytics. (*Commercial*)

About

A collection of datasets and other resources for legal text processing.

[openlegaldata.io](#)

nlp legal datasets legal-tech

Readme

Activity

Custom properties

155 stars

7 watching

24 forks

Report repository

Contributors 5



On the positive side...

Plenty of datasets available!

eg. cases addressed European Convention on Human Rights (ECHR), annotated by humans and LLMs, counting for the number and variety of citations (the citation context)....



Figure 3. Citation Analysis Tool incorporating rich citation context, including across repeated citations of the same authority by the same case.

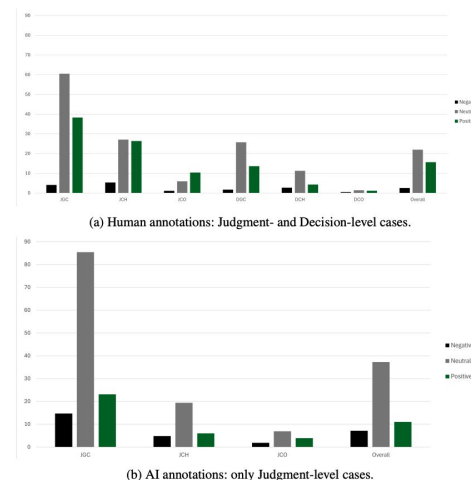


Figure 1. Distributions of authority-citations per case by judicial consideration label. Results are segmented by judicial output levels, where 'Overall' denotes the mean across all levels.

From JURIX 2025: Mumford, Jack, et al. "Context-Aware Citation Networks: A Human–AI Dataset, Analysis, and Tool." *Frontiers in Artificial Intelligence and Applications* (2025). <https://ebooks.iospress.nl/volumearticle/76891>

**Some (necessarily?)
open challenges**

Quantitative vs qualitative

- Economic vs Legal reasoning



Quantitative vs qualitative

- Economic vs Legal reasoning
- Axiological vs Deontological (in Ethics)
- Optimization vs Satisfiability (in Computer Science)
- Soft vs strong requirements (in Operational research)
- Principles (Balancing) vs Rules (in Law)
- Voting (Counting) vs Argumentation (in Decision-making)
- Metrics vs Topology (in Mathematics)
- Functional vs Script-based agents (in Agent-based Modelling)
- ...

Quantitative vs qualitative

- Economic vs Legal reasoning
- Axiological vs Deontological (in Ethics)
- Optimization vs Satisfiability (in Computer Science)
- Soft vs strong requirements (in Operational research)
- Principles (Balancing) vs Rules (in Law)
- Voting (Counting) vs Argumentation (in Decision-making)
- Metrics vs Topology (in Mathematics)
- Functional vs Script-based agents (in Agent-based Modelling)
- ...



These dimensions are dynamically entrenched, we cannot easily “collapse” one on the other without problems..

On probabilistic judgments

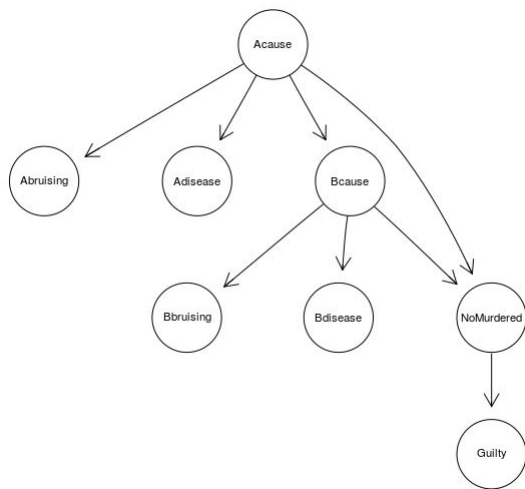
A dramatic, and unfortunately true story

“Starting in 1996, four British mothers were arrested for murder because two or more of their infants had died in their cribs. Three of the mothers were convicted and sentenced to life imprisonment. Yet by 2006, all four mothers were declared innocent of crimes that had never occurred. Bayes’ rule could have cleared them.

Sally Clark, a lawyer who lost her first son at 11 weeks and her second at 8 weeks, was convicted in 1999. A prominent pediatrician, Sir Roy Meadow, had testified for the prosecution about Sudden Infant Death Syndrome, known as SIDS in the U.S. and cot death in Britain. Citing a government study, Meadow said the incidence of one SIDS death was one in 8,500 in a family like Clark’s—stable, affluent, nonsmoking, with a mother more than 26 years old.

Then, despite the fact that some families are predisposed to SIDS, Meadow assumed erroneously that each sibling’s death occurred independently of the other. Multiplying 8,500 by 8,500, he calculated that the chance of two children dying in a family like Sally Clark’s was so rare—one in 73 million—that they must have been murdered. [...]”

A dramatic, and unfortunately true story



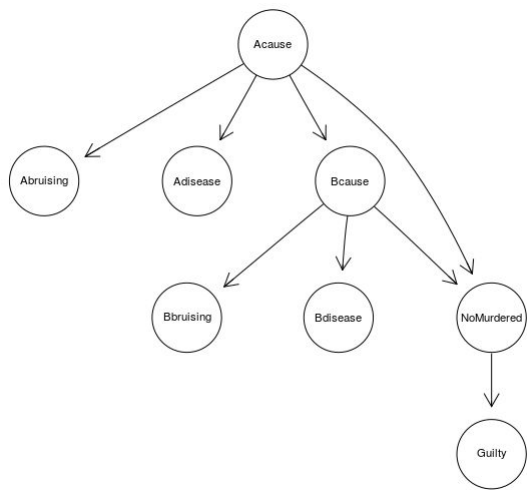
***Can we judge individuals
on the basis of frequencies
observable in a social
group?***

“Starting in 1996, four British mothers were arrested for murder because two or more of their infants had died in their cribs. Three of the mothers were convicted and sentenced to life imprisonment. Yet by 2006, all four mothers were declared innocent of crimes that had never occurred. Bayes’ rule could have cleared them.

Sally Clark, a lawyer who lost her first son at 11 weeks and her second at 8 weeks, was convicted in 1999. A prominent pediatrician, Sir Roy Meadow, had testified for the prosecution about Sudden Infant Death Syndrome, known as SIDS in the U.S. and cot death in Britain. Citing a government study, Meadow said the incidence of one SIDS death was one in 8,500 in a family like Clark’s—stable, affluent, nonsmoking, with a mother more than 26 years old.

Then, despite the fact that some families are predisposed to SIDS, Meadow assumed erroneously that each sibling’s death occurred independently of the other. Multiplying 8,500 by 8,500, he calculated that the chance of two children dying in a family like Sally Clark’s was so rare—one in 73 million—that they must have been murdered. [...]”

Can we apply bayesian inference in court?



***Can we judge individuals
on the basis of frequencies
observable in a social
group?***



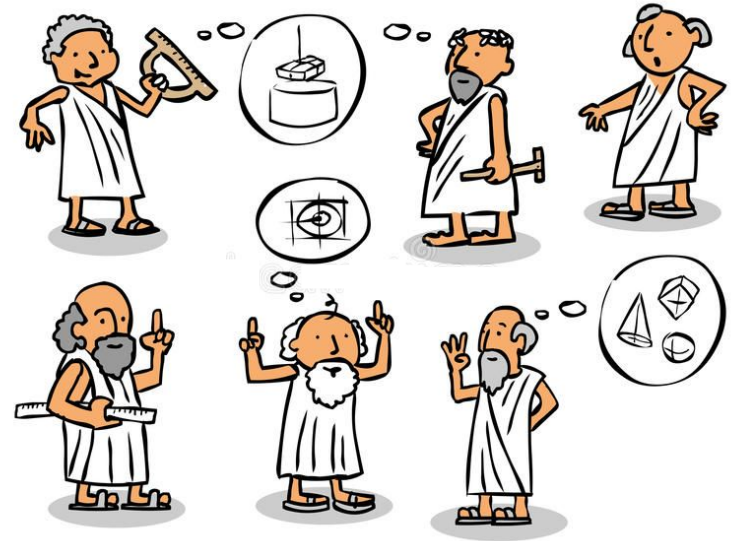
Judges usually argue that eg. people cannot be deemed to be in a certain place with a probability of 0.6 or 0.7.

Either they were there or they were not there.

On the difficulty of
normative modelling

What terms like duty or right mean?

- long-standing debate
- no shared agreement
- new semantics continuously released



What terms like duty or right mean?

- long-standing debate
- no shared agreement
- new semantics continuously released

And this holds even without considering ***open-textured*** terms (*bona fide*, reasonable person, etc.)!



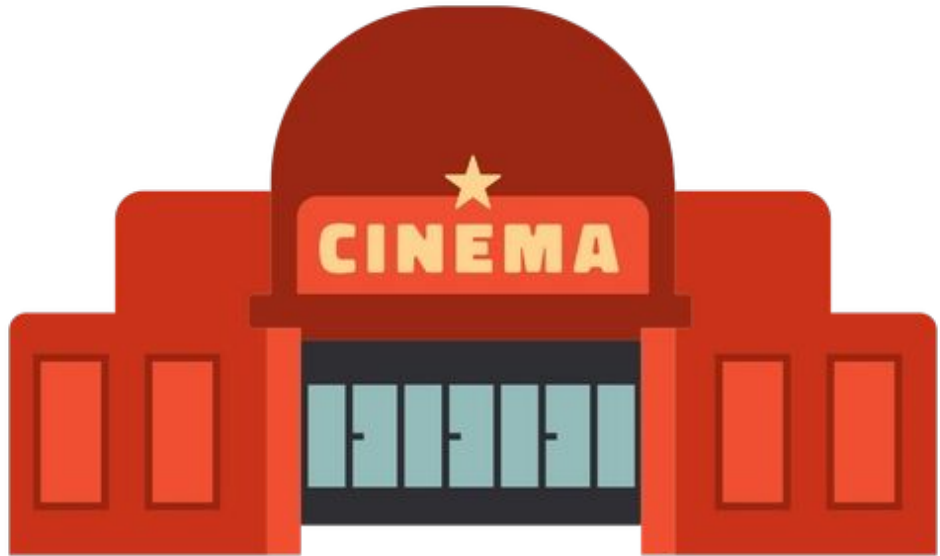
Example 1

- You are permitted to smoke.



Example 2

- You have to pay to see the film.



Example 3

- You are forbidden to use trolleys.



Example 3

- You are forbidden to use trolleys.



What terms like duty or right mean?

- long-standing debate
- no shared agreement
- new semantics continuously released

And this holds even without considering ***open-textured*** terms (*bona fide*, reasonable person, etc.)!



Text is not enough: you require extra-linguistic, contextual knowledge to form the right interpretation!

On case predictions

Judgment “prediction”?

- An increasing number of works state they reach a good accuracy in predicting court outcomes.



JURI SAYS:

I'm predicting judgments of the European Court of Human Rights with an accuracy of **87.2% over the *last month*.**

JURI reads published documents from previous years and decisions of the cases judged by the European Court of Human Rights and predicts decisions the Court will make. Every month it learns from its mistakes.



<https://jurisays.com/>

PLOS ONE

OPEN ACCESS PEER-REVIEWED
RESEARCH ARTICLE

A general approach for predicting the behavior of the Supreme Court of the United States

Daniel Martin Katz, Michael J. Bommarito II, Josh Blackman

Published: April 12, 2017 • <https://doi.org/10.1371/journal.pone.0174698>

Article	Authors	Metrics	Comments	Media Coverage
✓				

Abstract

Introduction
Research principles and prior work
Data and feature engineering
Model construction
Model testing and results
Conclusion and future research
Acknowledgments
Author Contributions

Abstract

Building on developments in machine learning and prior work in the science of judicial prediction, we construct a model designed to predict the behavior of the Supreme Court of the United States in a generalized, out-of-sample context. To do so, we develop a time-evolving random forest classifier that leverages unique feature engineering to predict more than 240,000 justice votes and 28,000 cases outcomes over nearly two centuries (1816-2015). Using only data available prior to decision, our model outperforms null (baseline) models at both the justice and case level under both parametric and non-parametric tests. Over nearly two centuries, we achieve 70.2% accuracy at the case outcome level and 71.9% at the justice vote level. More recently, over the past century, we outperform an *in-sample optimized null* model by nearly 5%. Our performance is consistent with, and improves on the general level of prediction demonstrated by prior work; however, our model is distinctive because it can be applied out-of-sample to the entire past and future of the Court, not a single term. Our results represent an important advance for the science of quantitative legal prediction and portend a range of other potential applications.

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0174698>

Judgment “prediction”?

- An increasing number of works state they reach a good accuracy in predicting court outcomes.
- Yet, besides the clear ethical challenge, often they have serious **methodological issues!**

<https://aclanthology.org/2023.nllp-1.9.pdf>

Legal Judgment Prediction: If You Are Going to Do It, Do It Right

Masha Medvedeva
eLaw Center for Law and Digital Technologies
& Department of Business Studies,
Faculty of Law,
Leiden University, the Netherlands
m.medvedeva@law.leidenuniv.nl

Pauline McBride
COHUBICOL,
Law, Science, Technology
and Society (LSTS),
Vrije Universiteit Brussel, Belgium
pauline.mcbride@vub.be

Abstract

The field of Legal Judgment Prediction (LJP) has witnessed significant growth in the past decade, with over 100 papers published in the past three years alone. Our comprehensive survey of over 150 papers reveals a stark reality: only ~7% of published papers are doing what they set out to do - predict court decisions. We delve into the reasons behind the flawed and unreliable nature of the remaining experiments, emphasising their limited utility in the legal domain. We examine the distinctions between predicting court decisions and the practices of legal professionals in their daily work. We explore how a lack of attention to the identity and needs of end-users has fostered the misconception that LJP is a near-solved challenge suitable

The media has speculated about the advent of robojudges and automated lawyers for years (Griffin, 2016; Markou, 2020; Kelly, 2022). Many in the justice sector are persuaded that machine learning technologies might usefully be employed to predict the outcomes of cases (CEPEJ, 2018). Some judges believe these technologies could even make “very minor” decisions.¹

Yet, these systems are not widely used. While there are a number of ethical considerations (Leins et al., 2020) including issues of bias in the data (Angwin et al., 2016) that could be the reason for the hesitancy, in this paper we will demonstrate that there may be a different root cause, namely that the setup of the experiments presented in the

Judgment “prediction”?

- An increasing number of works state they reach a good accuracy in predicting court outcomes.
- Yet, besides the clear ethical challenge, often they have serious **methodological issues!**
- Predictions are mostly based on **summaries, not on the brute facts given to the court!**

<https://aclanthology.org/2023.nllp-1.9.pdf>

Legal Judgment Prediction: If You Are Going to Do It, Do It Right

Masha Medvedeva
eLaw Center for Law and Digital Technologies
& Department of Business Studies,
Faculty of Law,
Leiden University, the Netherlands
m.medvedeva@law.leidenuniv.nl

Pauline McBride
COHUBICOL,
Law, Science, Technology
and Society (LSTS),
Vrije Universiteit Brussel, Belgium
pauline.mcbride@vub.be

Abstract

The field of Legal Judgment Prediction (LJP) has witnessed significant growth in the past decade, with over 100 papers published in the past three years alone. Our comprehensive survey of over 150 papers reveals a stark reality: only ~7% of published papers are doing what they set out to do - predict court decisions. We delve into the reasons behind the flawed and unreliable nature of the remaining experiments, emphasising their limited utility in the legal domain. We examine the distinctions between predicting court decisions and the practices of legal professionals in their daily work. We explore how a lack of attention to the identity and needs of end-users has fostered the misconception that LJP is a near-solved challenge suitable

The media has speculated about the advent of robojudges and automated lawyers for years (Griffin, 2016; Markou, 2020; Kelly, 2022). Many in the justice sector are persuaded that machine learning technologies might usefully be employed to predict the outcomes of cases (CEPEJ, 2018). Some judges believe these technologies could even make “very minor” decisions.¹

Yet, these systems are not widely used. While there are a number of ethical considerations (Leins et al., 2020) including issues of bias in the data (Angwin et al., 2016) that could be the reason for the hesitancy, in this paper we will demonstrate that there may be a different root cause, namely that the setup of the experiments presented in the

On algorithmic fairness

A super-recent example

- Studying 1,200 bail cases from Indian High Courts and the Supreme Court (1975–2025), researchers found that different ML algorithms behave differently both with respect to **performance** (accuracy) and **fairness** (always measured in terms of accuracy)...

Model	Acc.	Gender	Crime	Region
LegalBERT (Base)	93.9 ± 1.6	2.3 ± 0.9	22.8 ± 4.0	34.7 ± 18.1
LegalBERT (Red.)	82.2 ± 4.1	5.8 ± 5.2	33.6 ± 2.9	49.0 ± 15.7
Logistic Reg.	93.1 ± 0.9	3.9 ± 3.5	15.0 ± 2.1	34.7 ± 17.1
XGBoost	95.6 ± 0.5	$1.8 \pm 1.2^\dagger$	20.0 ± 5.6	25.0 ± 7.9

A super-recent example

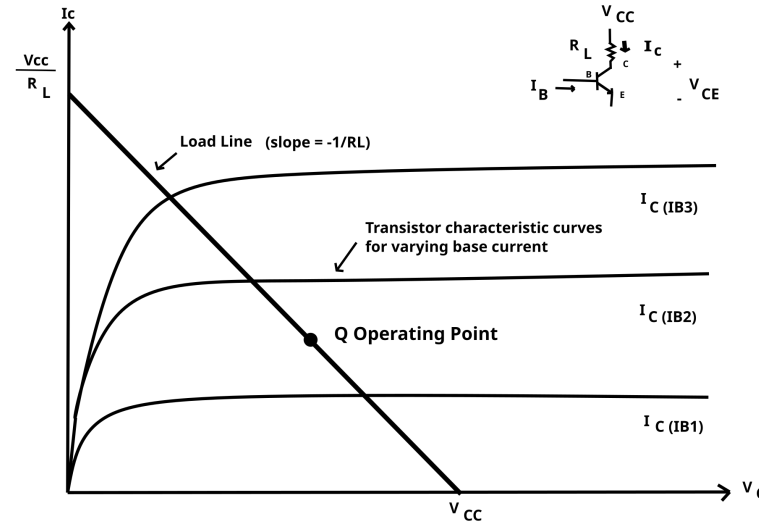
- Studying 1,200 bail cases from Indian High Courts and the Supreme Court (1975–2025), researchers found that different ML algorithms behave differently both with respect to **performance** (accuracy) and **fairness** (always measured in terms of accuracy)...

Model	Acc.	Gender	Crime	Region
LegalBERT (Base)	93.9 ± 1.6	2.3 ± 0.9	22.8 ± 4.0	34.7 ± 18.1
LegalBERT (Red.)	82.2 ± 4.1	5.8 ± 5.2	33.6 ± 2.9	49.0 ± 15.7
Logistic Reg.	93.1 ± 0.9	3.9 ± 3.5	15.0 ± 2.1	34.7 ± 17.1
XGBoost	95.6 ± 0.5	$1.8 \pm 1.2^\dagger$	20.0 ± 5.6	25.0 ± 7.9

**Yet, many things
are still unclear here!**

Bias

- As an electronic engineer, I was trained to see the term bias from a technical point of view: bias as variable (eg. current/tension) or **offset** that changes the functioning of a device.



(Social) Bias and Fairness

- More commonly, bias is used to denote social discrimination:
preferring someone over someone else.
- typically to cases of **negative discrimination** (disfavouring/punishing). but in principle it can be applied for **positive discrimination** (favouring/assisting).

(Social) Bias and Fairness

- More commonly, bias is used to denote social discrimination:
preferring someone over someone else.
- typically to cases of **negative discrimination** (disfavouring/punishing). but in principle it can be applied for **positive discrimination** (favouring/assisting).



affirmative action (“active effort to improve employment, educational, and other opportunities for members of groups that have been subjected to discrimination”) is positive discrimination, and by definition requires discrimination (= *differentiation*)

Bias and Algorithmic Fairness

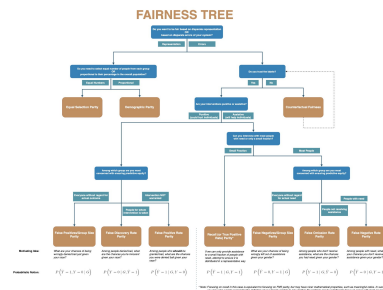
- Particularly in data-centred contexts, bias is measured as **statistical parity**
- Studies in algorithmic fairness have introduced several statistical measures (**demographic parity, equalized odds, ...**), in most cases *theoretically incompatible* with each other.

Bias and Algorithmic Fairness

- Particularly in data-centred contexts, bias is measured as **statistical parity**
- Studies in algorithmic fairness have introduced several statistical measures (**demographic parity, equalized odds, ...**), in most cases *theoretically incompatible* with each other.
- *Which one to choose?*
 - It depends on the specific decision context – eg. how much impact has a false negative error compared to a false positive error.

see eg. the AEQUITAS fairness tree:

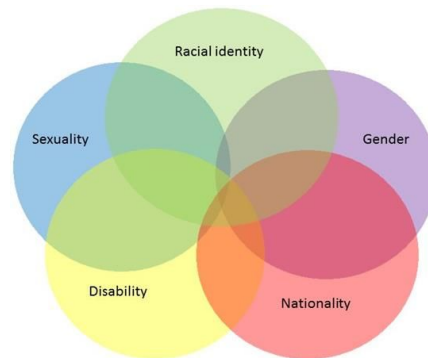
datasciencepublicpolicy.org/our-work/tools-guides/aequitas/



Parity between whom?

- statistical measures relies on identifying groups
- but which groups? generally protected groups groups defined by sensitive dimensions (eg. sex or gender, race or ethnicity)

➔ problem of *intersectionality*



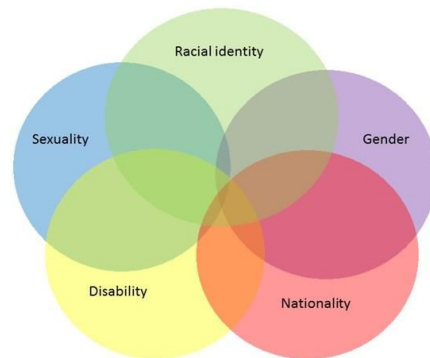
Parity between whom?

- statistical measures relies on identifying groups
- but which groups? generally protected groups groups defined by sensitive dimensions (eg. sex or gender, race or ethnicity)

 problem of *intersectionality*

Example:

- women vs men
- black women vs black men
- low-class black women vs low-class black men



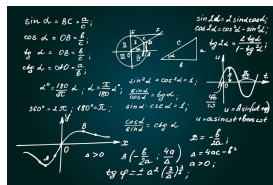
Which is the right granularity? (problem with affinities to Simpson's paradox!)

Example of Simpson's paradox

- The Simpson's paradox is an unintuitive property consequent to counting.



university



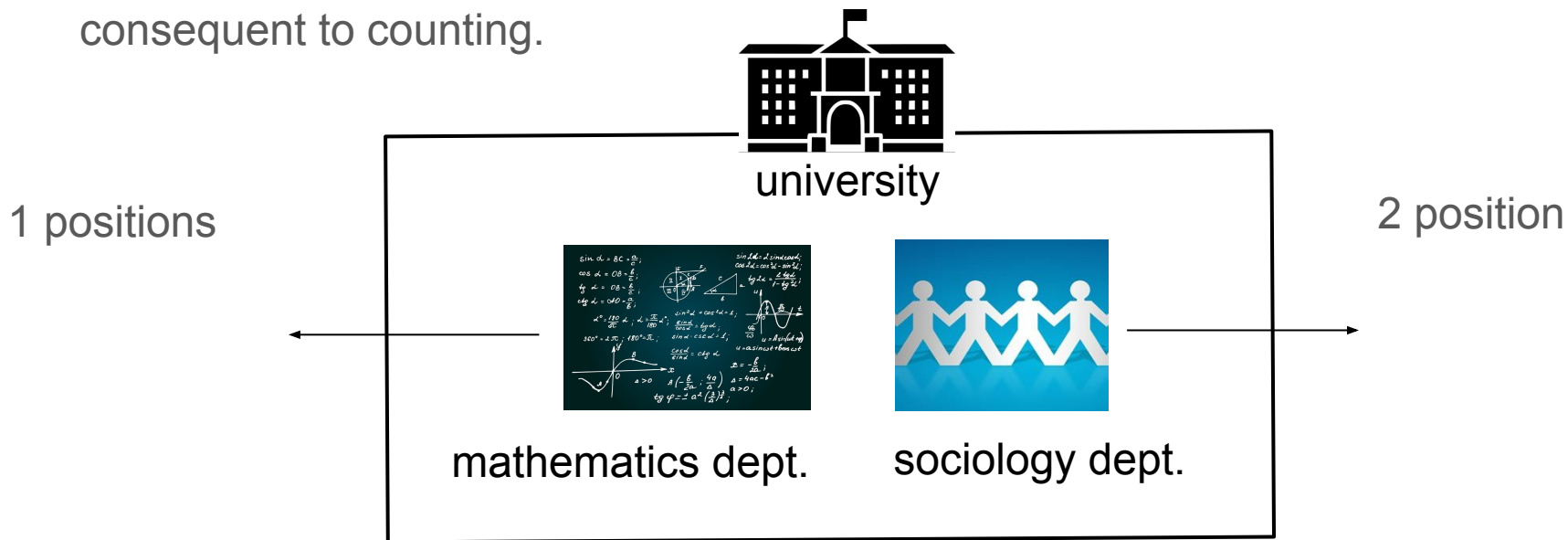
mathematics dept.



sociology dept.

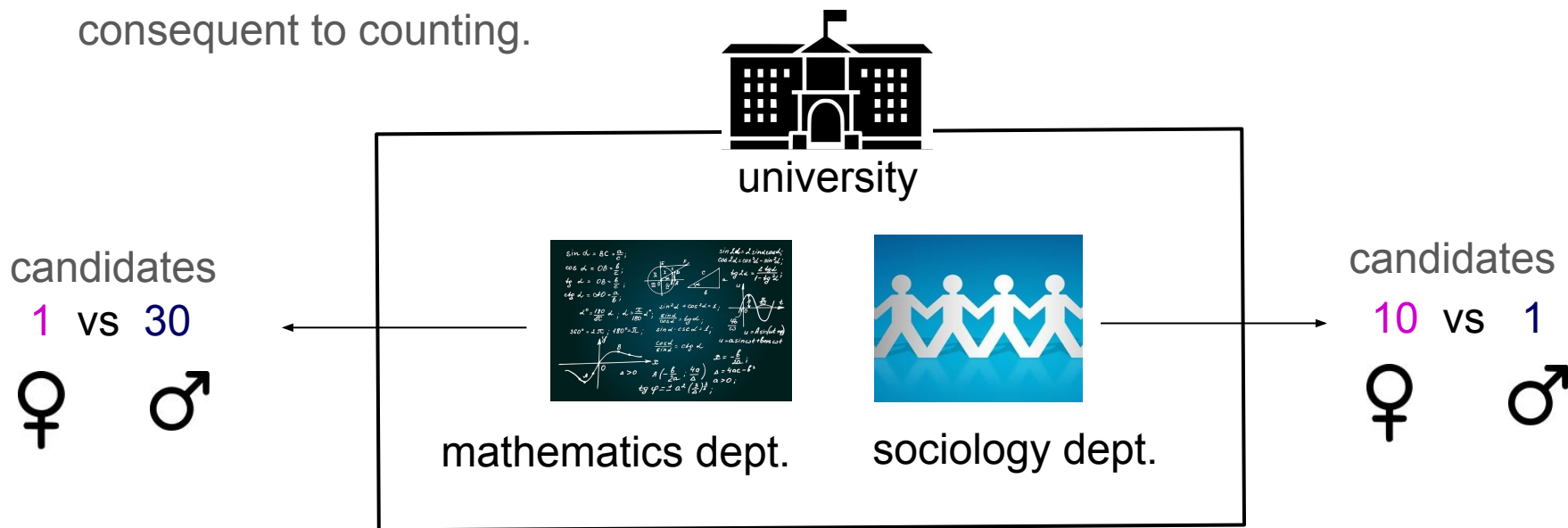
Example of Simpson's paradox

- The Simpson's paradox is an unintuitive property consequent to counting.



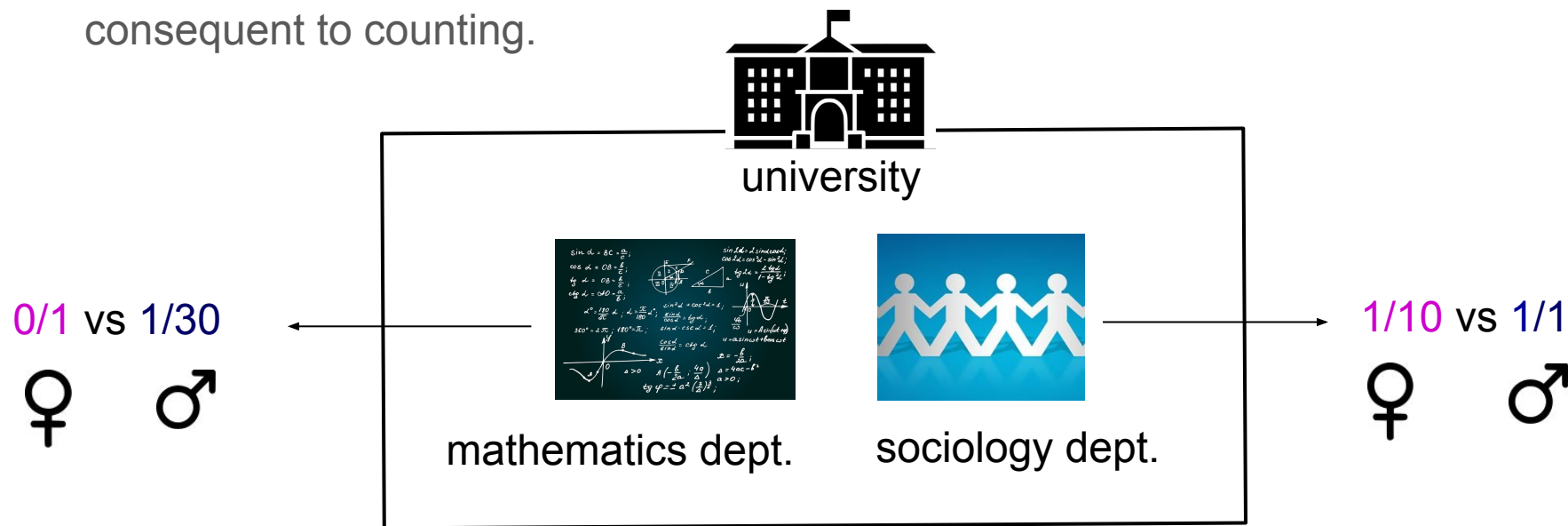
Example of Simpson's paradox

- The Simpson's paradox is an unintuitive property consequent to counting.



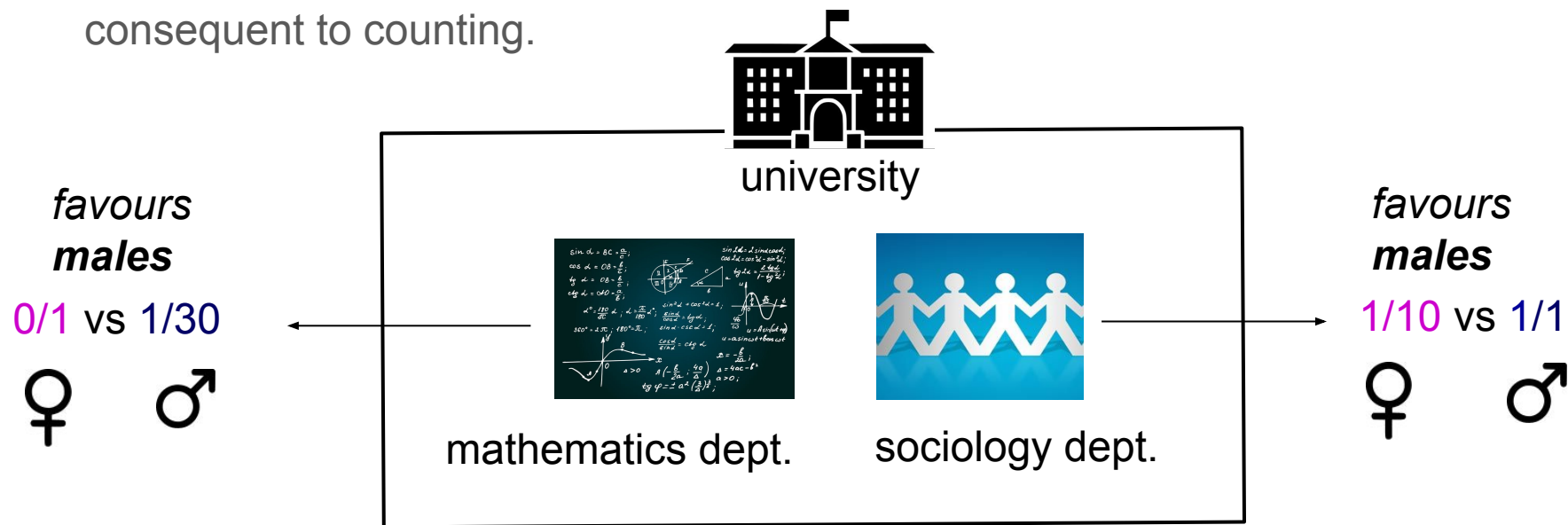
Example of Simpson's paradox

- The Simpson's paradox is an unintuitive property consequent to counting.



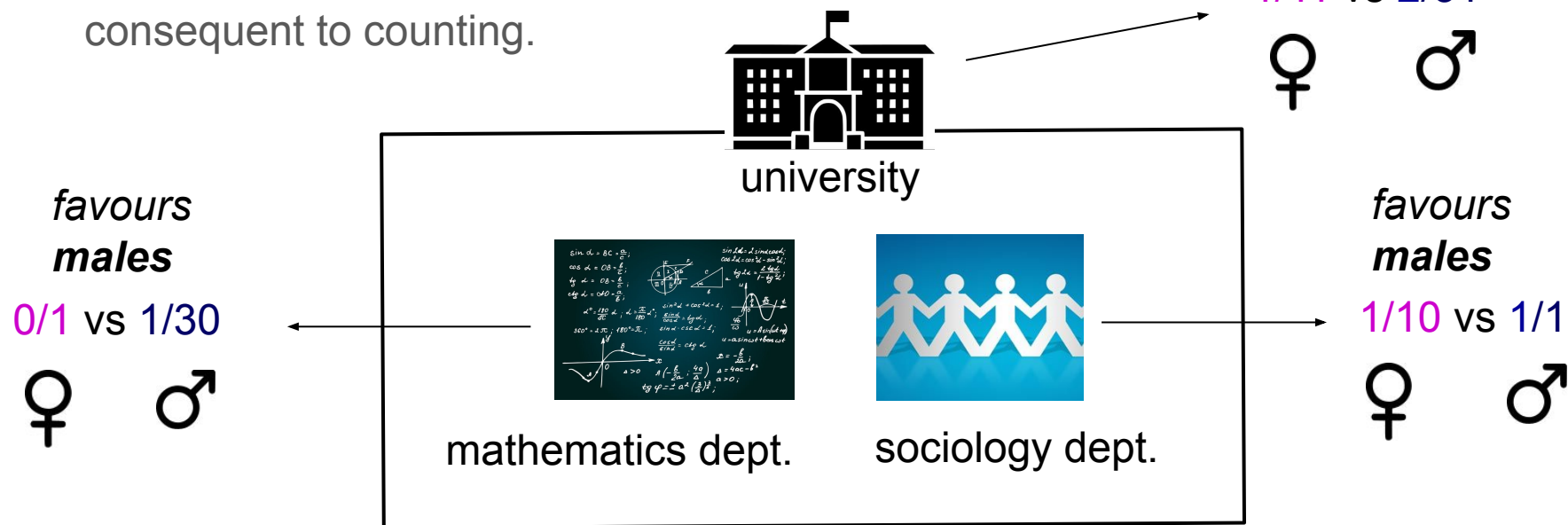
Example of Simpson's paradox

- The Simpson's paradox is an unintuitive property consequent to counting.



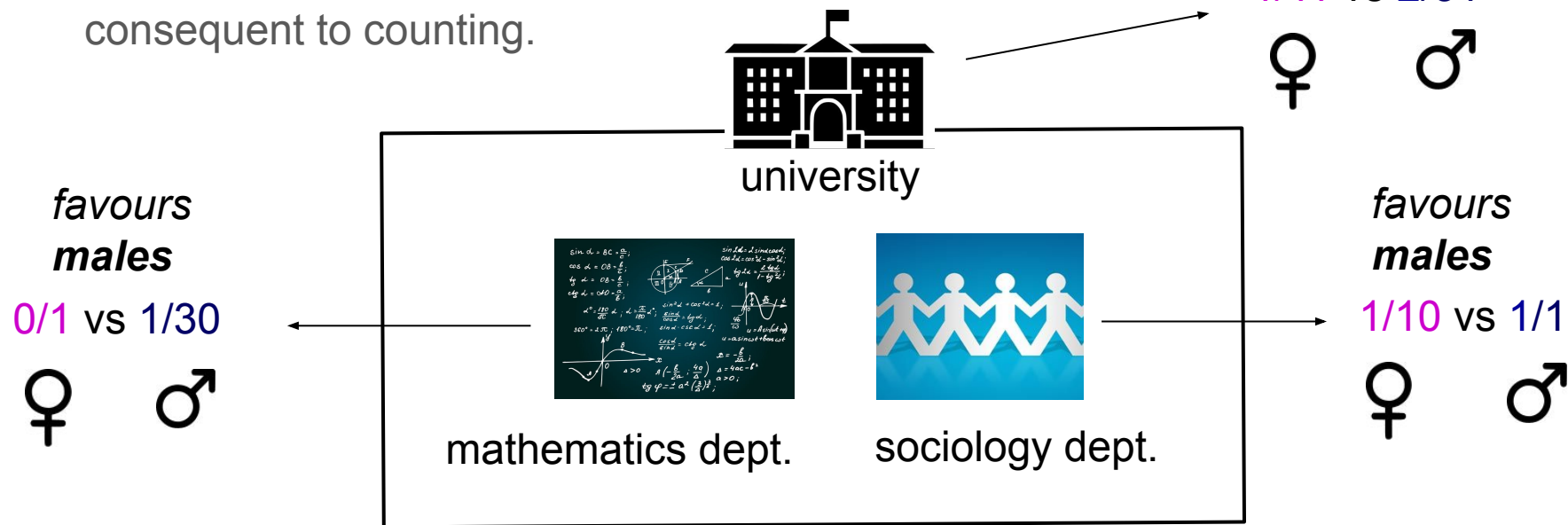
Example of Simpson's paradox

- The Simpson's paradox is an unintuitive property consequent to counting.



Example of Simpson's paradox

- The Simpson's paradox is an unintuitive property consequent to counting.

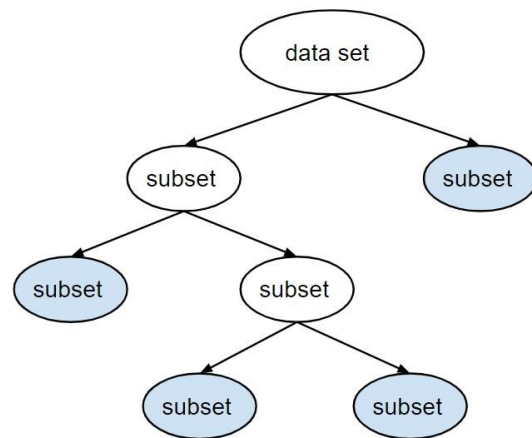


By changing the level of aggregation (eg. university vs dept.) you can reach opposite conclusions about the same phenomenon!

Parity between whom?

With a Master IS student we worked on inverting the problem:

- identifying descriptively relevant subgroups in the data space
- check how the algorithm works on these in general, and w.r.t. protected features



Some results on German credit dataset

– Note: here ML is used to unveil
historical (human) bias!

demographic disparity
against women
with low resources
when asking credit
for cars

(slight)
demographic disparity
against women
with low resources
when asking credit
for furniture/equipment

Profile	Age	Sex	Job	Housing	Saving accounts	Checking account	Credit amount	Duration	Purpose	Sample
0	26	male	2	rent	moderate	unknown	3577	9	car	859
1	37	male	2	own	unknown	unknown	7409	36	business	868
2	39	male	3	own	little	unknown	6458	18	car	106
3	26	male	2	own	little	little	4370	42	radio/TV	639
4	31	male	2	own	quite rich	unknown	3430	24	radio/TV	19
5	38	female	2	own	unknown	unknown	1240	12	radio/TV	135
6	43	male	1	own	little	little	1344	12	car	929
7	36	male	2	rent	little	little	2799	9	car	586
8	39	male	2	own	little	little	2522	30	radio/TV	239
9	31	male	2	own	little	moderate	1935	24	business	169
10	33	female	2	own	little	little	1131	18	furniture/equipment	166
11	26	male	1	own	little	moderate	625	12	radio/TV	220
12	23	male	2	own	unknown	moderate	1444	15	radio/TV	632
13	42	male	2	own	little	little	4153	18	furniture/equipment	899
14	29	male	2	own	unknown	unknown	3556	15	car	962
15	37	female	2	own	little	moderate	3612	18	furniture/equipment	537
16	27	female	2	own	little	little	2389	18	radio/TV	866
17	26	female	2	rent	little	unknown	1388	9	furniture/equipment	582
18	29	male	2	own	little	unknown	2743	28	radio/TV	426
19	53	male	2	free	little	little	4870	24	car	4
20	36	male	2	own	little	little	1721	15	car	461
21	38	male	2	own	little	unknown	804	12	radio/TV	997
22	29	male	2	own	little	moderate	1103	12	radio/TV	696
23	43	male	2	own	unknown	unknown	2197	24	car	406
24	27	male	2	own	little	little	3552	24	furniture/equipment	558
25	30	male	2	own	little	moderate	1056	18	car	580
26	24	female	2	own	little	moderate	2150	30	car	252
27	34	male	2	own	little	unknown	2750	12	furniture/equipment	452
28	24	female	2	rent	little	little	2124	18	furniture/equipment	761
29	34	male	2	own	little	moderate	5800	30	car	893
30	34	female	2	own	little	unknown	1493	12	radio/TV	638
31	30	female	2	own	little	unknown	1055	18	car	161
32	35	male	2	own	little	unknown	2346	24	car	654
33	35	male	2	own	unknown	unknown	1979	15	radio/TV	625

Some results on German credit dataset

– Note: here ML is used to unveil
historical (human) bias!

demographic disparity
against women
with low resources
when asking credit
for cars

(slight)
demographic disparity
against women
with low resources
when asking credit

Profile	Age	Sex	Job	Housing	Saving accounts	Checking account	Credit amount	Duration	Purpose	Sample
0	26	male	2	rent	moderate	unknown	3577	9	car	859
1	37	male	2	own	unknown	unknown	7409	36	business	868
2	39	male	3	own	little	unknown	6458	18	car	106
3	26	male	2	own	little	little	4370	42	radio/TV	639
4	31	male	2	own	quite rich	unknown	3430	24	radio/TV	19
5	38	female	2	own	unknown	unknown	1240	12	radio/TV	135
6	43	male	1	own	little	little	1344	12	car	929
7	36	male	2	rent	little	little	2799	9	car	586
8	39	male	2	own	little	little	2922	30	radio/TV	239
9	31	male	2	own	little	moderate	1935	24	business	169
10	33	female	2	own	little	little	1131	18	furniture/equipment	166
11	26	male	1	own	little	moderate	625	12	radio/TV	220
12	23	male	2	own	unknown	moderate	1444	15	radio/TV	632
13	42	male	2	own	little	little	4153	18	furniture/equipment	899
14	29	male	2	own	unknown	unknown	3556	15	car	962
15	37	female	2	own	little	moderate	3612	18	furniture/equipment	537
16	27	female	2	own	little	little	2389	18	radio/TV	866
17	26	female	2	rent	little	unknown	1388	9	furniture/equipment	582
18	29	male	2	own	little	unknown	2743	28	radio/TV	426
19	53	male	2	free	little	little	4870	24	car	4
20	36	male	2	own	little	little	1721	15	car	461
21	38	male	2	own	little	unknown	804	12	radio/TV	997
22	29	male	2	own	little	moderate	1103	12	radio/TV	696
23	43	male	2	own	unknown	unknown	2197	24	car	406
24	27	male	2	own	little	little	3552	24	furniture/equipment	558
25	30	male	2	own	little	moderate	1056	18	car	580
26	24	female	2	own	little	moderate	2150	30	car	252
27	34	male	2	own	little	unknown	2750	12	furniture/equipment	452
28	24	female	2	rent	little	little	2124	18	furniture/equipment	761
29	34	male	2	own	little	moderate	3800	30	car	893
30	34	female	2	own	little	unknown	1493	12	radio/TV	638
31	29	female	2	own	little	unknown	1055	18	car	161

In general, parity at aggregate level does not imply parity at subgroup level!

A super-recent example

- Studying 1,200 bail cases from Indian High Courts and the Supreme Court (1975–2025), researchers found that different ML algorithms behave differently both with respect to **performance** (accuracy) and **fairness** (always measured in terms of accuracy)...

Model	Acc.	Gender	Crime	Region
LegalBERT (Base)	93.9 ± 1.6	2.3 ± 0.9	22.8 ± 4.0	34.7 ± 18.1
LegalBERT (Red.)	82.2 ± 4.1	5.8 ± 5.2	33.6 ± 2.9	49.0 ± 15.7
Logistic Reg.	93.1 ± 0.9	3.9 ± 3.5	15.0 ± 2.1	34.7 ± 17.1
XGBoost	95.6 ± 0.5	$1.8 \pm 1.2^\dagger$	20.0 ± 5.6	25.0 ± 7.9

*Prediction based on
summary!*

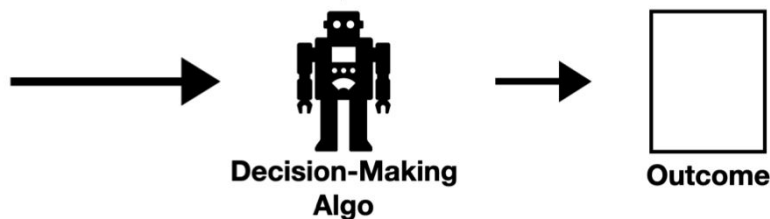
*What about world bias?
What about other fairness
measures? What about
intersectionality?*

On the fundamental problem of (un)fairness

(ongoing work)

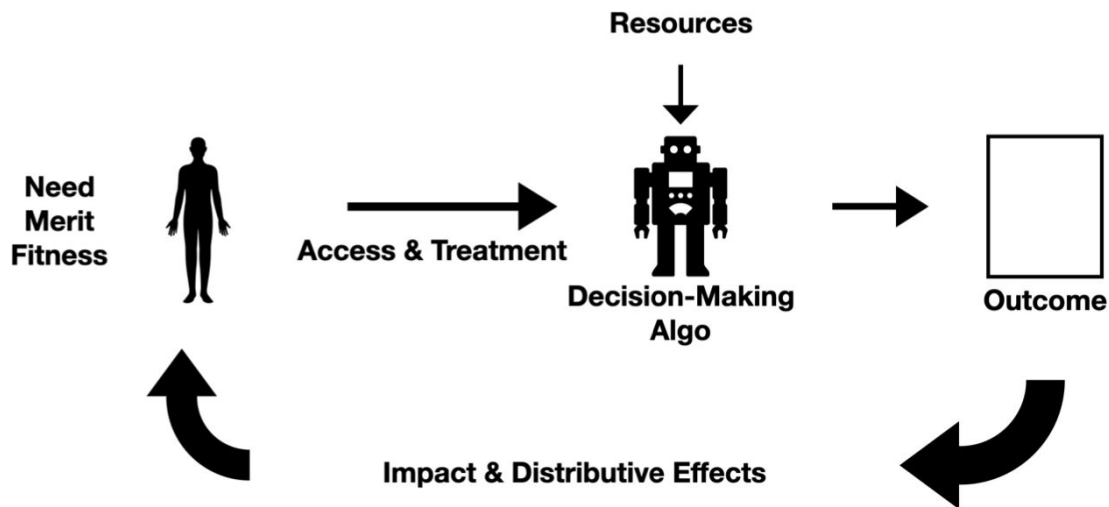
Background: Parity about what?

- Algorithmic approaches to fairness overlook the problem of defining of what fairness should be about (it is just about the decision)...



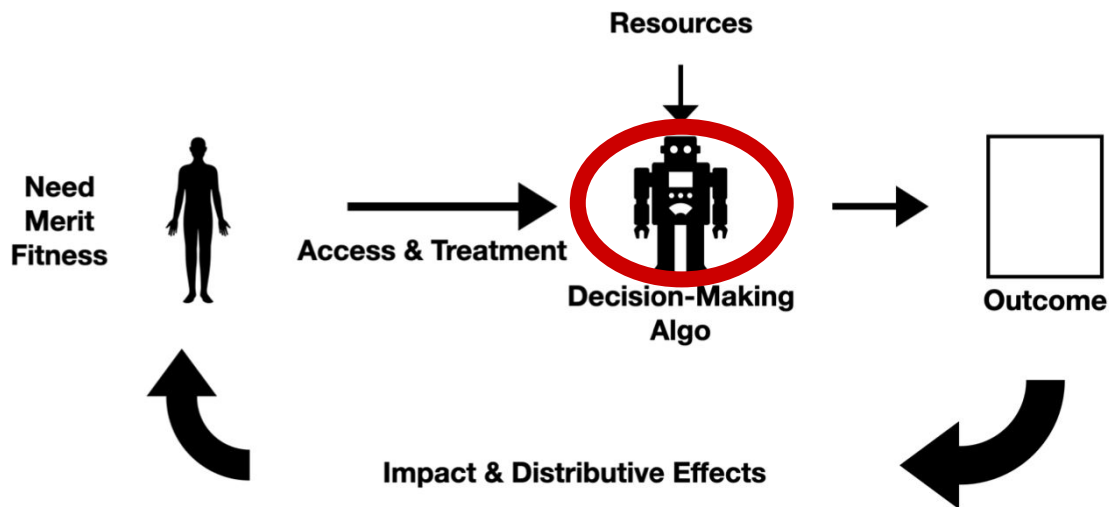
Background: Parity about what?

- Algorithmic approaches to fairness overlook the problem of defining of what fairness should be about (it is just about the decision). In contrast, philosophy has been debating for a long time which dimensions are relevant:



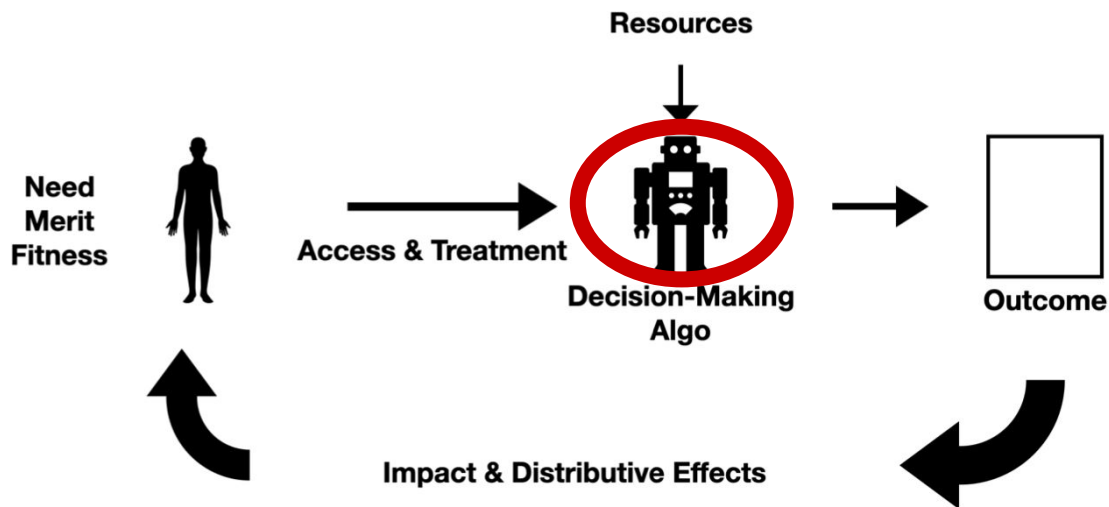
Background: Parity about what?

- Algorithmic approaches to fairness overlook the problem of defining of what fairness should be about (it is just about the decision). In contrast, philosophy has been debating for a long time which dimensions are relevant:



Background: Parity about what?

- Algorithmic approaches to fairness overlook the problem of defining of what fairness should be about (it is just about the decision). In contrast, philosophy has been debating for a long time which dimensions are relevant:

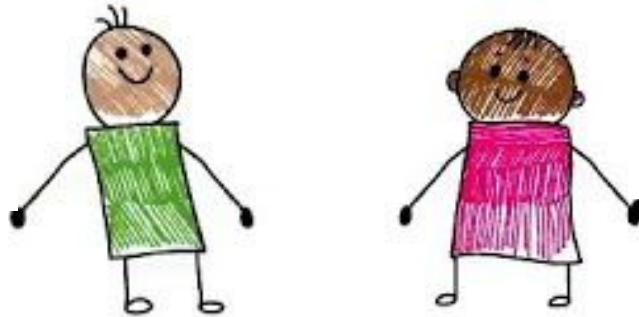


- In the wider system, Algorithmic fairness **may not bring fairer results!** (work of a Master AI student, still to be published).

Approaches to Fairness

There exists two main families of fairness models:

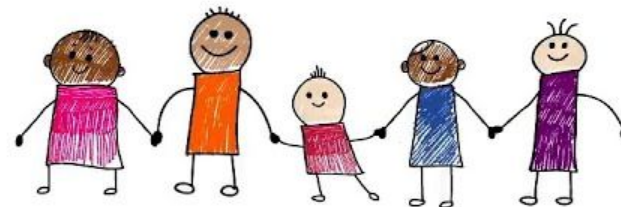
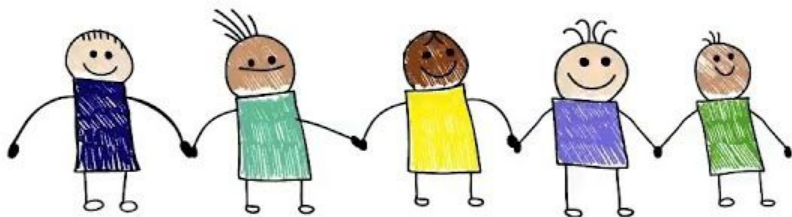
- **Individual fairness** compares treatment amongst individuals, based upon the principle that *“similar should be treated similarly”*



Approaches to Fairness

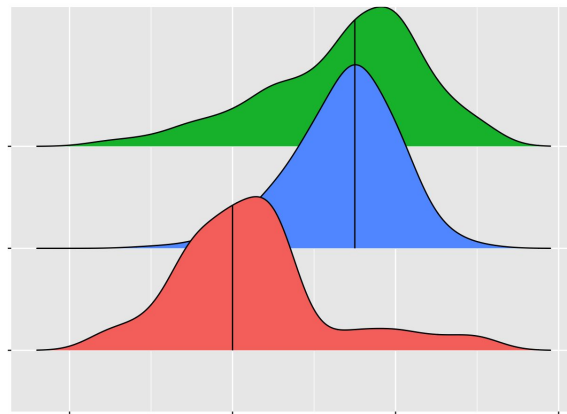
There exists two main families of fairness models:

- **Individual fairness** compares treatment amongst individuals, based upon the principle that “*similar should be treated similarly*”
- **Group fairness** compares treatment amongst groups (rather sub-groups of a given population) typically identified by sensitive attributes (e.g. *gender, ethnicity, age*, etc.), e.g. “women should be treated as men”.



Group fairness

- **Group fairness** is the one that has *by far* received more attention in the computational literature, as it provides a clear connection to *statistical methods*.



Group fairness vs Individual fairness

- **Group fairness** is the one that has *by far* received more attention in the computational literature, as it provides a clear connection to *statistical methods*.
- **Individual fairness** has traditionally attracted more attention in *non-computational* literature. At the end, claims of unfairness are primarily made about decisions towards individuals.



Group fairness vs Individual fairness

- **Group fairness** is the one that has *by far* received more attention in the computational literature, as it provides a clear connection to *statistical methods*.
- **Individual fairness** has traditionally attracted more attention in *non-computational* literature. At the end, claims of unfairness are primarily made about decisions towards individuals.
- An extension to individual fairness can be centred on **proportionality**, ie. that *differences in outcome of treatment should be proportional to differences between the two compared individuals*.

Fairness in law

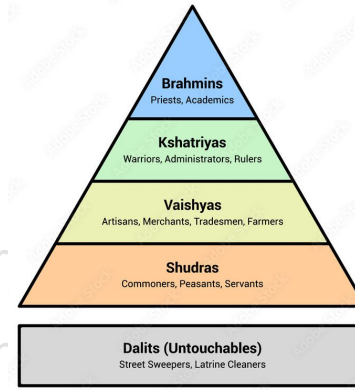
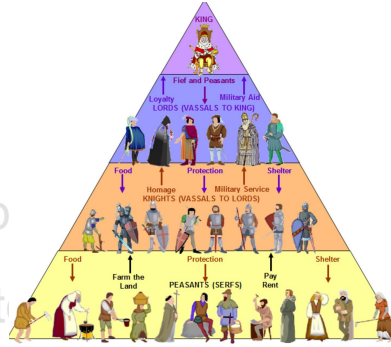
- Law addresses **categories** of people (*roles*), never *instances* (*individuals*).
- **Generality** is one of the accepted principles of sound legal systems.

Fairness in law

- Law addresses **categories** of people (*roles*), never *instances* (*individuals*).
- **Generality** is one of the accepted principles of sound legal systems
- Yet, legal systems may still be **unfair**, by privileging certain categories of people rather than others,
 - *explicitly*, by
 - disabling/enabling people selectively,
 - and/or distributing burdens unequally.

Fairness in law

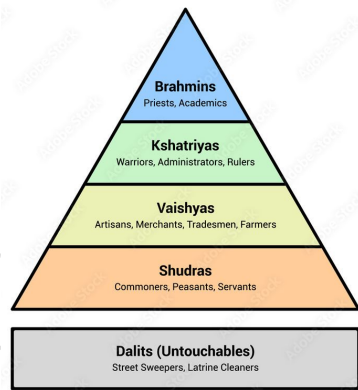
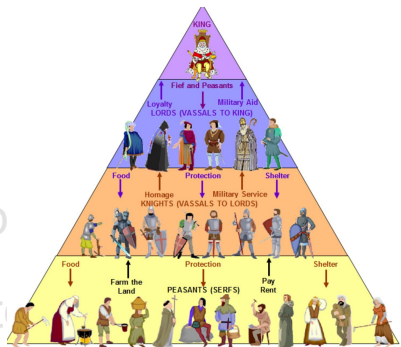
- Law addresses **categories** of people
- **Generality** is one of the accepted principles of law (applies to all individuals).
- Yet, legal systems may still be **unfair**, by privileging certain categories of people rather than others,
 - *explicitly*, by
 - disabling/enabling people selectively,
 - and/or distributing burdens unequally.



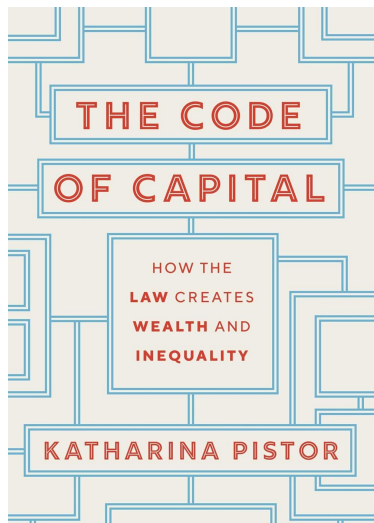
*class or
caste-based
systems*

Fairness in law

- Law addresses **categories** of people
- **Generality** is one of the accepted principles of law (applies to all individuals).
- Yet, legal systems may still be **unfair**, by privileging certain categories of people rather than others,
 - *explicitly*, by
 - disabling/enabling people selectively,
 - and/or distributing burdens unequally.
 - *implicitly*, for consequent non-trivial social effects

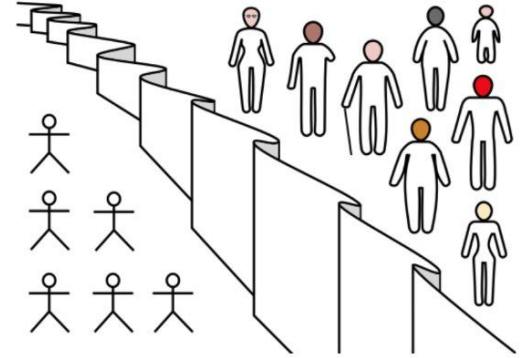


*class or
caste-based
systems*



Veil of ignorance?

- The *veil of ignorance* approach to “just laws” proposed by John Rawls states that a law should be deemed right if people would accept it even without knowing whether they are amongst the ones carrying its burden/its benefit.
- In his words: “*A just society is a society that if you knew everything about it, you’d be willing to enter it in a random place.*”



Unawareness...

- The *veil of ignorance* connects to the **Fairness Through Unawareness (FTU)** approach: by “forgetting” sensitive features, we are inhibited to take a biased decision (although this may still occur due to proxy features).

If we cannot see in what we are different, we cannot be judged differently.

Unawareness vs Awareness

- An approach opposite to FTU, called **Fairness through Awareness (FTA)**, has also been presented in the literature on fairness in classification.
- According to FTA, a classifier will be *fair* if the outcome distributions $\mathbf{M}(\mathbf{x})$ and $\mathbf{M}(\mathbf{y})$ concerning any two individuals $\mathbf{x}$ and $\mathbf{y}$ distant $d(\mathbf{x}, \mathbf{y})$, are statistically distant at $\max d(\mathbf{x}, \mathbf{y})$.

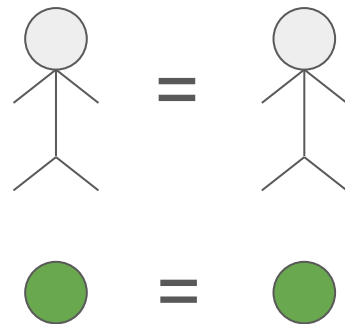
Unawareness vs Awareness

- An approach opposite to FTU, called **Fairness through Awareness (FTA)**, has also been presented in the literature on fairness in classification.
- According to FTA, a classifier will be *fair* if the outcome distributions $\mathbf{M}(\mathbf{x})$ and $\mathbf{M}(\mathbf{y})$ concerning any two individuals $\mathbf{x}$ and $\mathbf{y}$ distant $d(\mathbf{x}, \mathbf{y})$, are statistically distant at $\max d(\mathbf{x}, \mathbf{y})$.
- **Problem:** where this (similarity) metric comes from is identified as an open question by the authors.

Individual fairness (again)

- In the traditional formulation:

We are similar $\rightarrow$ we are treated similarly



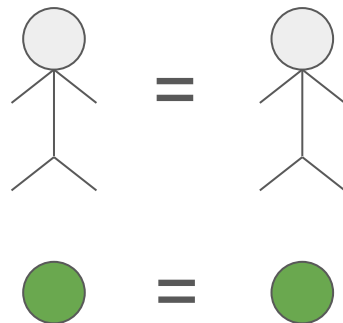
Individual fairness (again)

- In the traditional formulation:

We are similar $\rightarrow$ we are treated similarly

- We can write it in its contrapositive form:

We are not treated similarly $\rightarrow$ we are not similar



Individual fairness (again)

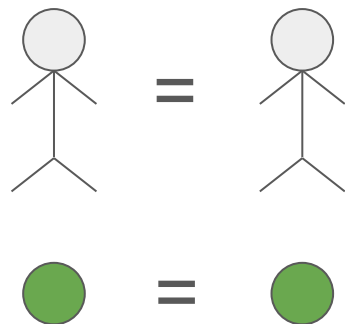
- In the traditional formulation:

We are similar $\rightarrow$ we are treated similarly

- We can write it in its contrapositive form:

We are not treated similarly $\rightarrow$ we are not similar

➡ The contrapositive makes clear that the decision has consequences on determining the similarity metric. In other words, a classifier presupposes classes of equivalence applicable to the dataset.



Individual fairness (again)

- In the traditional formulation:

We are similar $\rightarrow$ we are treated similarly

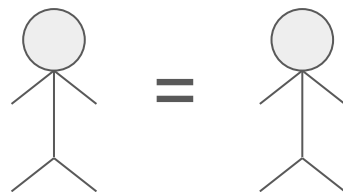
- We can write it in its contrapositive form:

We are not treated similarly $\rightarrow$ we are not similar

 The contrapositive makes clear that the decision has consequences on determining the similarity metric. In other words, a classifier presupposes classes of equivalence applicable to the dataset.

- **Unfairness** is defined as the negation of fairness:

We are similar, yet we are not treated similarly.



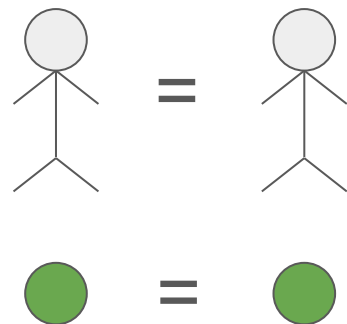
Individual fairness (again)

- In the traditional formulation:

We are similar $\rightarrow$ we are treated similarly

- We can write it in its contrapositive form:

We are not treated similarly $\rightarrow$ we are not similar



➡ The contrapositive makes clear that the decision has consequences on determining the **similarity metric**. In other words, a classifier presupposes **classes of equivalence** applicable to the dataset.

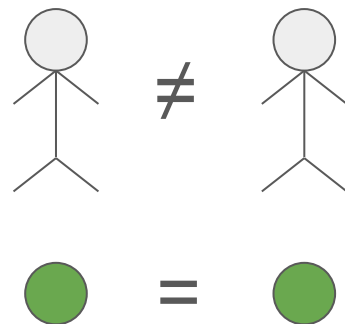
OOOPS: METRICS + TOPOLOGY !!!

Group fairness (again)

- Group fairness has a different formulation than individual fairness. Here, we have sensitive attributes (gender, etc.) that should not be used to reach a conclusion.

Group fairness (again)

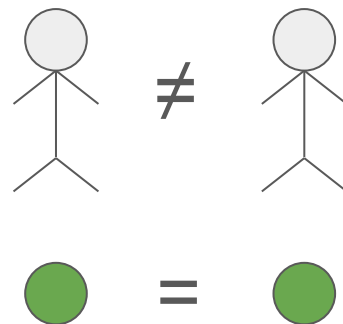
- Group fairness has a different formulation than individual fairness. Here, we have sensitive attributes (gender, etc.) that should not be used to reach a conclusion.
- Therefore, even if we are different with respect to a certain sensitive feature, we should be treated the same, because this feature has to be treated as ***irrelevant*** in the decision process:



Even if we are different for a sensitive feature, we are treated the same.

Group fairness (again)

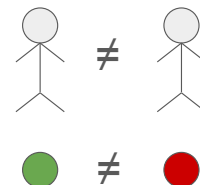
- Group fairness has a different formulation than individual fairness. Here, we have sensitive attributes (gender, etc.) that should not be used to reach a conclusion.



- Therefore, even if we are different with respect to a certain sensitive feature, we should be treated the same, because this feature has to be treated as **irrelevant** in the decision process:

Even if we are different for a sensitive feature, we are treated the same.

- Unfairness** is defined as the negation of the above rule:



We are different for a feature, AND we are treated differently.

Individual vs Group fairness

- If individual fairness makes reference to variation on input having effects on output, group fairness appeals to absence of effects on output
- The keyword unifying the two views is “relevant”:
 - Individual: *We are RELEVANTLY similar, yet we are not treated similarly.*
 - Group: *We are IRRELEVANTLY different, yet we are not treated similarly.*

Individual vs Group fairness

- If individual fairness makes reference to variation on input having effects on output, group fairness appeals to absence of effects on output.
- The keyword unifying the two views is “relevant”:
 - Individual: *We are RELEVANTLY similar, yet we are not treated similarly.*
 - Group: *We are IRRELEVANTLY different, yet we are not treated similarly.*



Group-fairness seems easier to apply as we tend to prescribe explicitly irrelevant features!

Individual vs Group fairness

- If individual fairness makes reference to variation on input having effects on output, group fairness appeals to absence of effects on output.
- The keyword unifying the two views is “relevant”:
 - Individual: *We are RELEVANTLY similar, yet we are not treated similarly.*
 - Group: *We are IRRELEVANTLY different, yet we are not treated similarly.*

 Group-fairness seems easier to apply as we tend to prescribe explicitly irrelevant features!

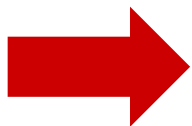
 Yet, irrelevance depends on context. For instance gender is not necessarily problematic when suggesting dresses on a web-shop.

Fairness vs Unfairness?

- All our elaboration so far brought to the foreground that the difficulty of defining what fairness is may be a side-effect of the fact that we generally react against *unfairness*.

Fairness vs Unfairness?

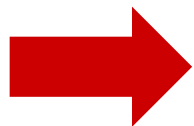
- All our elaboration so far brought to the foreground that the difficulty of defining what fairness is may be a side-effect of the fact that we generally react against *unfairness*.



Fairness is therefore a derived notion, **resulting from the absence of unfairness**, rather than having a positive description.

Fairness vs Unfairness?

- All our elaboration so far brought to the foreground that the difficulty of defining what fairness is may be a side-effect of the fact that we generally react against *unfairness*.



Fairness is therefore a derived notion, **resulting from the absence of unfairness**, rather than having a positive description.

fair = `not` **unfair**

where `not` works as the default negation or negation as failure (NAF) in logic programming: *it returns true if unfair cannot be proven*

Unfair with respect to what?

- We have seen that philosophers have discussed several dimensions relevant to fairness (need, treatment, outcome, access, ...

these are **all possible dimensions where we can perceive unfairness.**

Fairness as an argumentative matter

- **(Division of the loot/Sharing of the spoils)** *Let us imagine to have N pirates. The usual fair division of the loot, whose total value is T , would suggest T/N for each of them.*



Fairness as an argumentative matter

- **(Division of the loot/Sharing of the spoils)** *Let us imagine to have N pirates. The usual fair division of the loot, whose total value is T , would suggest T/N for each of them.*
- *But then... one pirate says: “that’s not fair, I fought more bravely than others”. Another say, “that’s not fair, I have only one leg, I could not go as fast as you”. Another says, “that’s not fair, I have plenty of kids waiting at home, you have no family to feed”. A last one says, “nobody woke me up!”*



Fairness as an argumentative matter

- (Division of the loot/Sharing of the spoils) *Let us imagine to have N pirates. The usual fair division of the loot, whose total value is T , would suggest T/N for each of them.*
- *But then... one pirate says: “that’s not fair, I fought more bravely than others”. Another say, “that’s not fair, I have only one leg, I could not go as fast as you”. Another says, “that’s not*



unfairness becomes primarily a **matter of argumentation**, as parties bring their position in the foreground, making clear in what they differ. Because the number of variations is open-ended, we cannot cover it extensionally...

The core mechanism (1)

- A decision is taken by considering certain input factors, and determining a certain output. The context is however always wider than the input, as well as impact is wider than the decision output.

The core mechanism (1)

- A decision is taken by considering certain input factors, and determining a certain output. The context is however always wider than the input, as well as impact is wider than the decision output.
- Schematically:

CONTEXT $\rightarrow$ INPUT $\rightarrow$ OUTPUT $\rightarrow$ IMPACT

with CONTEXT $\supset$ INPUT, and IMPACT $\supset$ OUTPUT

The core mechanism (1)

- A decision is taken by considering certain input factors, and determining a certain output. The context is however always wider than the input, as well as impact is wider than the decision output.

- Schematically:

CONTEXT $\rightarrow$ INPUT $\rightarrow$ OUTPUT $\rightarrow$ IMPACT

with **CONTEXT $\supset$ INPUT**, and **IMPACT $\supset$ OUTPUT**

- Applying proportionality to the wider picture, we have:

$\Delta\text{CONTEXT} \sim \Delta\text{IMPACT}$

The core mechanism (2)

- Unfairness occurs when $\Delta\text{IMPACT} - \Delta\text{CONTEXT} \neq 0$, so in two cases:
 - no variation of impact with variation of context
 - variation of impact with no variation of context

The core mechanism (2)

- Unfairness occurs when $\Delta\text{IMPACT} - \Delta\text{CONTEXT} \neq 0$, so in two cases:
 - **no variation of impact with variation of context**
 - variation of impact with no variation of context

The core mechanism (2)

- Unfairness occurs when $\Delta\text{IMPACT} - \Delta\text{CONTEXT} \neq 0$, so in two cases:
 - **no variation of impact with variation of context**
 - variation of impact with no variation of context
- **First case:** the decision-maker has seemingly no access to relevant categories (**over-simplification of the environment**)
 - The *burden of generating arguments* lies upon the patient of the decision



eg. legal recourse

The core mechanism (2)

- Unfairness occurs when $\Delta\text{IMPACT} - \Delta\text{CONTEXT} \neq 0$, so in two cases:
 - no variation of impact with variation of context
 - **variation of impact with no variation of context**

The core mechanism (2)

- Unfairness occurs when $\Delta\text{IMPACT} - \Delta\text{CONTEXT} \neq 0$, so in two cases:
 - no variation of impact with variation of context
 - **variation of impact with no variation of context**
- **Second case:** the patient had no access to relevant categories used by the decision-maker (**over-simplification of the decision-maker**).
 - The *burden of generating arguments* lies upon the decision-maker.



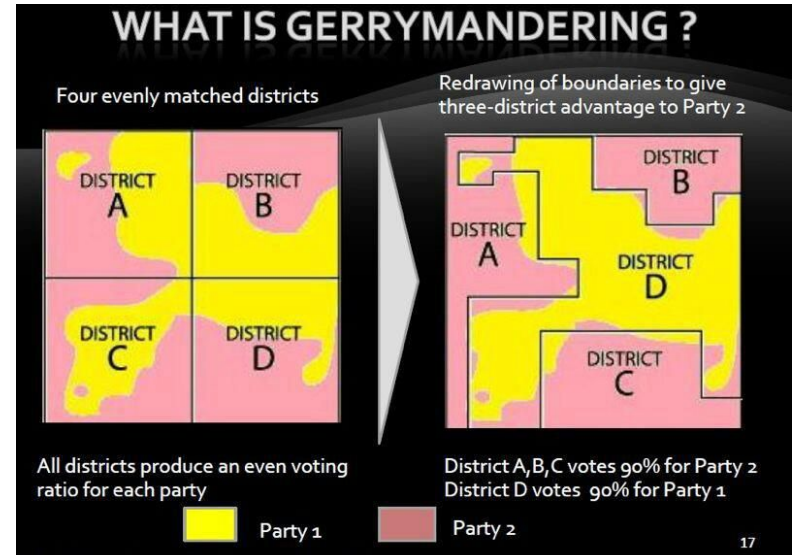
eg. distinguishing action

...ongoing and future work!

- I sketched here only the main lines of our arguments.
- In the full contribution, we will rely on Simplicity Theory, a non-probabilistic theory of cognition based on Kolmogorov complexity (description complexity)

...ongoing and future work!

- I sketched here only the main lines of our arguments.
- In the full contribution, we will rely on Simplicity Theory, a non-probabilistic theory of cognition based on Kolmogorov complexity (description complexity)
- We are making connections with ***gerrymandering*** problems, and with the various forms of justice (retributive, restorative, distributive, procedural, ...).



...ongoing and future work!

- I sketched here only the very first part of t
- In the full contribution, we will rely on Sim theory of cognition based on Kolmogorov
- We are making connections with the *gerrymandering* problems, and with the various forms of justice (retributive, restorative, distributive, procedural, ...).
- **Intuition:** at fundamental level, legal systems unfold as such because of underlying human cognitive mechanisms!

Simplicity Bias in Human-generated Data

Jean-Louis Dessalles (dessalles@telecom-paris.fr)

LTCI, Institut Polytechnique de Paris
France

Giovanni Sileno (g.sileno@uva.nl)

University of Amsterdam
the Netherlands

Abstract

Texts available on the Web have been generated by human minds. We observe that simple patterns are over-represented: *abcdef* is more frequent than *arfbxg* and 1000 appears more often than 1282. We suggest that word frequency patterns can be predicted by cognitive models based on complexity minimization. Conversely, the observation of word frequencies offers an opportunity to infer particular cognitive mechanisms involved in their generation.

Keywords: Simplicity; Kolmogorov complexity; Web frequency; Zipf's law

Introduction

The observation of word or number frequencies reveals a few spectacular patterns. The most obvious one is that simple items tend to be more frequent. For example, round numbers such as 1000 or consecutive strings like *abcde* appear orders of magnitude more frequently than items such as 1282) or *arfbxg*). Intuition suggests that some items are more frequent *because* they are descriptively simple. This form of causality, however, requires hypotheses about the way words

Caglioti, & Loreto, 2002) or semantic d words (Cilibrasi & Vitényi, 2007). In this paper, we study the frequency of numbers and words. We will attempt to support the idea that complexity at the cognitive level may produce some of the frequency patterns that can be observed in texts.

The paper is organized as follows. We begin with the notion of simplicity as it is applied in cognitive science and we discuss the way it is usually linked to word frequencies. Then we make three attempts to offer causal explanations for word frequencies. First, we propose a basic model to explain why the frequency of alphabetic sequences and passwords obeys a simplicity bias. Second, we examine number frequency distributions and propose an alternative model that partly accounts for their remarkable simplicity. Third, we examine word statistics; we show that the analysis on word frequencies points to certain hypotheses on complexity minimization in word generation.

Conclusion: Relevance to this RPA?

- **Polarization** may aggregate (communications of) supposed **unfairness**, which may be connected to **(perceived) inequality**, possibly caused by **segregation** processes.
Circularly, *polarization may cause further segregation*



Conclusion: Relevance to this RPA?

- **Polarization** may aggregate (communications of) supposed **unfairness**, which may be connected to **(perceived) inequality**, possibly caused by **segregation** processes.
Circularly, *polarization may cause further segregation*
- ***What about studying at scale arguments concerning unfairness?***



Conclusion: Relevance to this RPA?

- **Polarization** may aggregate (communications of) supposed **unfairness**, which may be connected to **(perceived) inequality**, possibly caused by **segregation** processes.
Circularly, *polarization may cause further segregation*

- ***What about studying at scale arguments concerning unfairness?***
- Identifying the sources of “contrast” (on impact/context) means supporting identifying where to dedicate resources and/or set-up mitigating policies.



The “Artificial” Intelligence of Law

RPA research group on Segregation, Polarisation and Inequality
Amsterdam, 18 December, 2025



Giovanni Sileno
g.sileno@uva.nl

University of Amsterdam