INTERDISCIPLINARY STUDIES OF LAW [METAJURIDICA]
VRIJE UNIVERSITEIT BRUSSEL

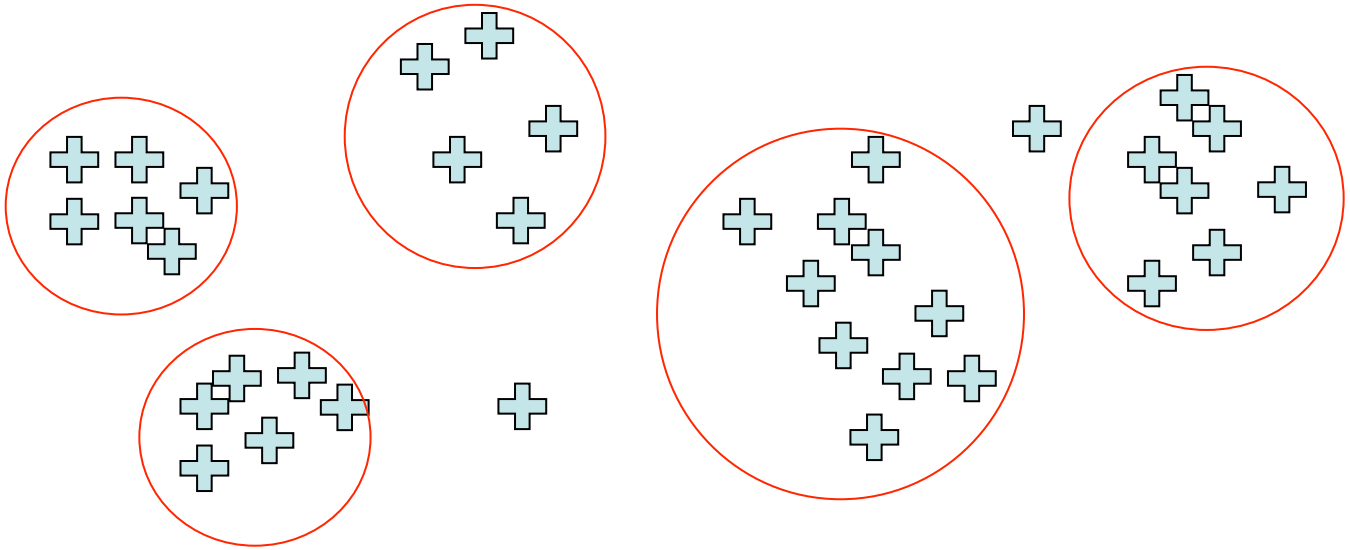
LSTS
LAW, SCIENCE, TECHNOLOGY & SOCIETY STUDIES
VRIJE UNIVERSITEIT BRUSSEL
BELGIUM

iCIS | Digital Security

Radboud Universiteit Nijmegen

Some legal conundrums in relation to machines that produce differences and similarities.

Algorithmic transparency and the thin line between justified differentiations /unjustified discriminations.

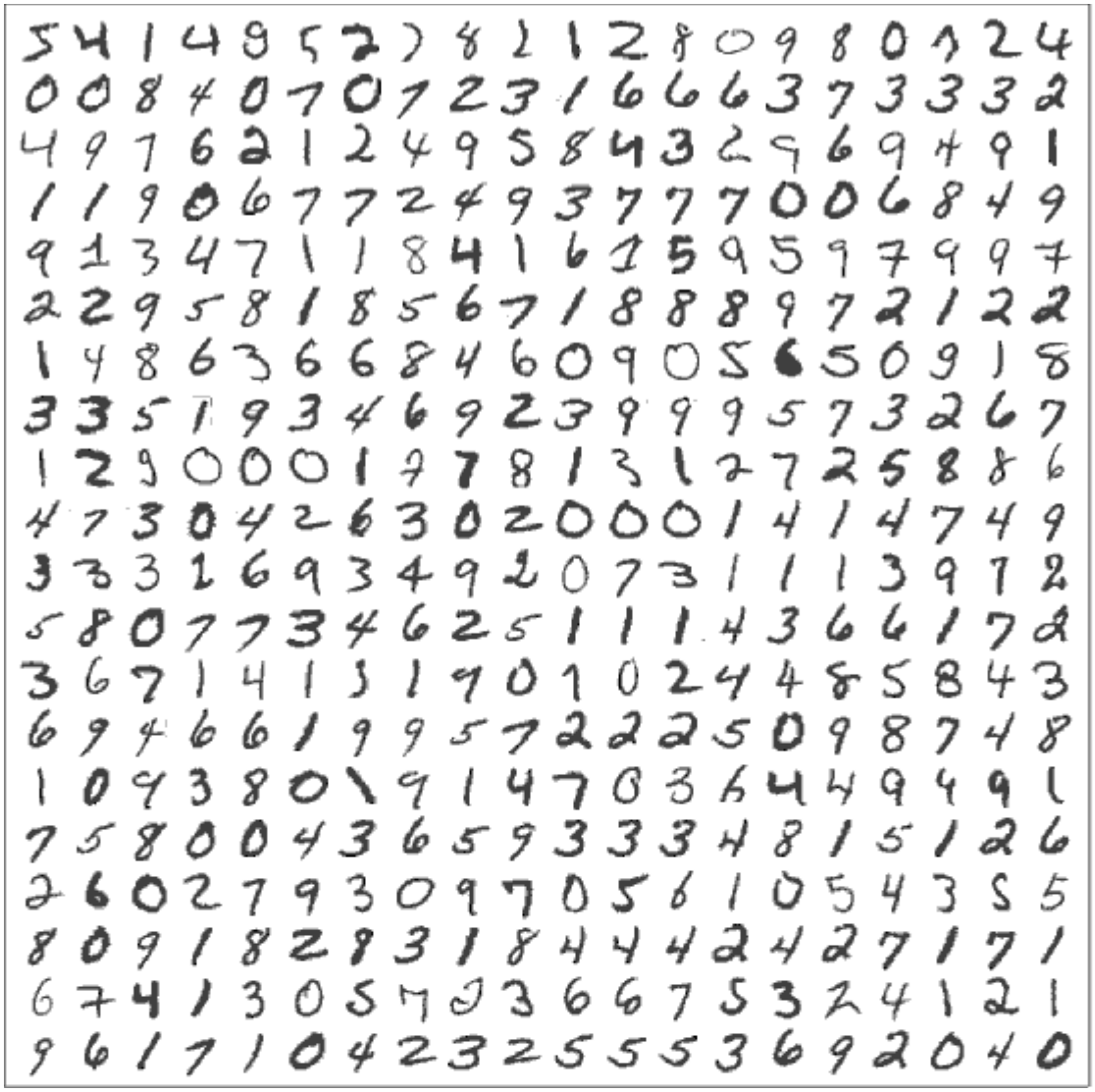


Katja de Vries (iCIS, RUN and LSTS, VUB)

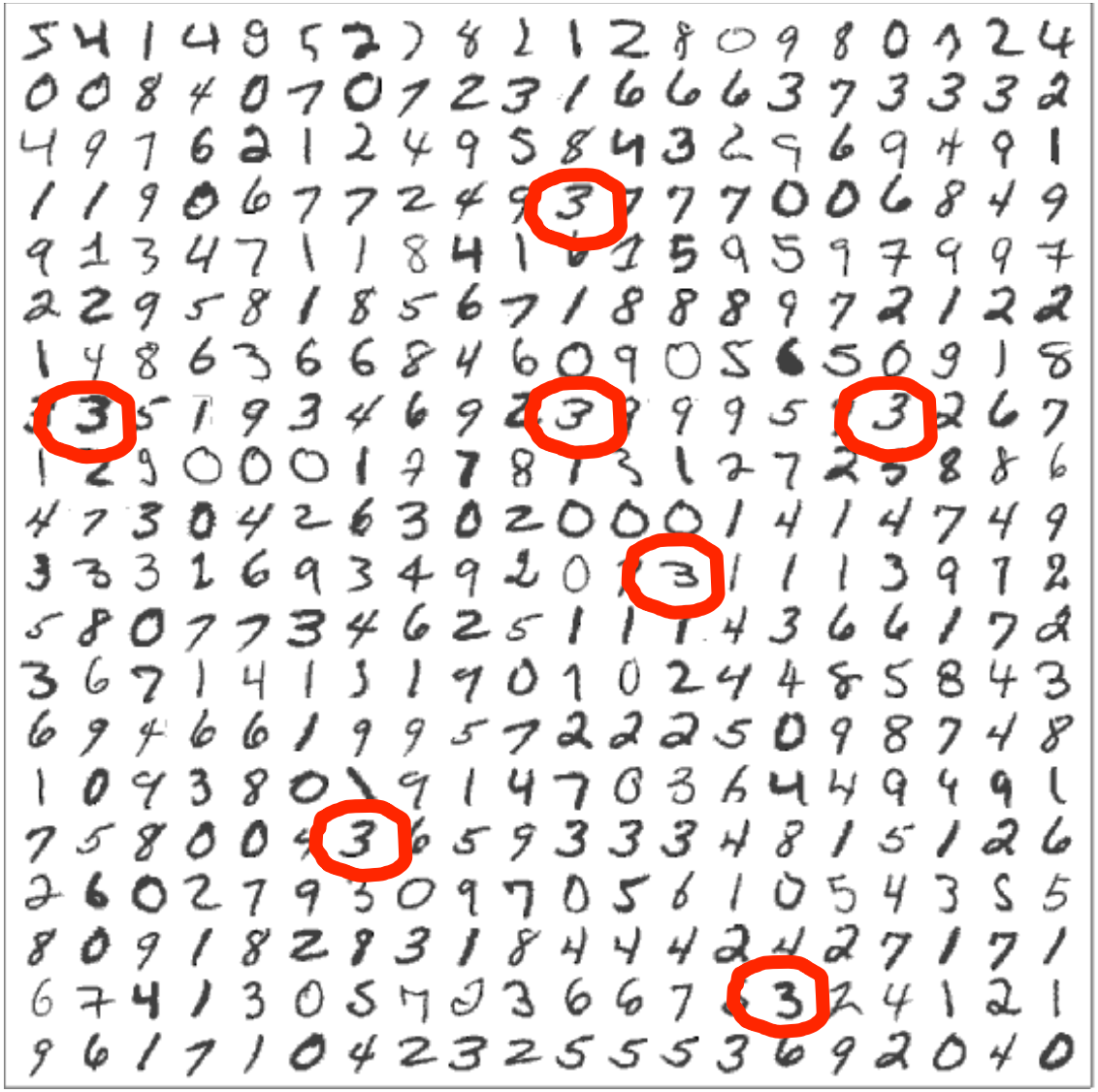
26 May 2016 – ETHOS, ITU, Copenhagen

Prologue

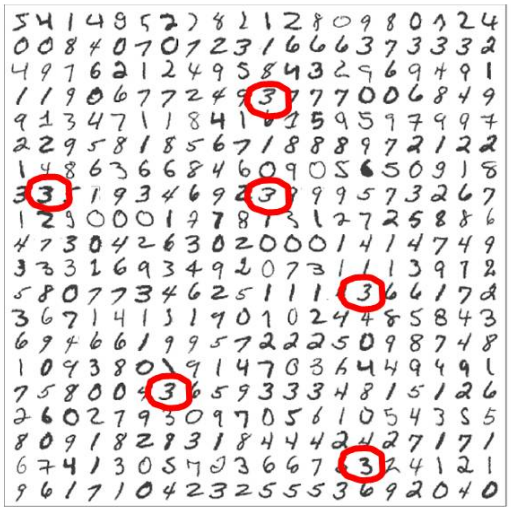
"... never do we find two eggs or two leaves or two blades of grass in a garden that are perfectly similar. And thus, perfect similarity is found only in incomplete and abstract notions [...]" G.W. Leibniz. *Logical-metaphysical principles* (spring-summer 1689?). In: Strickland L, editor. 'The shorter Leibniz texts'. London: Continuum; 2006. p. 48-52.

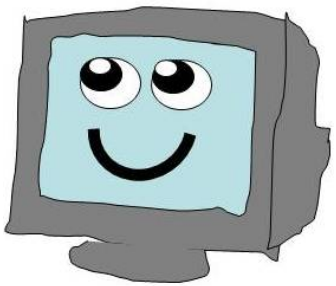


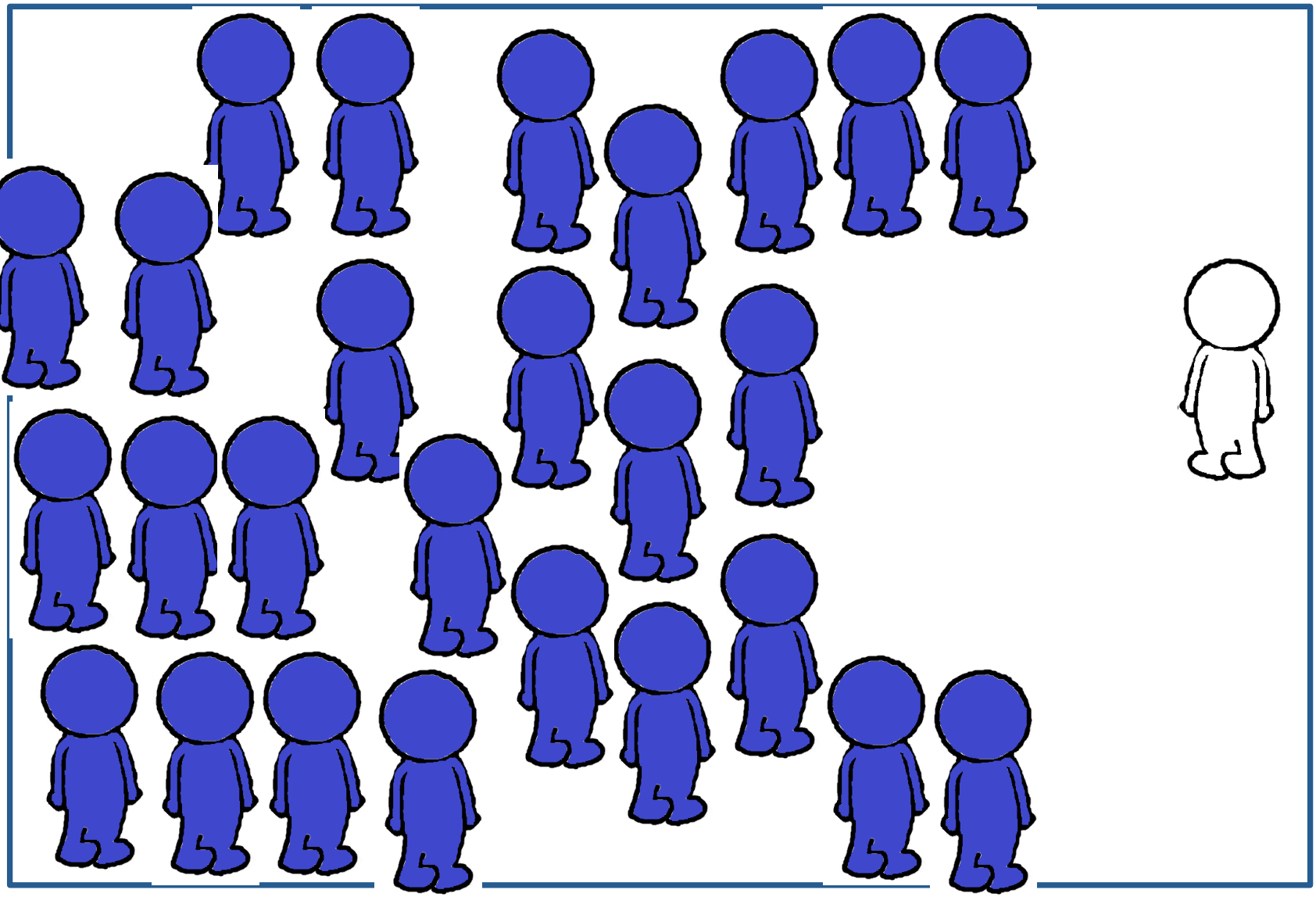
The MNIST database of handwritten digits. By Yann LeCun and Corinna Cortes.
Image source: exerpt from the MNIST database as reproduced in MacCormick, J. (2012). *Nine Algorithms That Changed the Future. The Ingenious Ideas That Drive Today's Computers*. Princeton: Princeton University Press, p.83

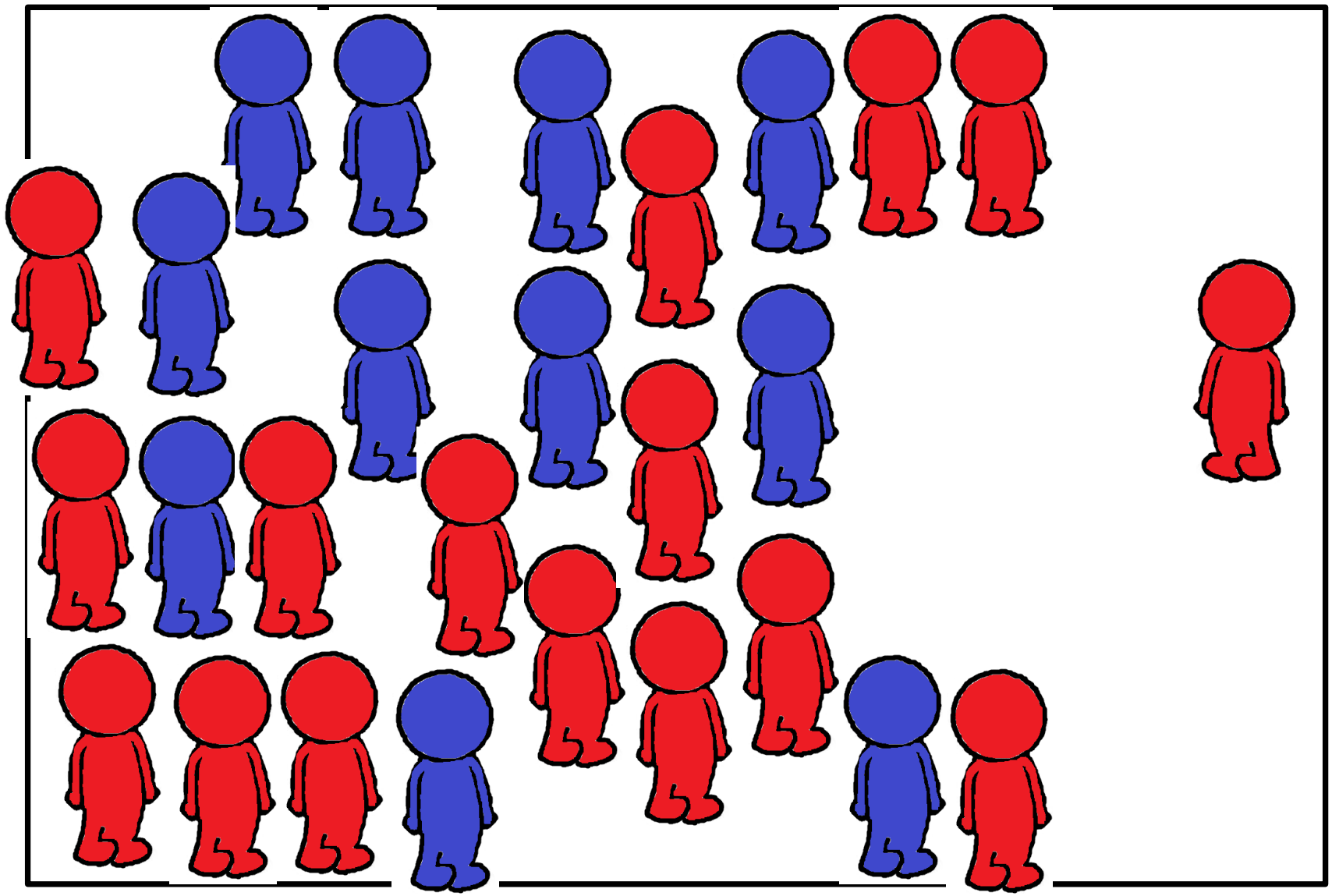


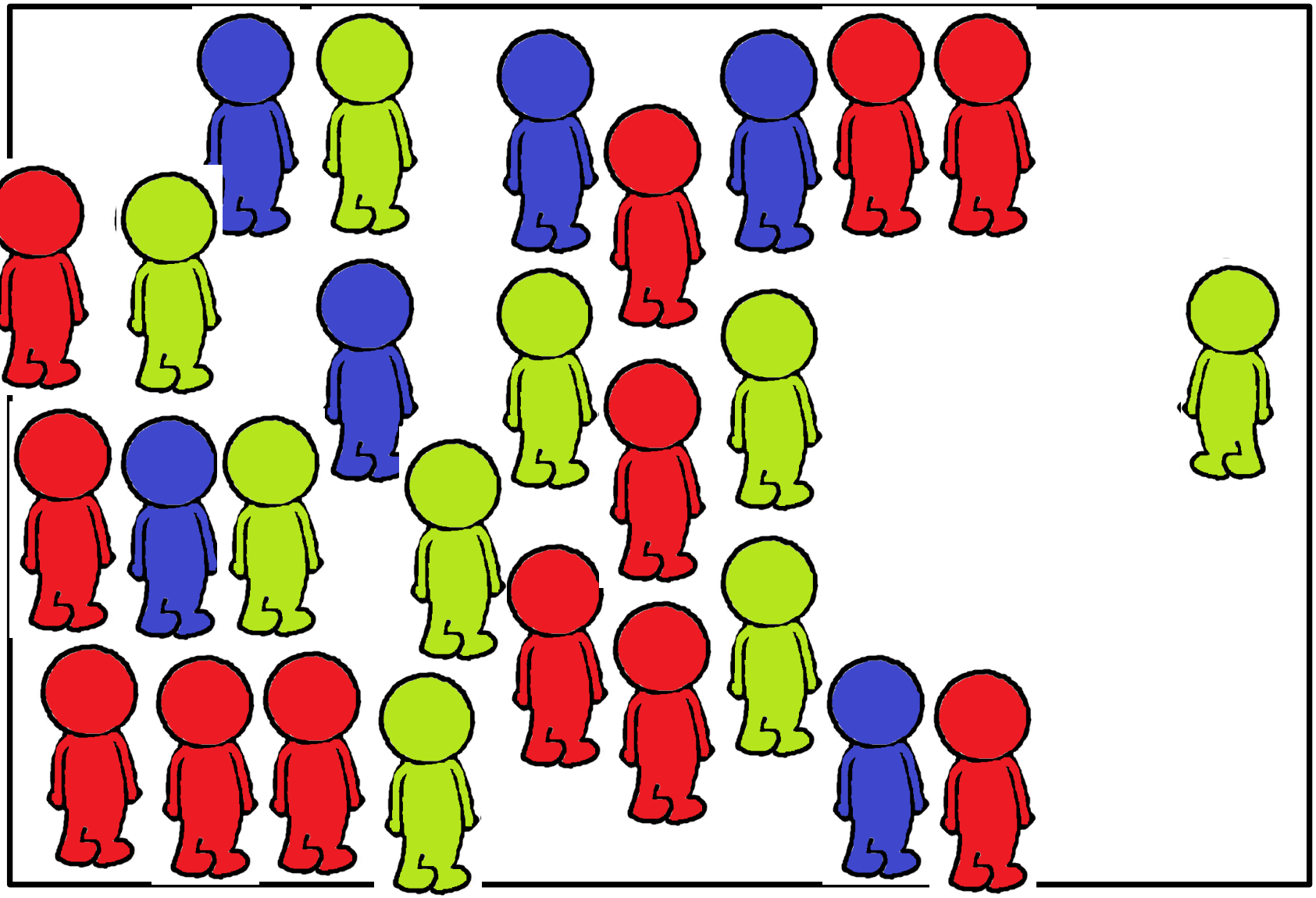
= 3

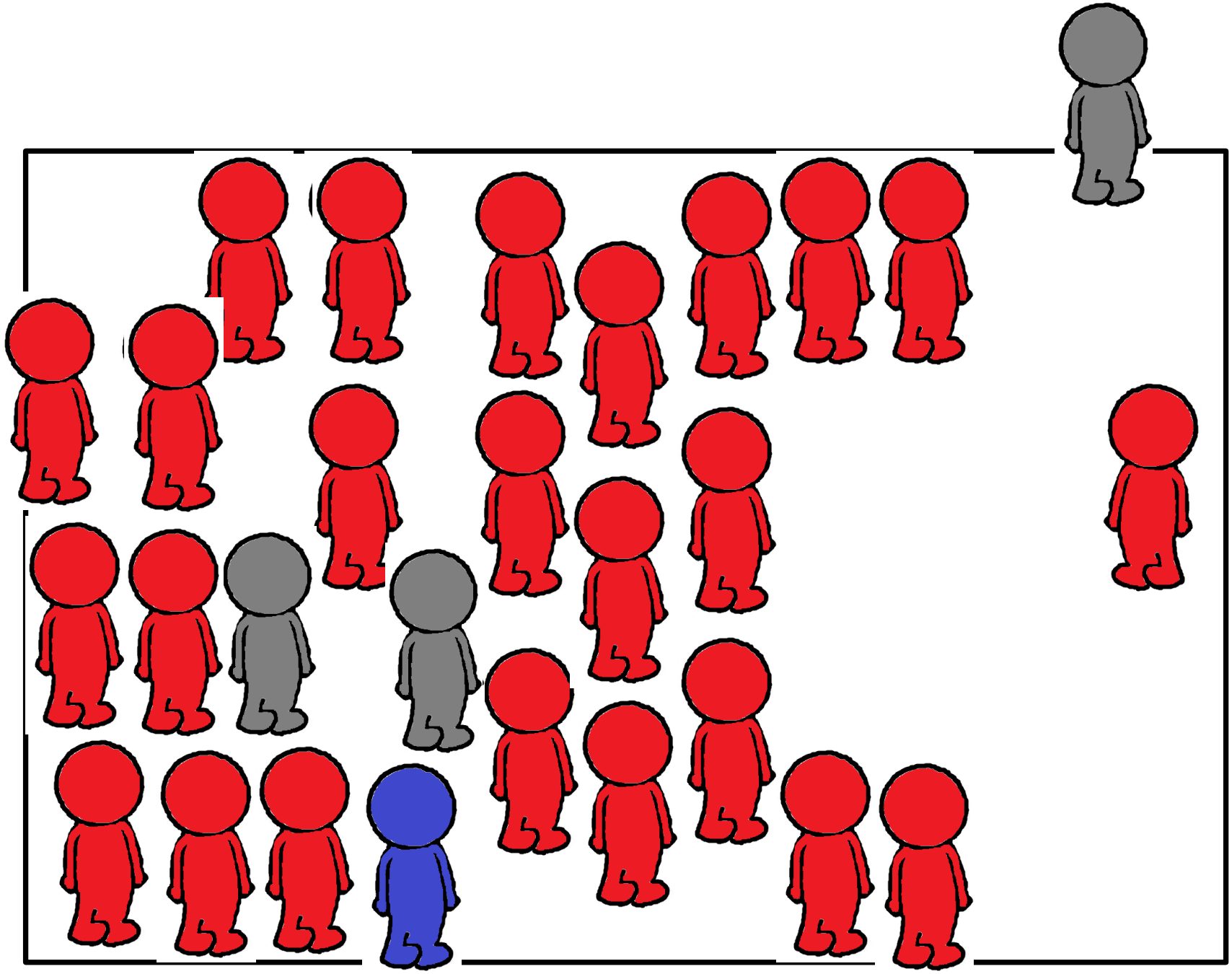












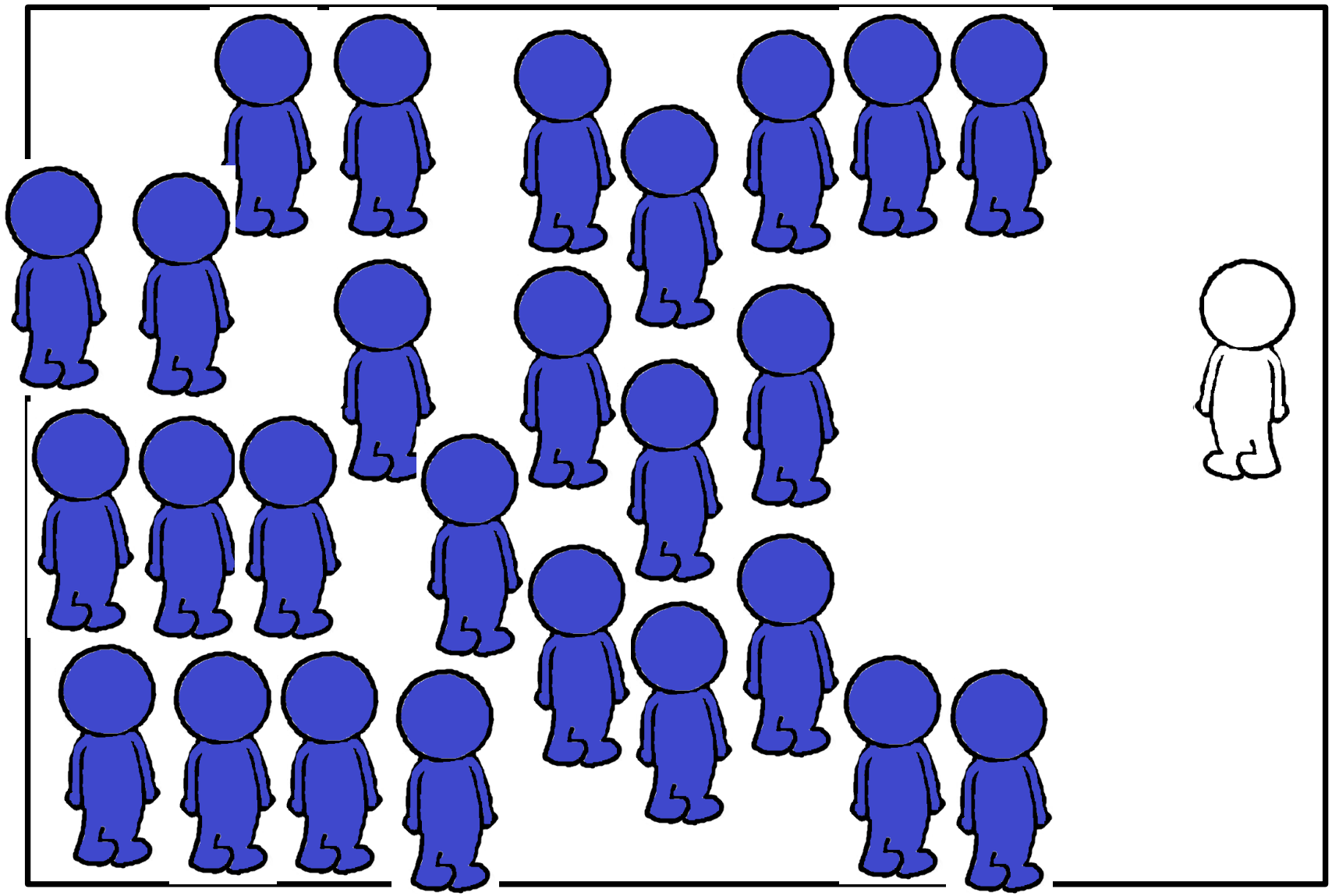
We see something *as something*
by generalizing former experiences

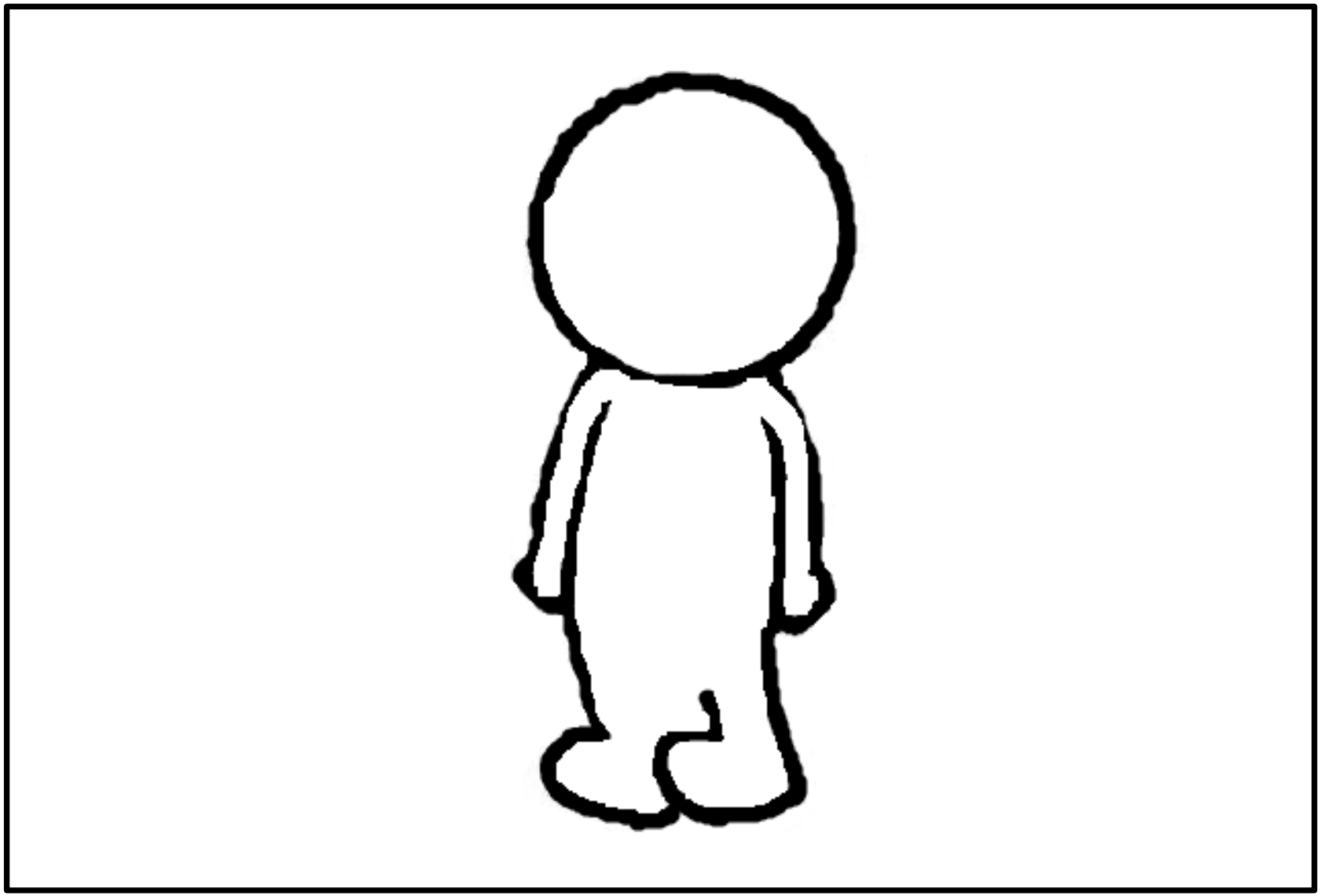
Not a one-way road

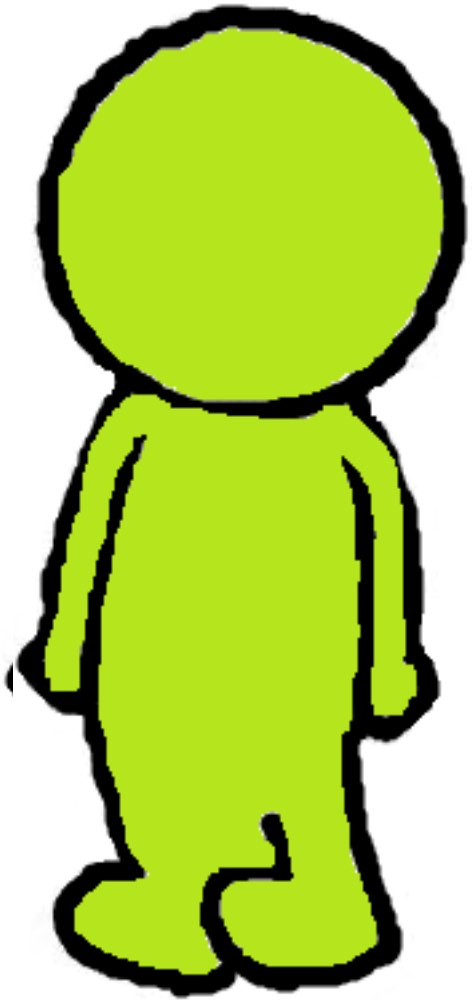
Profiling

AND

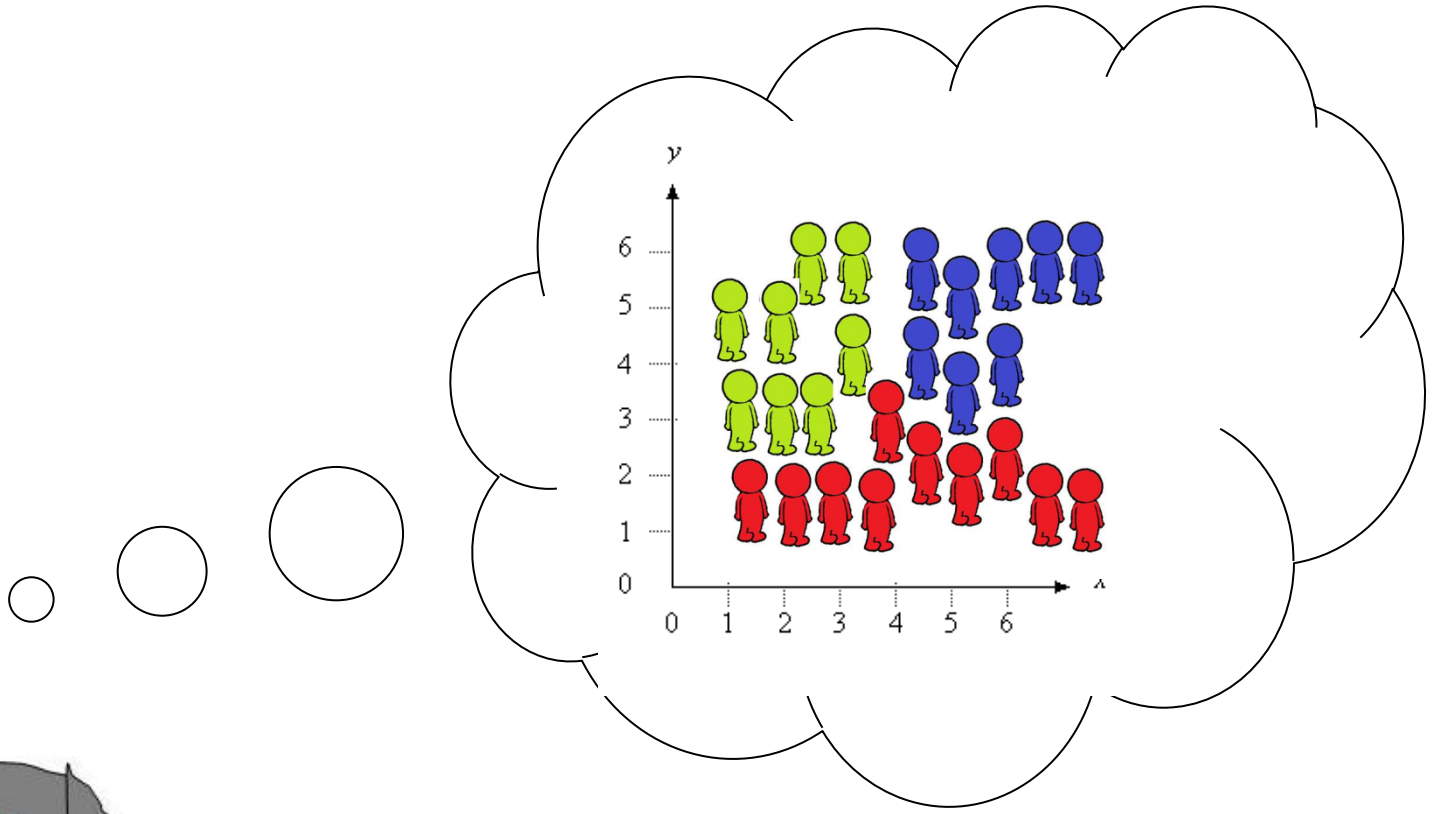
Counter profiling

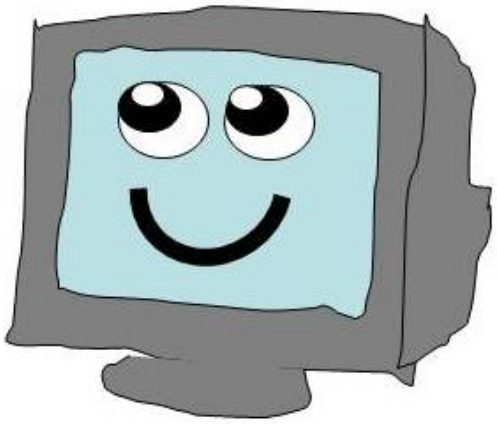






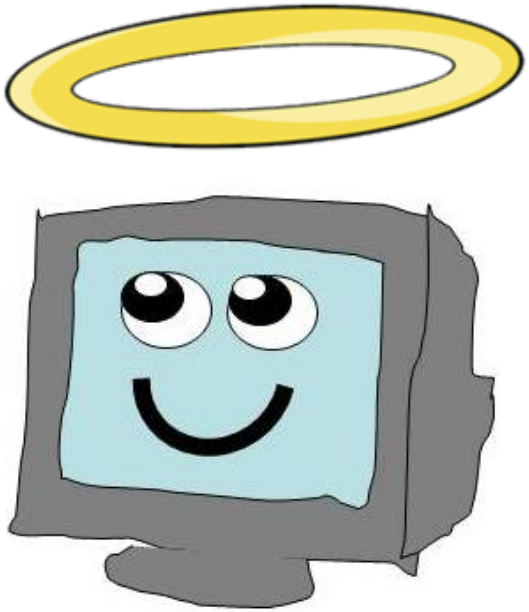
Machine learning applied to humans

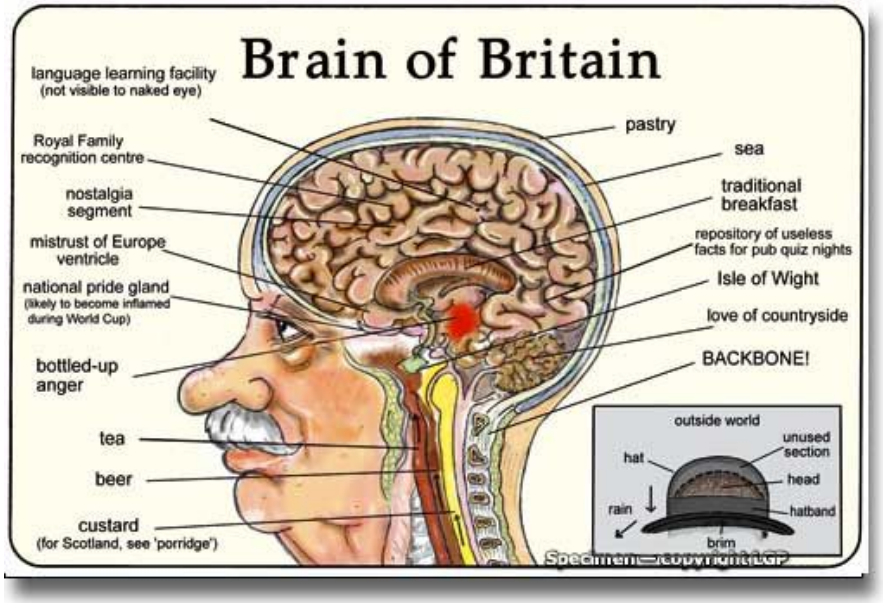




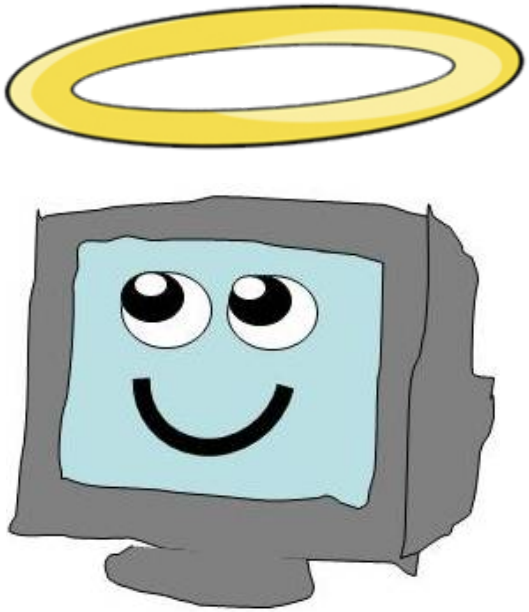
Two persistent myths about machine differentiations

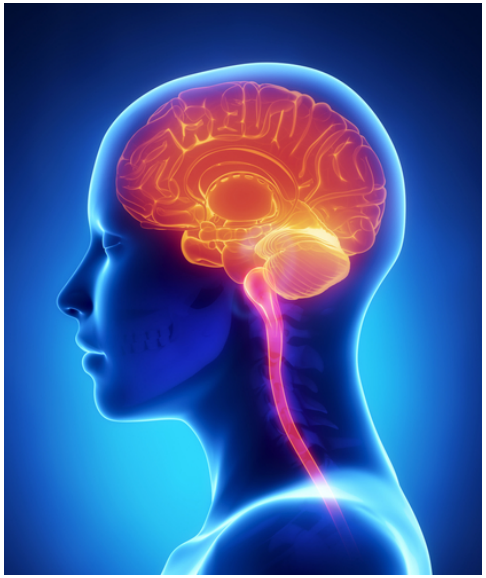
1. A machine is more neutral, rational and objective in its differentiations



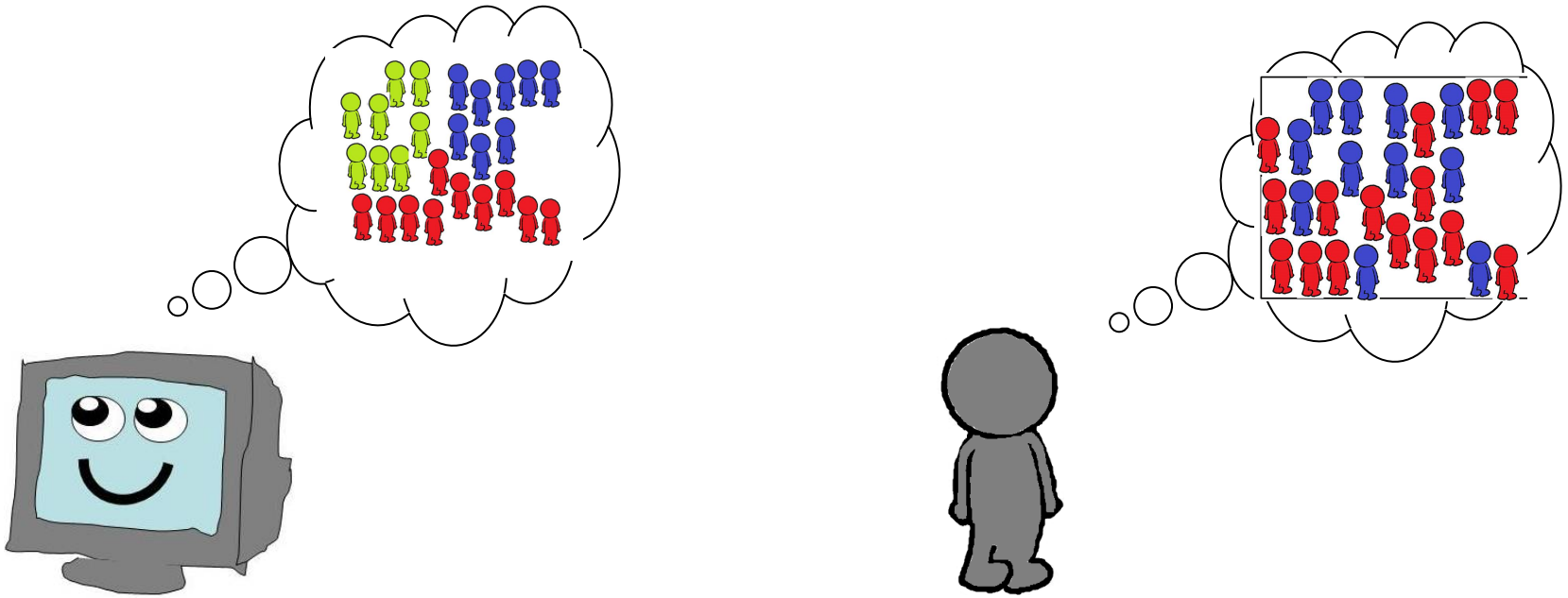


2. Machine differentiations are more transparent (*because you never know what goes on in the head of another human*)



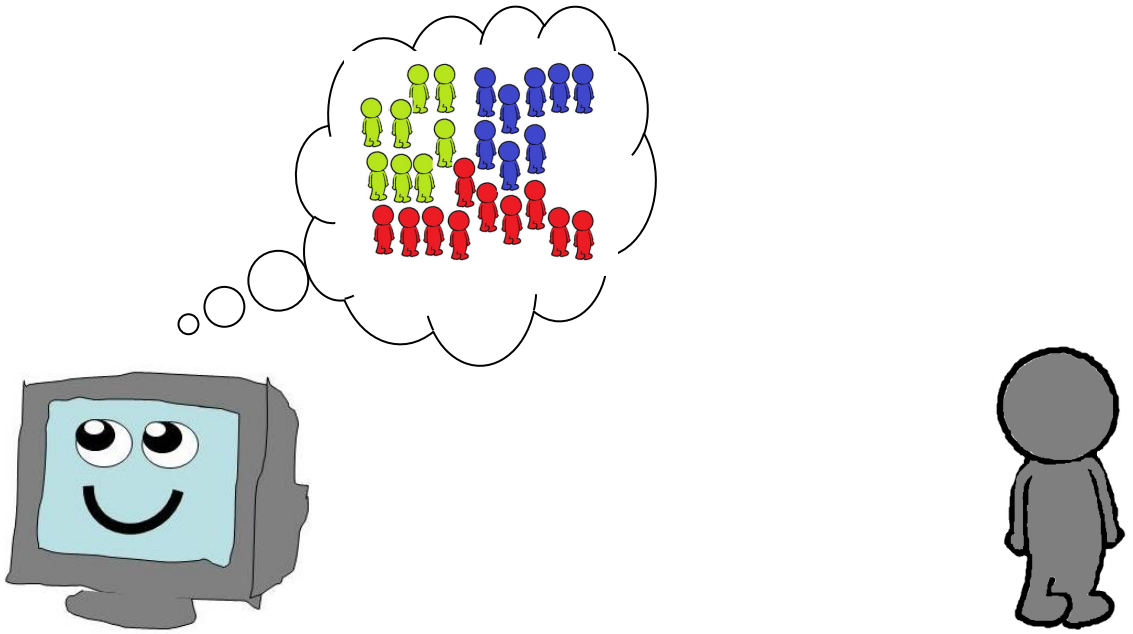


Not better or worse – just different



The more interesting question is:

How can we best co-exist with the 'perceptions' of machines?



To answer that question you need to think about:

I. How can we know what goes on in the ‘head’ of a machine?

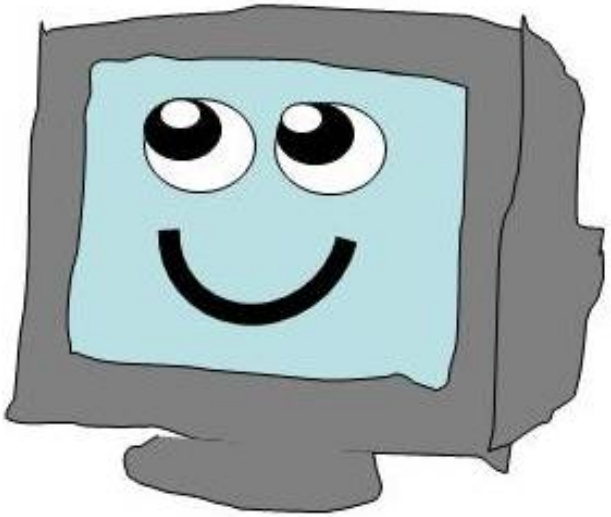
II. What is a ‘good’ machine differentiation?

I. How can we know what goes on in the 'head' of a machine?

II. What is a 'good' machine differentiation?

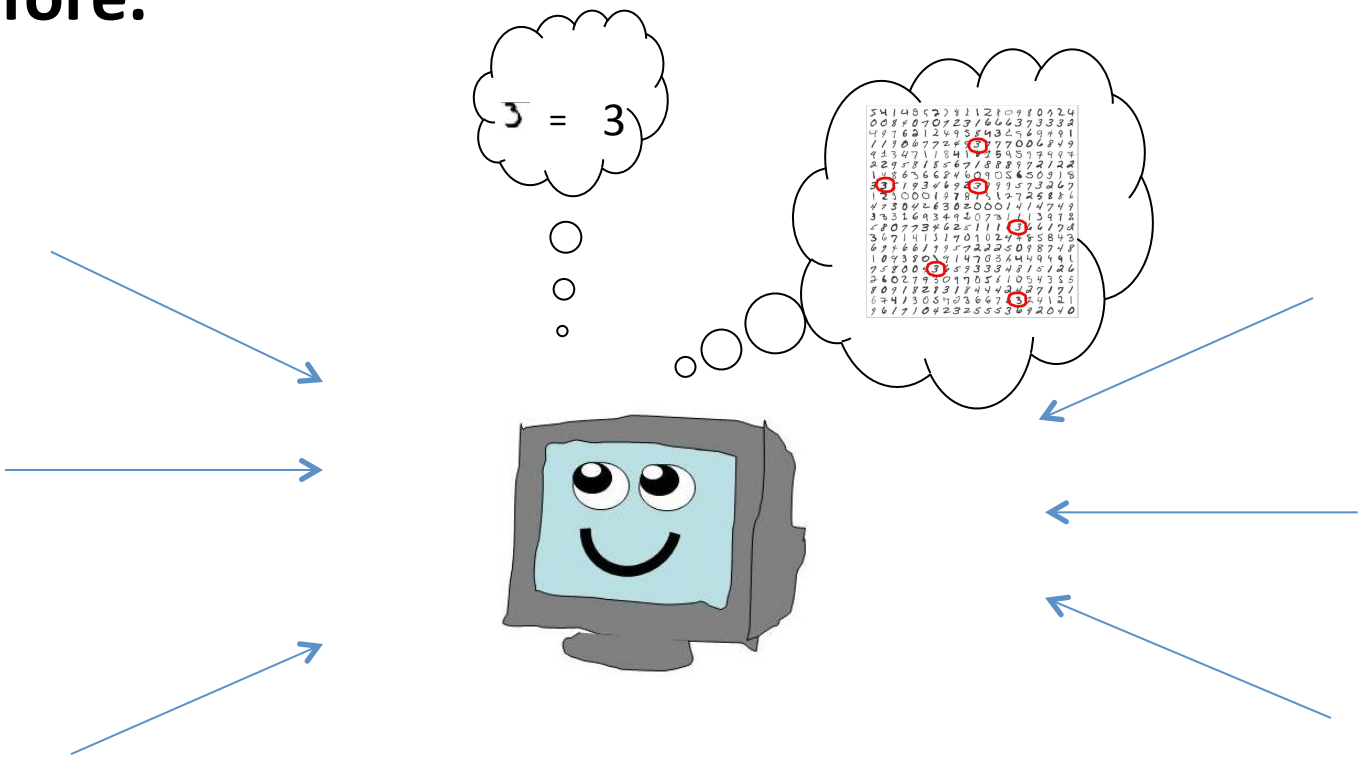
You will have to open up the black-box of machine differentiations

Computer says ‘no’!
I predict you will be
a bad customer.

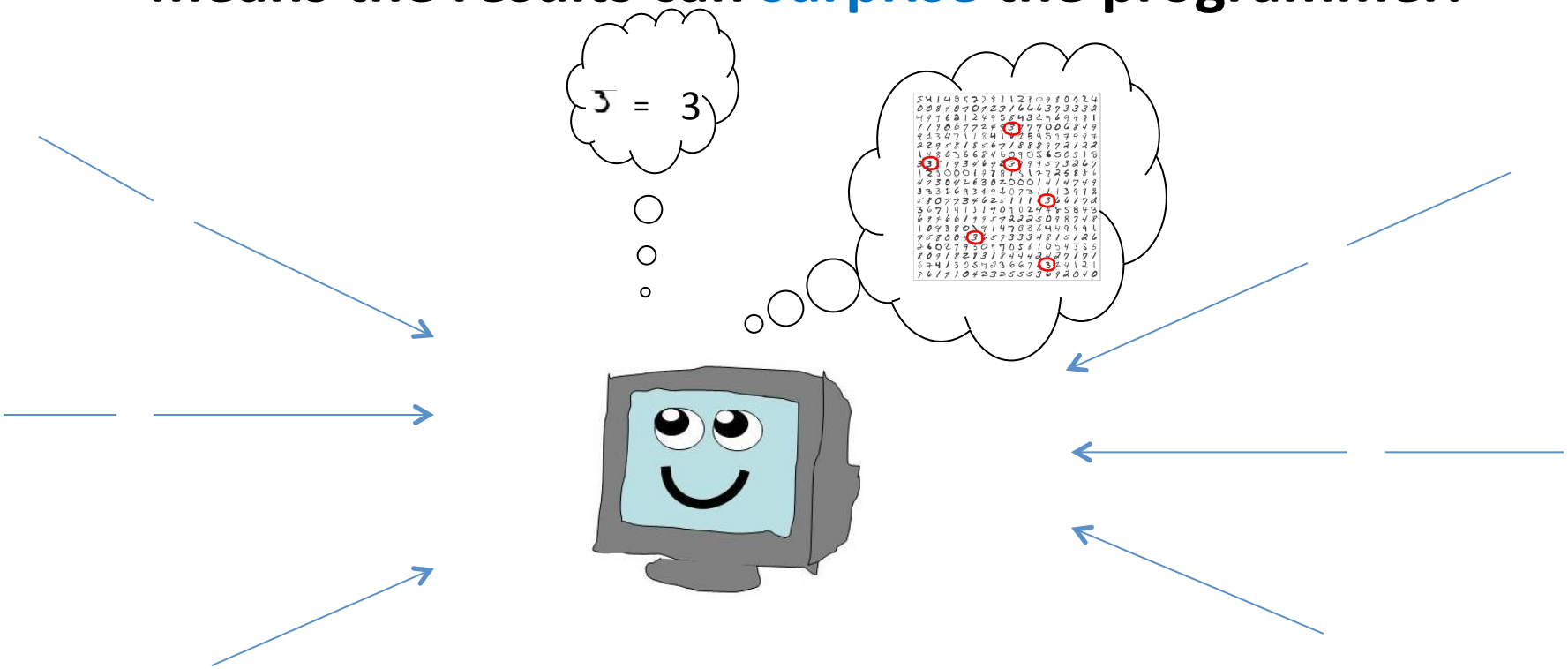


Why????

(1) A machine never just objectively-neutrally represents reality. A lot of human-all-too-human decisions, intentions, stakes and mistakes are folded into it – and that needs to be brought to the fore.

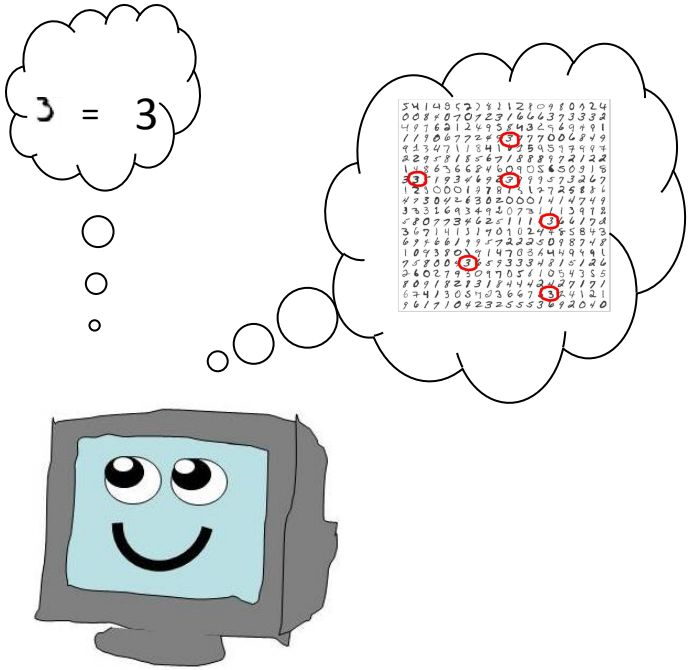


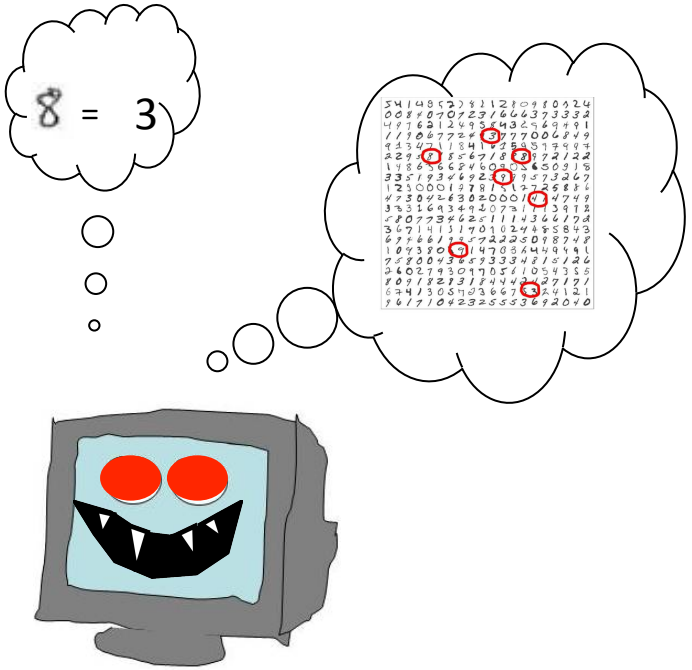
(2) Yet, unfolding all that went in is not enough: machine learning is an indirect way of programming (*instructing a machine with a rule on how to derive a model/rule from examples*) which means the results can surprise the programmer.



I. How can we know what goes on in the ‘head’ of a machine?

II. What is a ‘good’ machine differentiation?

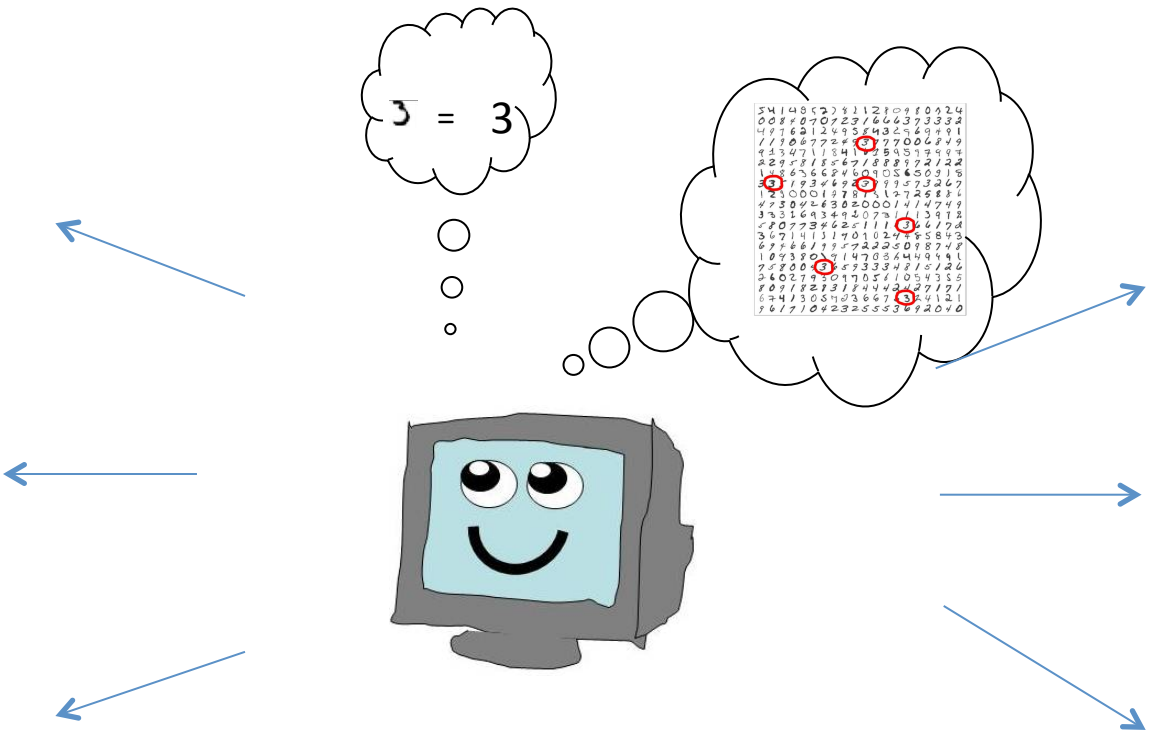

3 = 3


8 = 3

There are many concerns to take into account when you want to decide whether machine differentiations are 'good' or 'bad'.

It's not just about the amount of correct predictions on a standardized database. The machine differentiations have to 'work' within a real, living world.

What is the world the machine differentiations creates? Is that a world we want to live in?



INTERDISCIPLINARY STUDIES OF LAW [METAJURIDICA]
VRIJE UNIVERSITEIT BRUSSEL

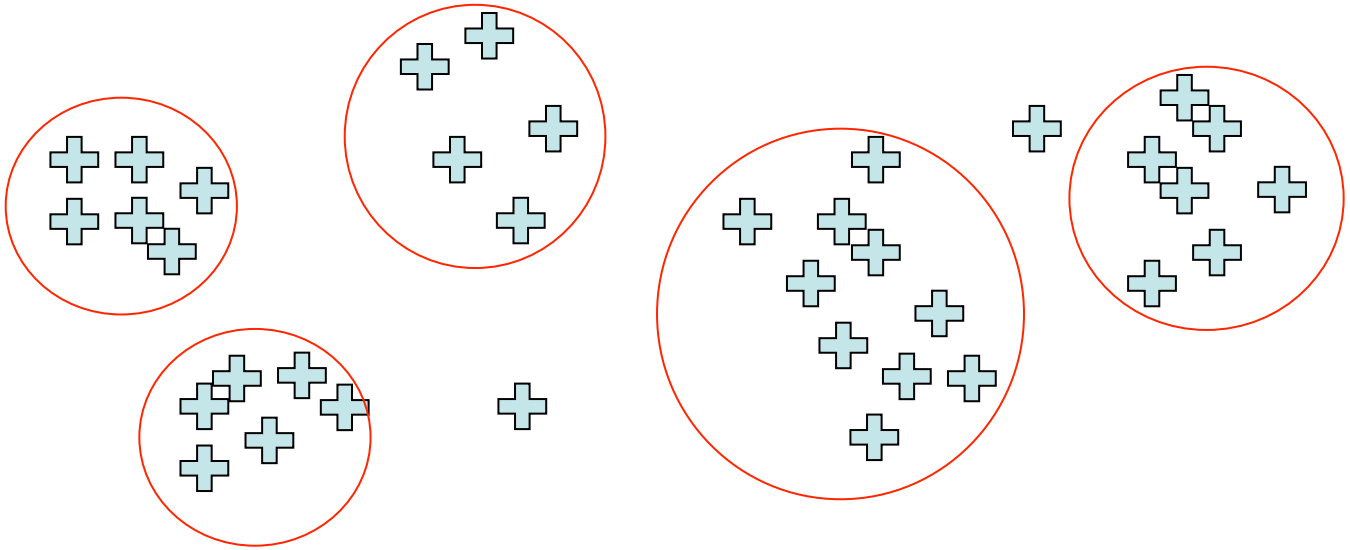
LSTS
LAW, SCIENCE, TECHNOLOGY & SOCIETY STUDIES
VRIJE UNIVERSITEIT BRUSSEL
BELGIUM

iCIS | Digital Security

Radboud Universiteit Nijmegen

Some legal conundrums in relation to machines that produce differences and similarities.

Algorithmic transparency and the thin line between justified differentiations /unjustified discriminations.



Katja de Vries (iCIS, RUN and LSTS, VUB)

26 May 2016 – ETHOS, ITU, Copenhagen

1. Machine learning
2. Some examples
3. Privacy
4. Data Protection law,
5. Anti-discrimnation law
6. Transparency in practice
7. Accountability & discrimination

1. Machine learning
2. Some examples
3. Privacy
4. Data Protection law,
5. Anti-discrimnation law
6. Transparency in practice
7. Accountability & discrimination

“Programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort”.

(Samuel, 1959)

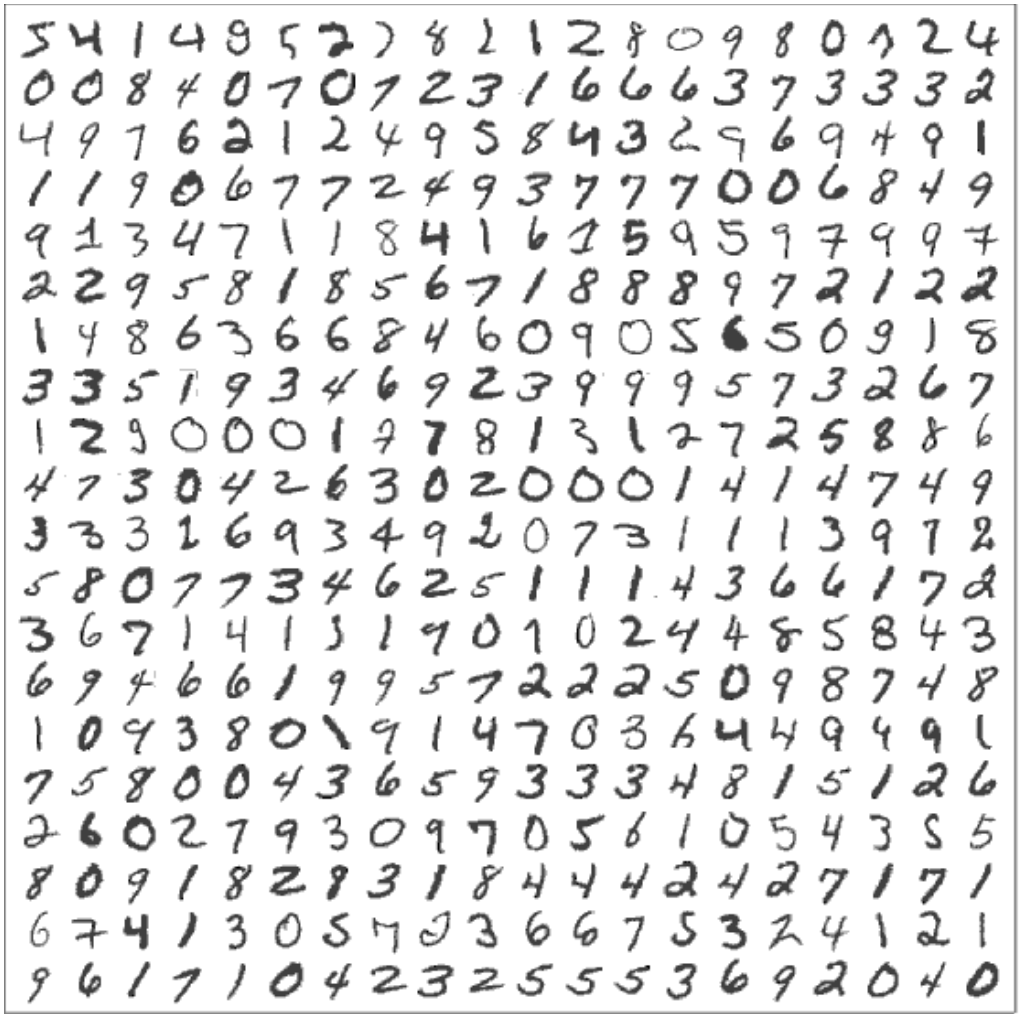
Classical programming:

Explicit instructions (“*it is a dog if conditions x, y and z are fulfilled*”)

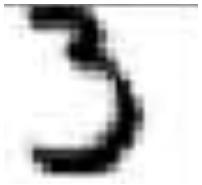
Programming a machine learning algorithm is to give to a machine:

(1) examples
(2) instructions how to extract ‘patterns’, a ‘rule’ or a ‘model’ from these examples
(3) sometimes: feedback (‘you classified this wrongly’)

Unlabelled examples (input *without* the desired output)



Labelled examples (input labelled with the desired output)

 = 3

 = dog

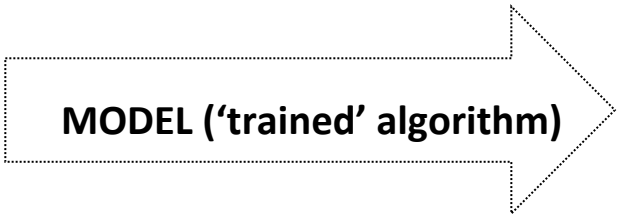
 = bad customer, good employee

Machine learning algorithm

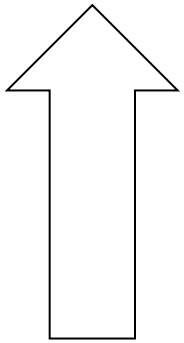
indirect

- examples
- feedback
- general principles about how to discover patterns

A precise recipe that specifies the exact sequence of steps required to solve a problem (MacCormick, 2012).



3



general instructions (‘untrained’ algorithm) on how to discover structures and patterns

(+ examples)

Some of the human-all-too-human elements folded into the machine learning model:

(a) How “true” is the “ground truth”? (labelled examples)

(b) How is the optimal output defined?

As long as you can define what it means to do better or worse on a task*

you can optimize

* This is called the “objective function” or “cost function”.

Clustering intuition:

Minimize the variation within cluster, maximize the variation between clusters

K-means

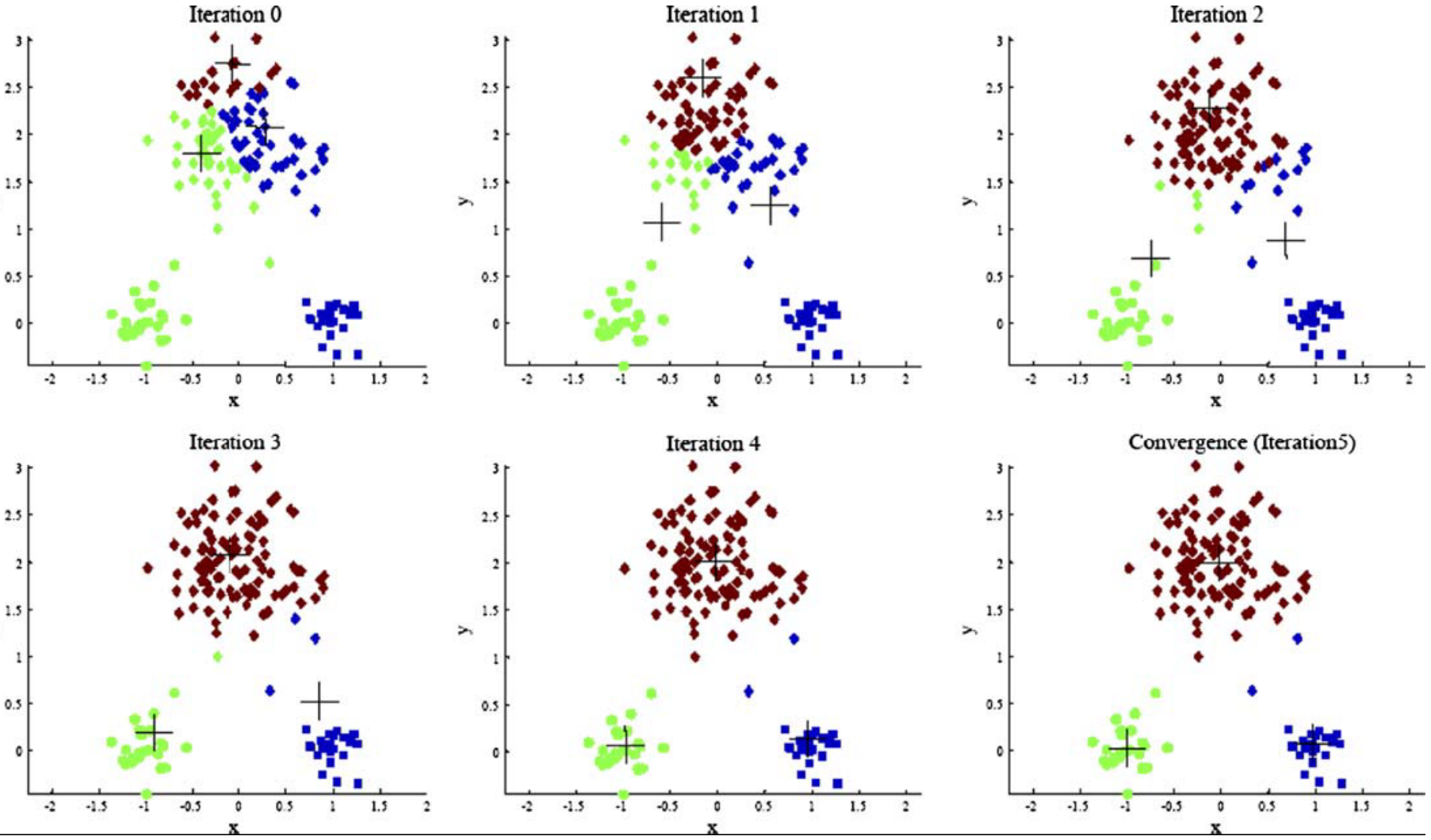
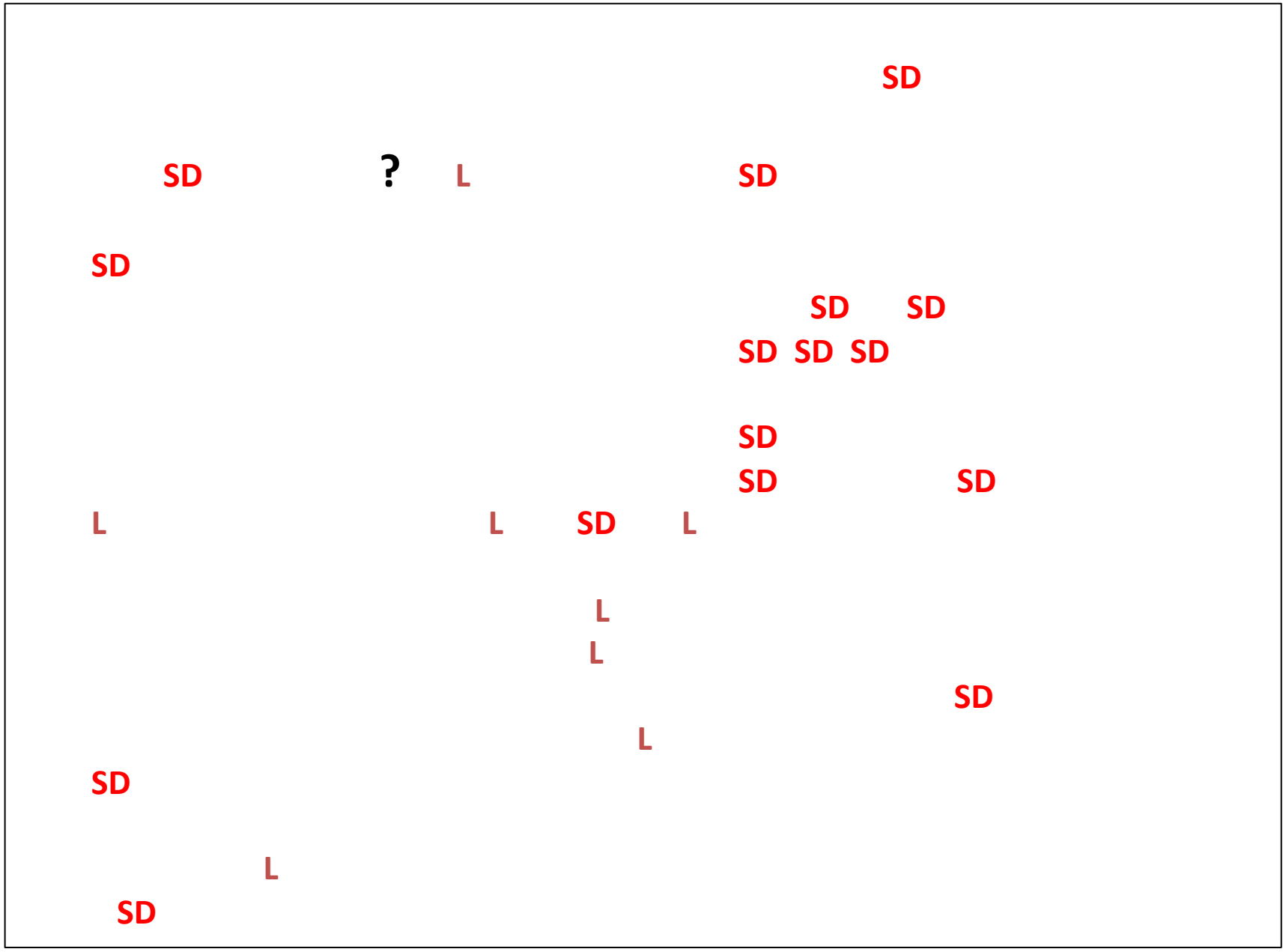


Image source: Wu *et al* (2008). Top 10 algorithms in data mining, *Knowl Inf Syst 14*:p. 1–37


SD
SD
?
L
SD
SD
SD
SD
SD
SD
SD
SD
SD
L
L
SD
L
L
L
SD
L
SD
L
SD

(c) Construct validity of the dependent variable:

how is the dependent variable (‘good employee’, ‘bad customer’, etc.) operationalized?

(d) How representative is the set of training examples?

(e) What relation between newer and older examples is desirable?

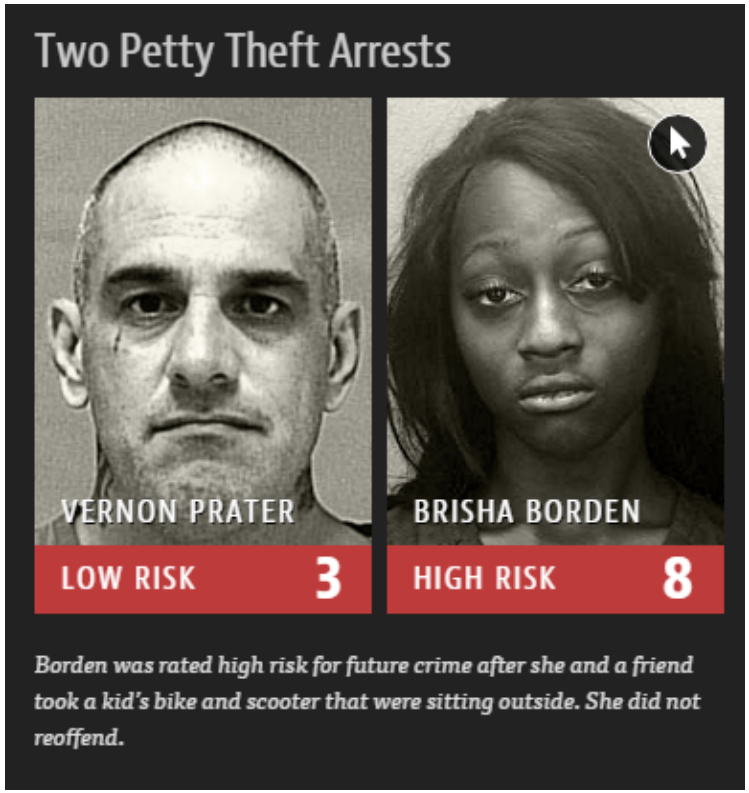
(e) What type of algorithm is chosen?

1. Machine learning
2. Some examples
3. Privacy
4. Data Protection law,
5. Anti-discrimnation law
6. Transparency in practice
7. Accountability & discrimination

Is there something (ethically? legally?) 'wrong' with the following examples of machine learning applications?

Example 1

Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks.



Source: **Machine Bias.** There's software used across the country to predict future criminals. And it's biased against blacks. *By Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica*May 23, 2016
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

"Northpointe's software is among the most widely used assessment tools in the country. The company does not publicly disclose the calculations used to arrive at defendants' risk scores, so it is not possible for either defendants or the public to see what might be driving the disparity. (On Sunday, Northpointe gave ProPublica the basics of its future-crime formula — which includes factors such as education levels, and whether a defendant has a job. It did not share the specific calculations, which it said are proprietary.)"

“Northpointe’s core product is a set of scores derived from 137 questions that are either answered by defendants or pulled from criminal records. Race is not one of the questions. The survey asks defendants such things as: “Was one of your parents ever sent to jail or prison?” “How many of your friends/acquaintances are taking drugs illegally?” and “How often did you get in fights while at school?” The questionnaire also asks people to agree or disagree with statements such as “A hungry person has a right to steal” and “If people make me angry or lose my temper, I can be dangerous.””

Example 2

In some cities, Amazon Prime same-day delivery serves white ZIP codes better than black ones.

Source: http://www.bloomberg.com/graphics/2016-amazon-same-day/ and http://bloom.bg/1UZ7cYj

Example 3

Klarna and Wonga – deciding who can 'pay later?'

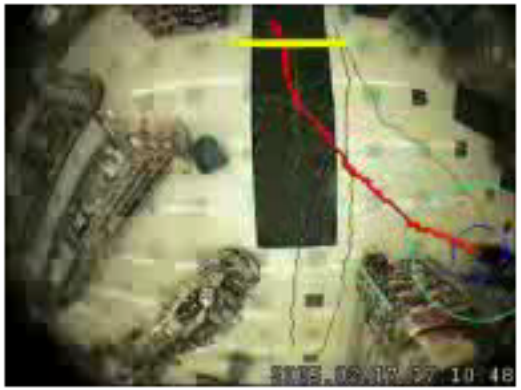
Examples of factors which are taken into consideration by the algorithms:

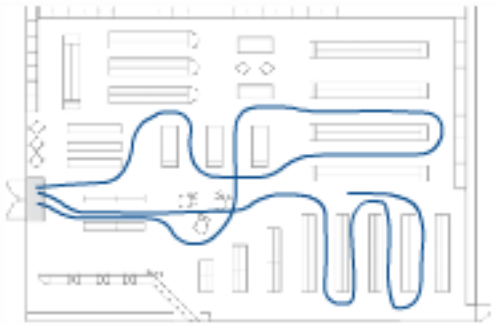
- time of purchase
- whether the consumer's name and address were typed or copied in (the latter is more likely to signal fraud).
- having a mobile phone with a contract

http://www.economist.com/node/18396166 and
http://www.economist.com/news/finance-and-economics/21690106-payments-unicorn-seeks-become-dray-horse-bank-getting-more-ambitious

Example 4

Examples of VideoMining Technologies in Action





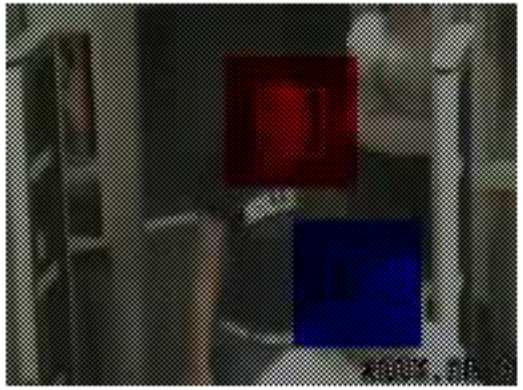


Image source: *VideoMining Corporation*

Categorization

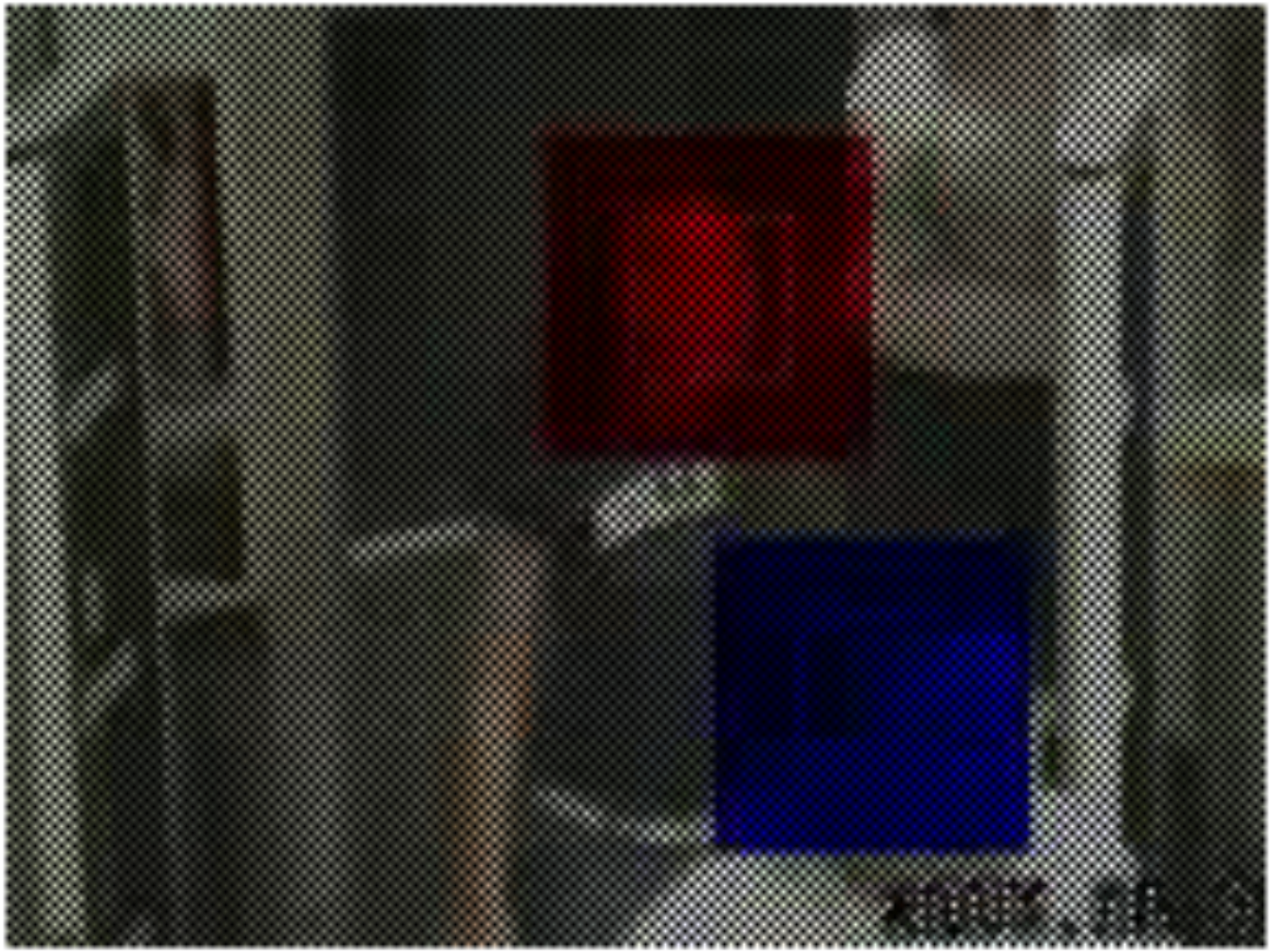


Image source: *VideoMining Corporation*)

Example 5

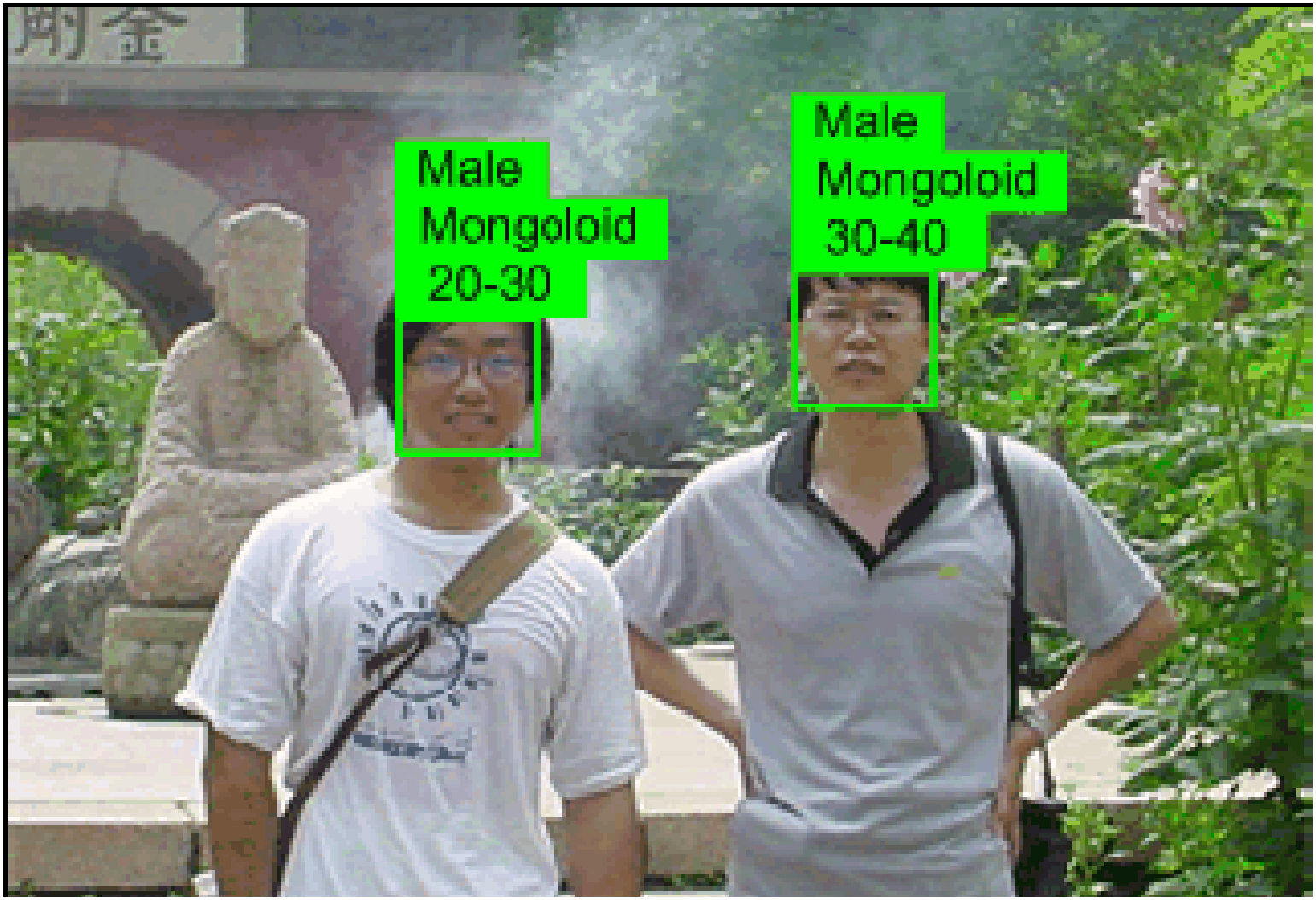


Image source: Bo Wu, Haizhou Ai, Chang Huang, "Facial Image Retrieval Based on Demographic Classification," *icpr*, vol. 3, pp.914-917, 17th International Conference on Pattern Recognition (ICPR'04) - Volume 3, 2004
Online available at:
http://media.cs.tsinghua.edu.cn/~imagevision/papers/204_WU_B.pdf

Example 6

https://www.youtube.com/watch?v=hGKCD0dUxOA

Example 7

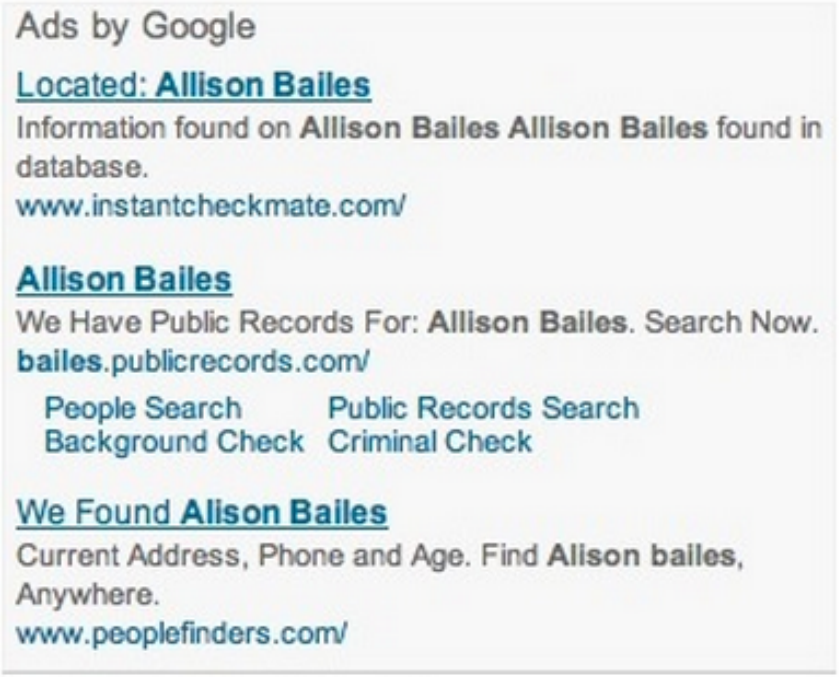
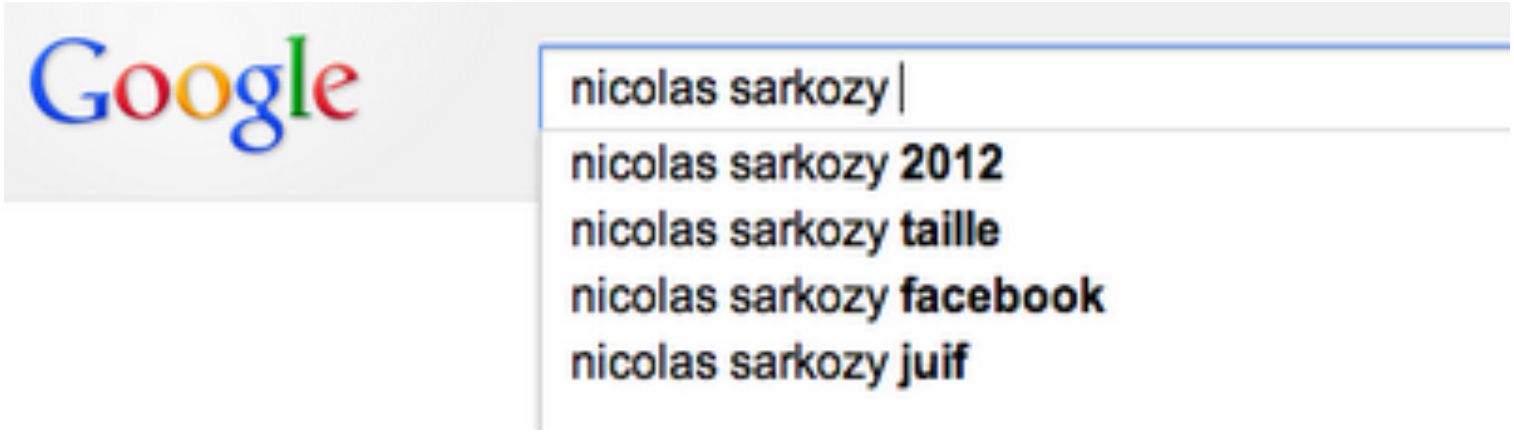


Image source: Latanya Sweeney (2012). *Discrimination in Online Ad Delivery*

Example 8


Google
France
carla bruni est
carla bruni est elle juive
carla bruni est une prostituée
carla bruni est moche
carla bruni est elle enceinte
carla bruni est elle refaite
En savoir plus


Google
nicolas sarkozy
nicolas sarkozy 2012
nicolas sarkozy taille
nicolas sarkozy facebook
nicolas sarkozy juif


Google
francois ho
françois hollande
françois hollande juif
françois hollande 2012
françois hollande dijon
Recherche

Example 9

Example 10

Sorting CV's and job applications

Source: https://www.erim.eur.nl/research/news/detail/3797-phd-defence-colin-lee/ and http://www.nrc.nl/next/2016/04/06/het-algoritme-zegt-geschikt-1605960

Some (extra-legal) problems

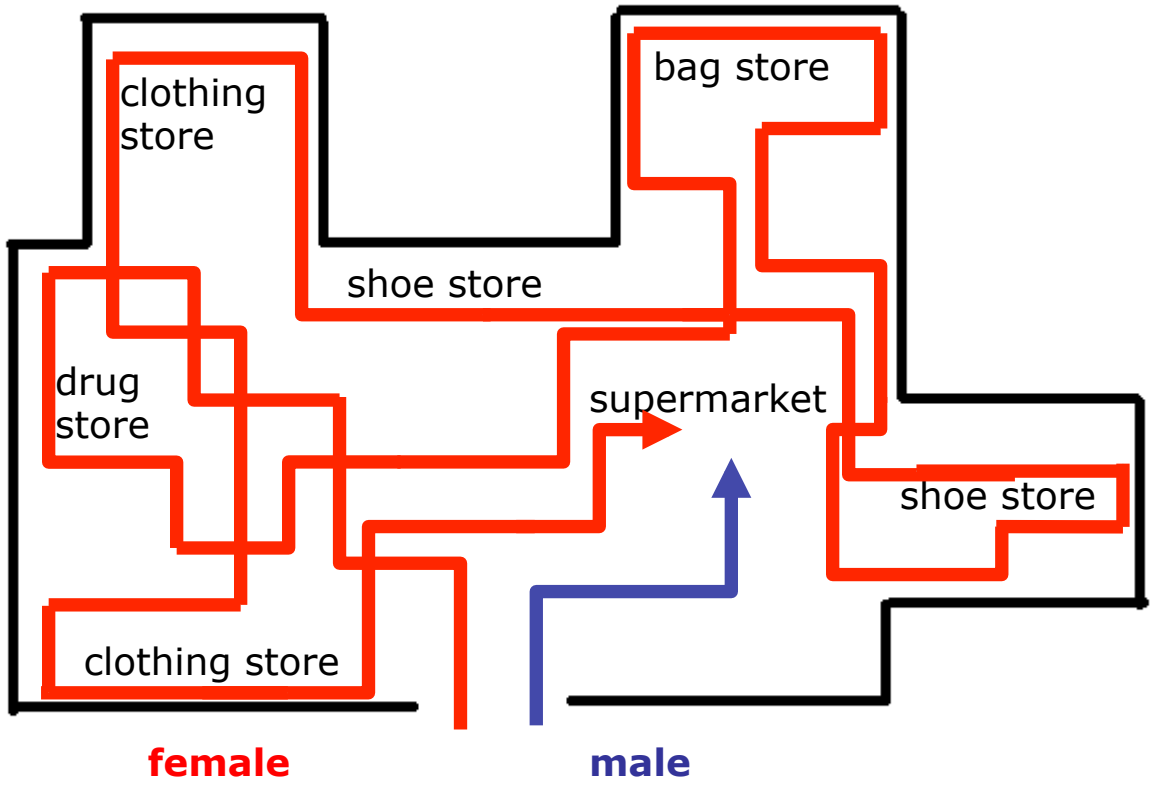
Assignment:
Go to the shopping centre and buy one carton of eggs

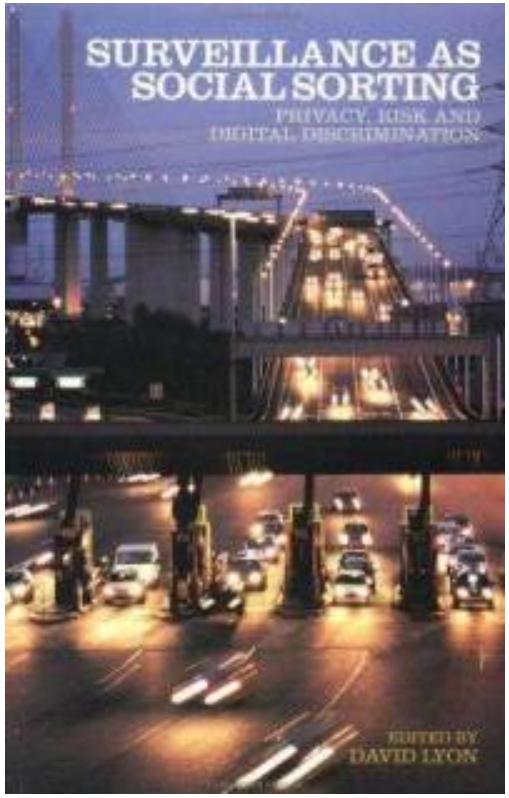
Result woman

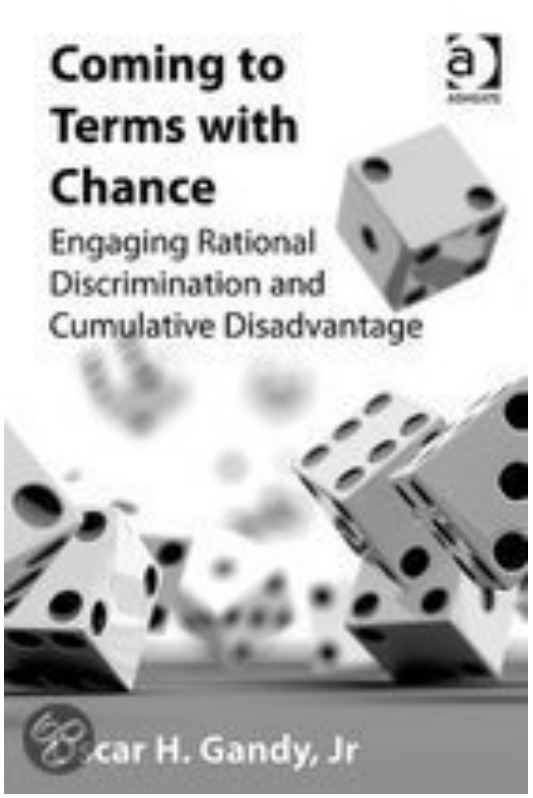
Price total: 2067 SEK

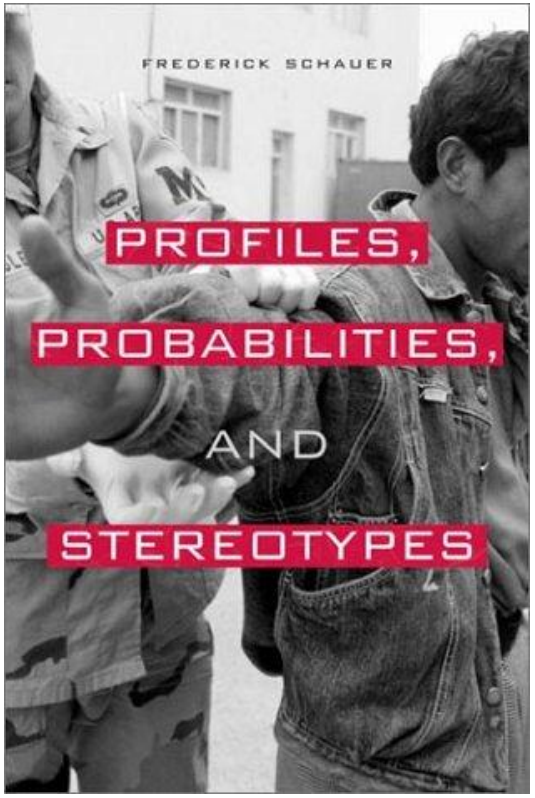
Result man

Price total: 12 SEK




SURVEILLANCE AS SOCIAL SORTING
PRIVACY, RISK AND DIGITAL DISCRIMINATION
EDITED BY DAVID LYON


Coming to Terms with Chance
Engaging Rational Discrimination and Cumulative Disadvantage
Oscar H. Gandy, Jr


FREDERICK SCHAUER
PROFILES,
PROBABILITIES,
AND
STEREOTYPES


Republic.com 2.0
Revenge of the Blogs
Cass R. Sunstein


The Filter Bubble
What
the
Internet
Is
Hiding
From
You
Eli Pariser

Image source: image*after

Queries in autocomplete are algorithmically determined based on a number of factors (including search term popularity) without manual intervention.

http://support.google.com/websearch/bin/answer.py?hl=en&answer=106230

1. Machine learning
2. Some examples
3. Privacy
4. Data Protection law,
5. Anti-discrimnation law
6. Transparency in practice
7. Accountability & discrimination

Art. 8 ECHR

Art. 8 of the *European Convention on Human Rights* (ECHR)
Right to respect for private and family life

Everyone has the right to respect for his private and family life, his home and his correspondence.

There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

1. Machine learning
2. Some examples
3. Privacy
4. Data Protection law
5. Anti-discrimnation law
6. Transparency in practice
7. Accountability & discrimination

Current data protection law: Directive 95/46/EC

24 May 2016: new law (General Data Protection Regulation 2016/679) entered into force. Shall apply from **25 May 2018**.

http://ec.europa.eu/justice/data-protection/reform/index_en.htm

New Data Protection Law (replacing Directive 95/46):
REGULATION (EU) 2016/679

http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=RO

The core actants and basic principles stay the same…..

Source: http://upload.wikimedia.org/wikipedia/commons/2/20/IBM_Electronic_Data_Processing_Machine_-_GPN-2000-001881.jpg

Art 4 (1) GDPR. Definitions.

'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

Art 4 (7) GDPR. Definitions.

'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data;

Art 5 GDPR. Principles relating to processing of personal data

Personal data shall be:

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

(b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes (‘purpose limitation’);

(c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (‘data minimisation’);

(d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay (‘accuracy’);

(e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');
(f) processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

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

‘Article 9 GDPR. Processing of special categories of personal data

1.Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.

2.Paragraph 1 shall not apply if one of the following applies: (a) the data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject;

Some of the new stuff in the GDPR

- Expanded territorial reach

- Accountability and privacy by design: this also includes lots of Data Protection Officers and Data Protection Impact Assessments

- More bite: (1) consent, (2) sanctions

- Not so very new but more explicit: right to be forgotten

Relevant for machine learning applied to humans:

- Profiling

- More attention to (discriminatory) consequences of profiling ("..*the significance and the envisaged consequences of such processing* ")

Article 4(4) GDPR. Definitions

'Profiling' means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements

Transparency

Articles 13(2f) and 14(2g) GDPR (Information to be provided to the datasubject):

the controller shall […] provide the data subject with the following […] information [….]:

- the **existence of automated decision-making, including profiling**, […] and, at least in those cases,
- **meaningful information about the logic involved**,
- as well as **the significance and the envisaged consequences** of such processing for the data subject.

** Also see 15(1h) GDPR (Right of acces by data subject)*

Transparency

Recital 60 GDPR.

… The principles of fair and transparent processing require that the data subject be informed of the existence of the processing operation and its purposes. The controller should provide the data subject with any further information necessary to ensure fair and transparent processing taking into account the specific circumstances and context in which the personal data are processed. Furthermore, the data subject should be **informed of the existence of profiling** and **the consequences of such profiling**. Where the personal data are collected from the data subject, the data subject should also be informed whether he or she is obliged to provide the personal data and of the consequences, where he or she does not provide such data. **That information may be provided in combination with standardised icons in order to give in an easily visible, intelligible and clearly legible manner, a meaningful overview of the intended processing**. Where the icons are presented electronically, they should be machine-readable.

Article 22 GDPR.
Automated individual decision-making, including profiling

1.The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

2.Paragraph 1 shall not apply if the decision: (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller; (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or (c) is based on the data subject's explicit consent.

3.In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

4.Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

Art 6(1)(e) and (f). Right to object

(e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
(f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

Article 21 **Right to object** 1.The data subject shall have the right to object, on grounds relating to his or her particular situation, at any time to processing of personal data concerning him or her which is based on point (e) or (f) of Article 6(1), including profiling based on those provisions. The controller shall no longer process the personal data unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims. 2.Where personal data are processed for direct marketing purposes, the data subject shall have the right to object at any time to processing of personal data concerning him or her for such marketing, which includes profiling to the extent that it is related to such direct marketing. 3.Where the data subject objects to processing for direct marketing purposes, the personal data shall no longer be processed for such purposes.

1. Machine learning
2. Some examples
3. Privacy
4. Data Protection law,
5. Anti-discrimination law
6. Transparency in practice
7. Accountability & discrimination

Antidiscrimination (art. 14 ECHR: *Prohibition of discrimination with regard to*
the exercise other human rights) [an "ancillary" right, but less so since the introduction of protocol 12 in 2009]

Art. 14 ECHR *Prohibition of discrimination*
The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.

EU Anti-Discrimination Law

- Treaty level:
  - Title III of the EU Charter of Fundamental Rights
  - art. 18-25 *Treaty on the Functioning of the European Union*

- Directive level:
  - 2000/43/EC: race in employment, social welfare, access to goods & services
  - 2000/78/EC: religion, belief, age, disability, sexual orientation in employment.
  - 2006/54/EC: gender in employment
  - 2004/113/EC: gender in access G&S.
  - CFD 2008/913/JHA: on combating certain forms and expressions of racism and xenophobia by means of criminal

The fact that a differentiation is based on "relevant and accurate actuarial and statistical data“ does not make it a legitimate ground for discrimination.

ECJ, *Test Achats v. Council*, C-236/09, Judgment of 1 March 2011.

- life insurance
- car insurance

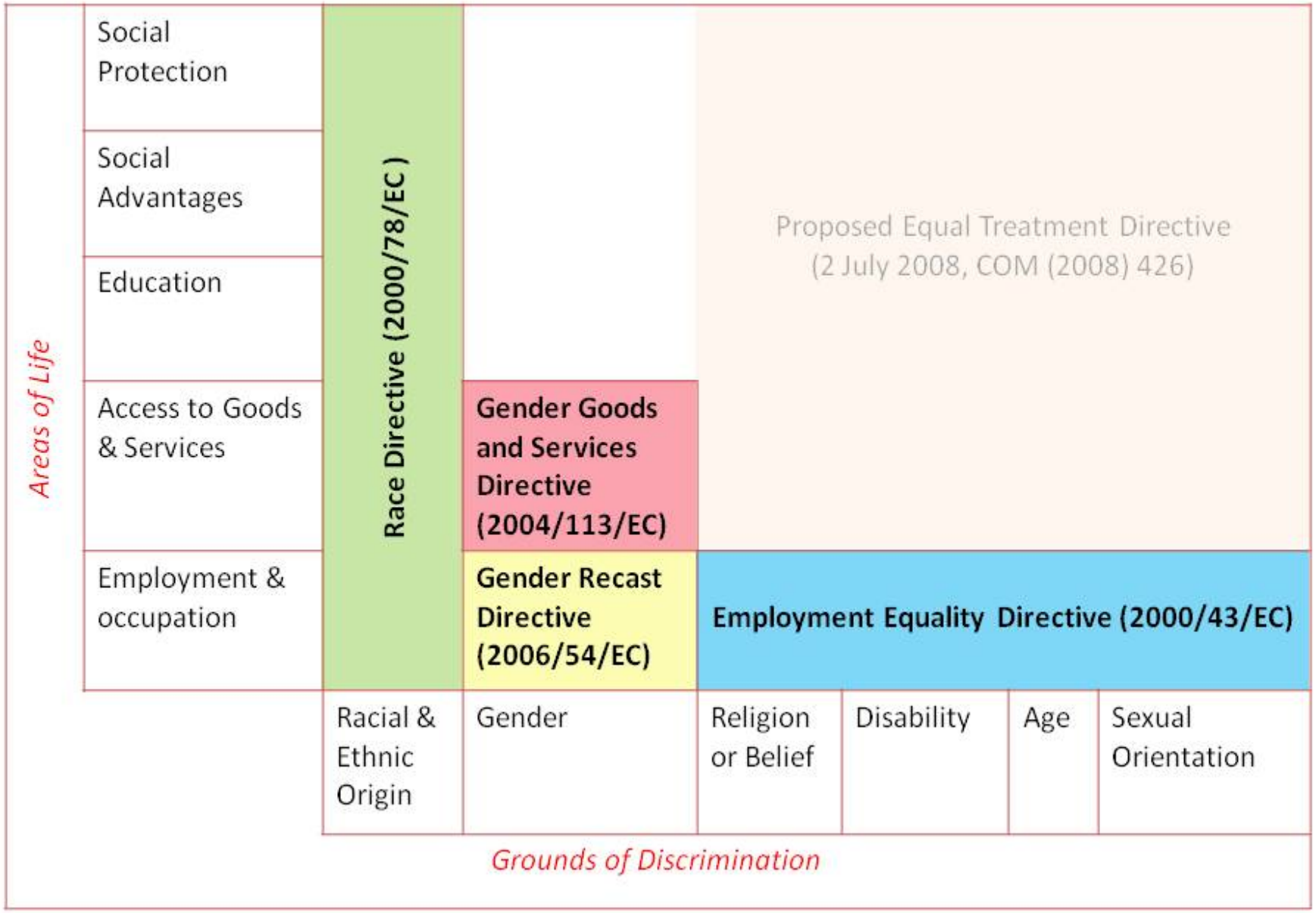

Areas of Life
Social Protection
Social Advantages
Education
Access to Goods & Services
Employment & occupation
Race Directive (2000/78/EC)
Gender Goods and Services Directive (2004/113/EC)
Gender Recast Directive (2006/54/EC)
Proposed Equal Treatment Directive
(2 July 2008, COM (2008) 426)
Employment Equality Directive (2000/43/EC)
Racial & Ethnic Origin
Gender
Religion or Belief
Disability
Age
Sexual Orientation
Grounds of Discrimination

Image source: http://manipuronline.com

Direct discrimination*:

Based on a protected ground (racial or ethnic, sex, religion or belief, disability, age or sexual orientation), one person is treated less favorably than another is/has been/would be treated in a comparable situation.

See: Directives 2000/43; 2000/78 art.2

Indirect discrimination*:

Where **an apparently neutral provision**, criterion or practice would put persons having a particular racial or ethnic origin, religion or belief, sex, disability, age or sexual orientation at a particular disadvantage compared with other people

Except if: this provision, criterion or practice is **objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary**

Can you show it's a **proportionate means** of **achieving** a **legitimate aim**?

See: Directives 2000/43; 2000/78 art.2

1. Machine learning
2. Some examples
3. Privacy
4. Data Protection law,
5. Anti-discrimnation law
6. Transparency in practice
7. Accountability & discrimination

Types of transparency the data controller can provide

Algorithmic transparency:

- the algorithm is easy interpretable (e.g. a decision tree or a straightforward linear function)
- the algorithms not easy interpretable (e.g. neural net) but the data controller creates an more easily interpretable model based on the input-output. See e.g.: http://arxiv.org/abs/1602.04938)

Narrative transparency:

- 'telling' how it works (but think e.g. about the Latanya Sweeney case!)

Types of "transparency" third parties can provide to support data subjects in the exercise of their informational rights

Input analysis ('speculative' insights in the possibilities of the state of the art):

- USEMP/DataBait.

The USEMP (http://www.usemp-project.eu/) project which will result in a transparency tool that shows users of social networks which (commercially interesting) information can be derived from their data (http://databait.eu).

Input-output matching:

- - Input-output matching ('blackbox testing') like you have in Xray (http://xray.cs.columbia.edu/) , AdFisher (https://www.cs.cmu.edu/~mtschant/ife/), or SunLight (https://columbia.github.io/sunlight/ or www.cs.columbia.edu/~djhsu/papers/sunlight.pdf) .

https://databait.hwcomms.com

DataBait

Language

The quest to reclaim online data

In order to participate in the DataBait Research, you will have to register for a DataBait account. This is necessary in order to sign the digital contract and to install a connection to your Facebook account (It is necessary to have a **Facebook account** to participate to Databait).

First time on DataBait?

Databait: what, why, how

Register Now

If you already have a Databait account please provide your details below and Sign In

Username (your registered email)

Password

Sign In

Forgot Password

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no611596

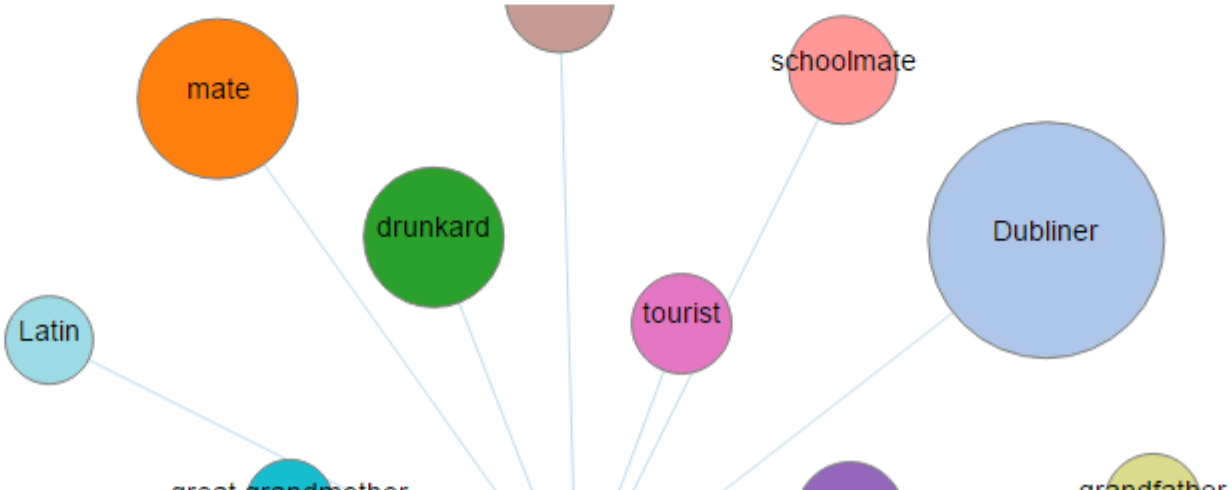
My Disclosures

Overview Location Insights Photo Insights Brands Insights

Click on concept to see the related photos.

buddy Dubliner mate lover drunkard
Byelorussian employer schoolmate flatm

Top 20 Visual Concepts Detected



1. Machine learning
2. Some examples
3. Privacy
4. Data Protection law,
5. Anti-discrimnation law
6. Transparency in practice
7. Accountability & discrimination

- Ex-ante
- Ex-post