



# Cross Cutting Issues in Horizon Europe: Ethics and Integrity

Vincent C. Müller

TU Eindhoven  
[www.sophia.de](http://www.sophia.de)  
12/2021

# ... what we don't do

Tuesday December 22, 2020 *Ethics Class*



DILBERT.COM @SCOTTADAMSSAYS



12-22-20 2020 Scott Adams, Inc./Dist. by Andrews McMeel



## ... what we may achieve

- Ability to recognise an **ethical problem** and develop a basic analysis of it (norms & consequences)
- Understand Descriptive vs. normative statements:  
“A **is** B” vs. “A **should be** B”  
“What should I do?”
- Motivated to be a “good person” and a “good engineer”

# Ethics



- Etymology: “mores”, “ethos” (ἠθος) = **customs**, character, ‘what one does’
- **Ethics** (Morals): systematic reflection on the normative
- Basic question of ethics:
  - **“What should I do?”** (I. Kant)
    - What should be my next action, e.g. my next move in this game of chess? (Ethics is part of an advanced theory of rational choice.)

## Descriptive vs. Normative

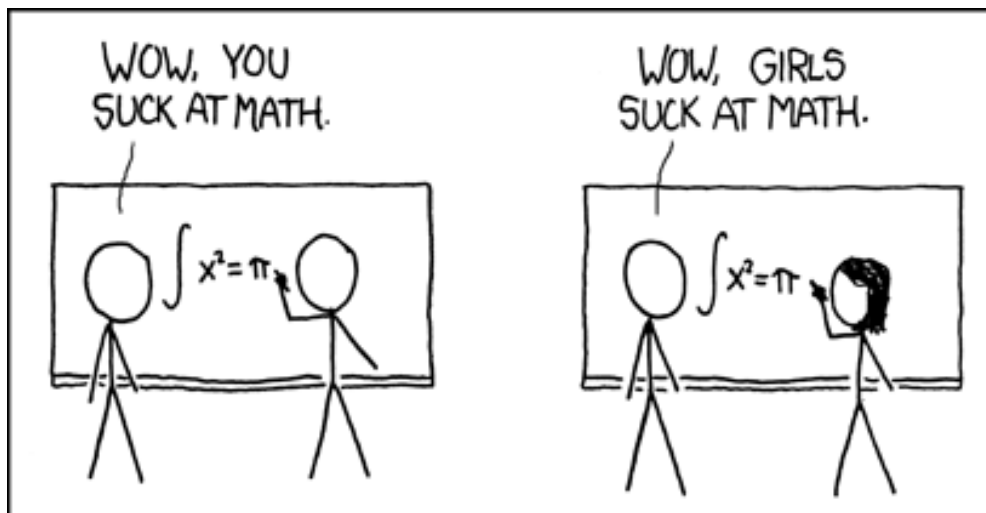
- We now have an unprecedented amount of data.
- Data is likely to change the world.
- People have values and follow to norms,

vs.

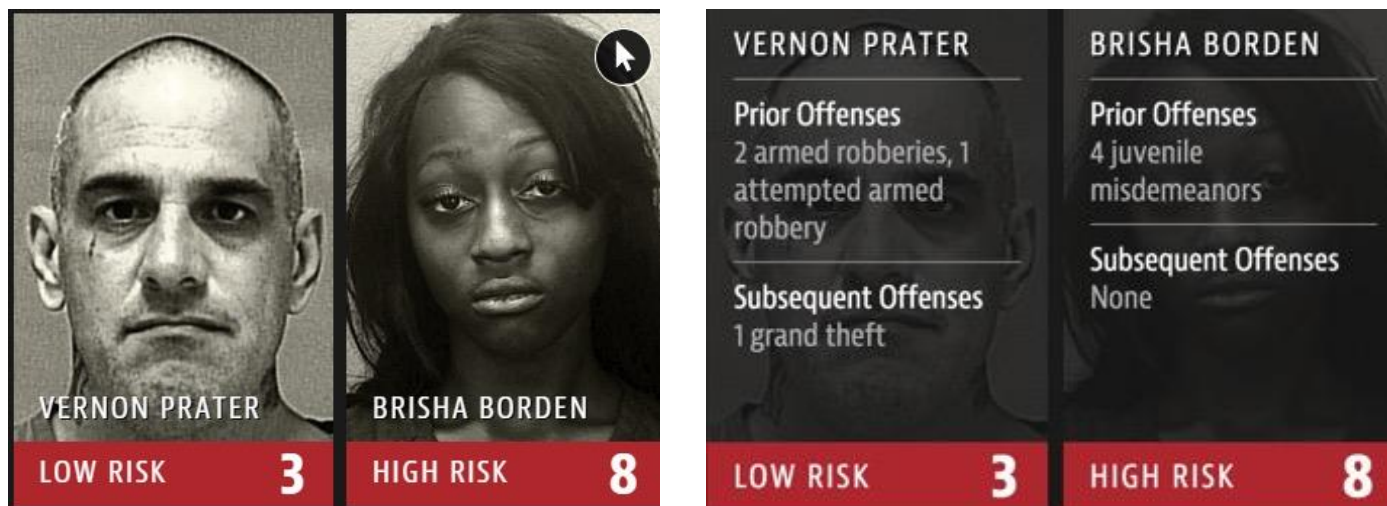
- You should have told me that my girlfriend is cheating on me.
- Survival is more important than privacy.
- An engineer should act in the public interest.

# Bias & AI – Example 1: Decision systems & predictive analytics

- **Bias:** judgments are influenced by a characteristic that is **actually irrelevant** to the matter at hand - typically a discriminatory preconception about members of a group. (Implicit or explicit)
  - **Cognitive bias:** Tendencies for thinking errors, e.g. anchoring, confirmation bias, ... (Kahneman: thinking fast & slow). Bias as heuristics?
    - Hofstadter's Law: "It always takes longer than you expect, even when you take into account Hofstadter's Law."
  - **Statistical bias:** dataset does not represent reality, e.g. selection bias, historical bias, ... [you know]



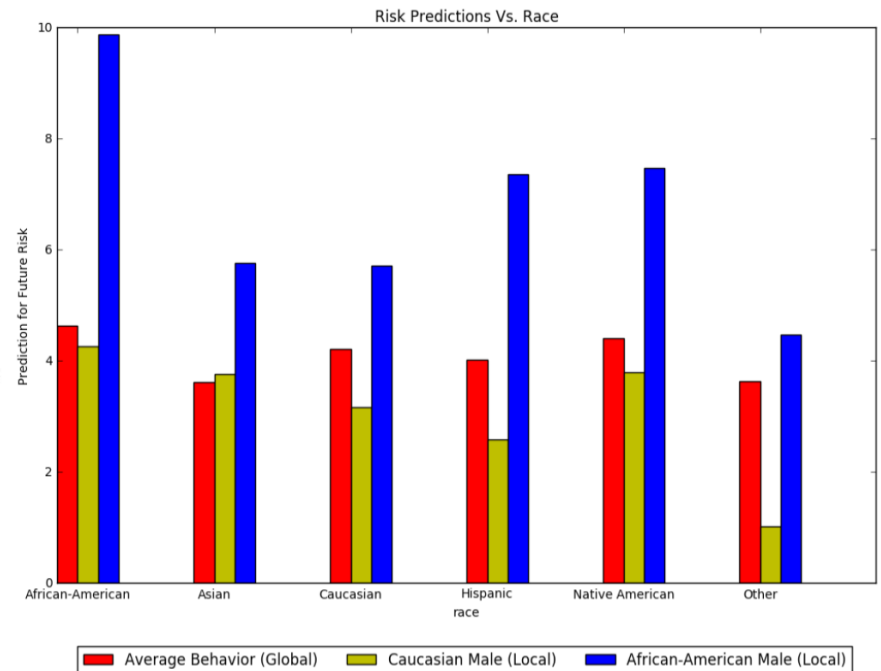
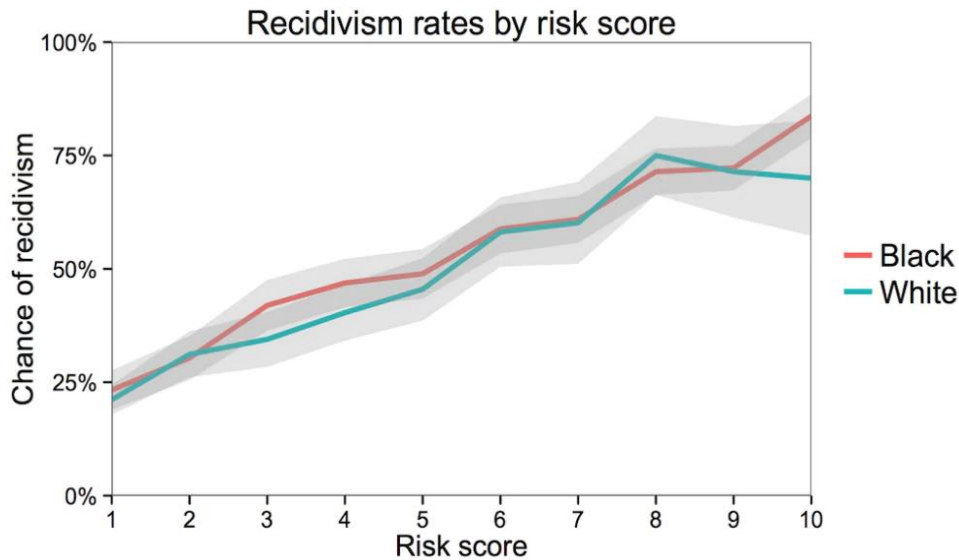
# Example: Northpoint COMPAS Parole Recommendation System



Black offenders were almost **twice** as likely as white offenders to be labelled a higher risk (while the actual re-offending rate was identical)

"There's software used across the country to predict future criminals. And it's biased against blacks", ProPublica 2016

# Northpoint's Response & A Counter-Response



: A technical report: "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity" (2016) - <https://www.equivant.com/response-to-propublica-demonstrating-accuracy-equity-and-predictive-parity/>

Patrick Hall & Navdeep Gill: Debugging the Black-Box COMPAS Risk Assessment Instrument to Diagnose and Remediate Bias (Draft 2017)



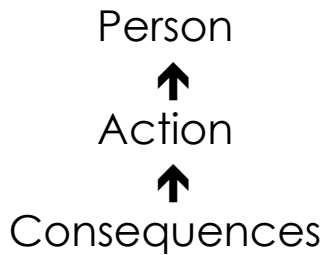
# What would a good engineer do?

- Aims
  - System compliant to legal demands (GDPR)
  - Fair system
  - Transparent system
  - Explainable decisions
  - ...? [you know better]
- How?
  - Select datasets
  - Enable transparent & explainable learning
  - Be aware of cognitive biases
  - Be aware of societal biases
  - ...? [you know better]

“Algorithms are opinions embedded in code”

Cathy O’Neil,  
*The era of blind faith in big data must end*,  
TED Talk, 2017

## 2. Three basic ethical theories



3 types of things to evaluate:

- What kind of person would do this? Positive character traits, *virtues*

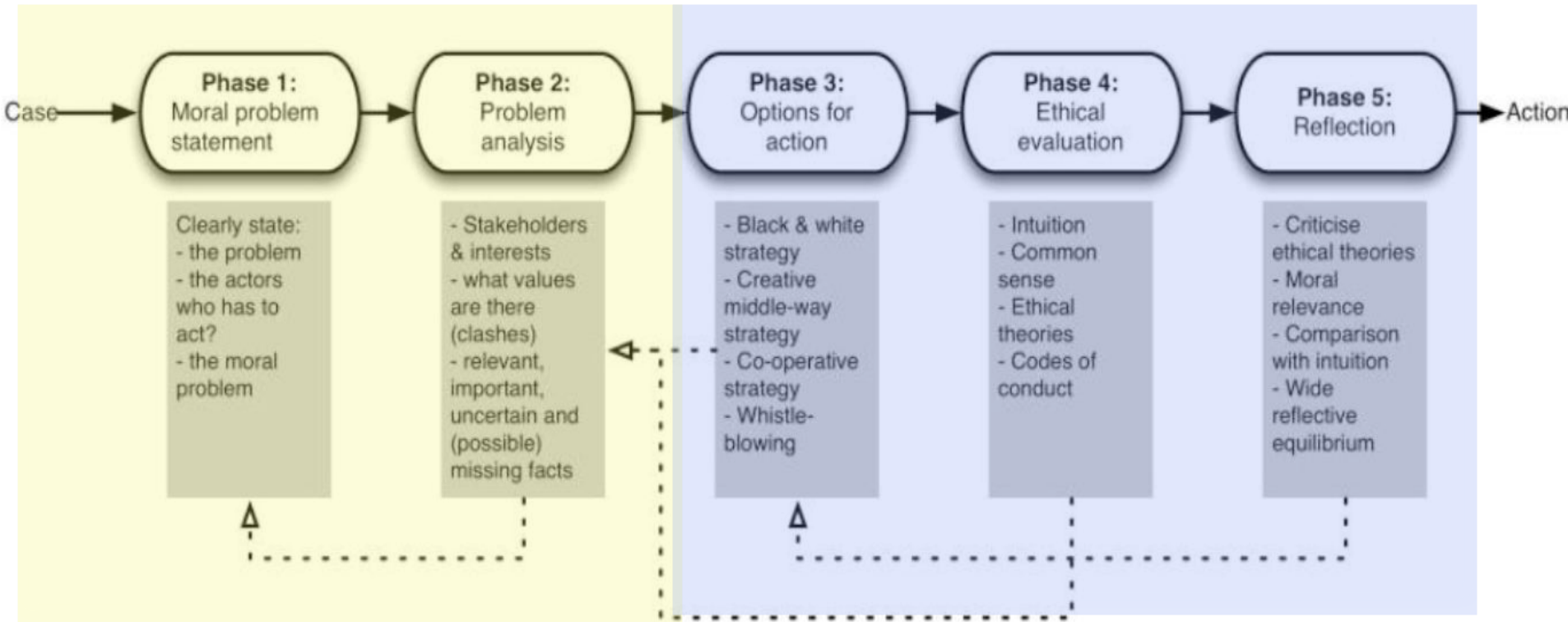


- What kind of action is it? Rules and *norms*



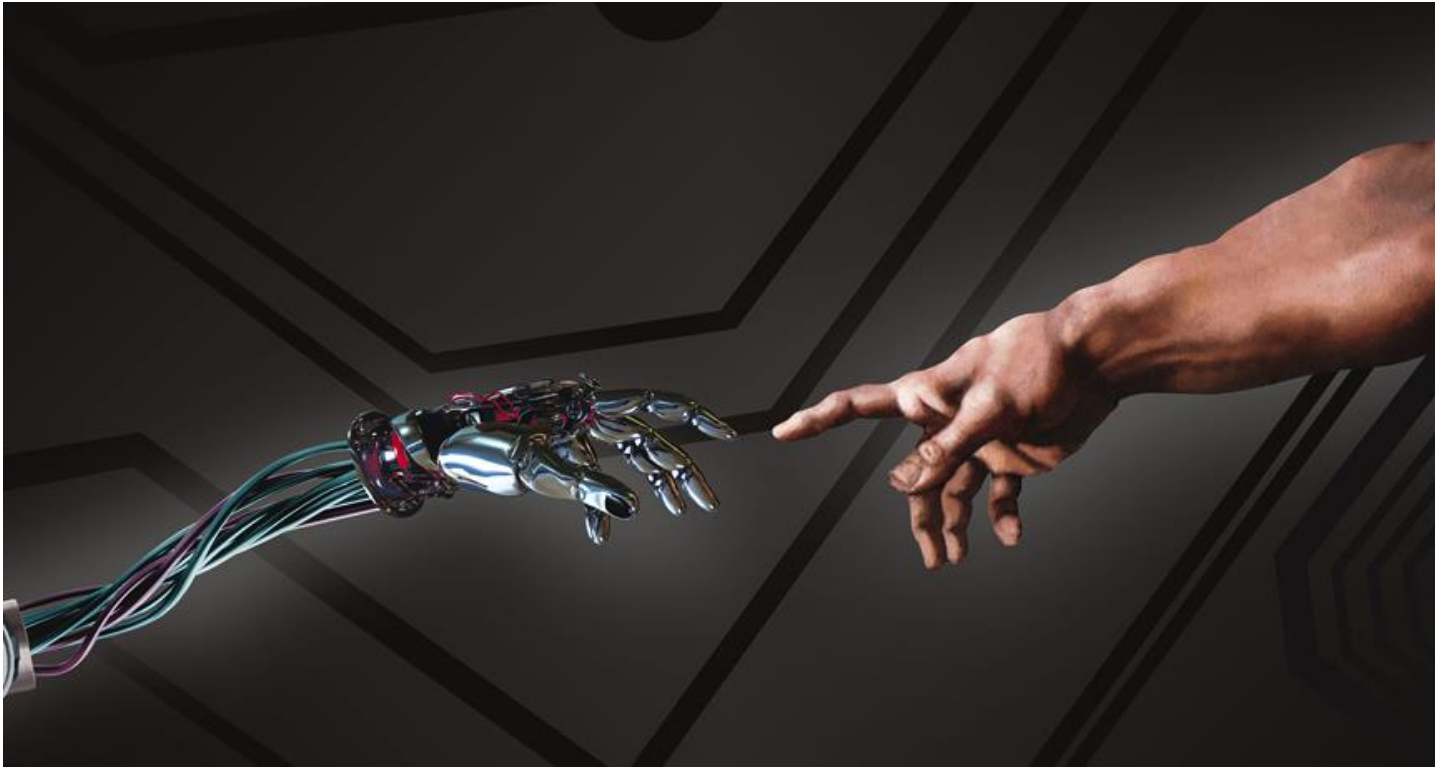
- Consequences of actions—positive and negative, *utility*

# The ethical “cycle”



# Take-home message

1. Ethics: "What should I do?"
  - Normative (not descriptive)
2. Evaluating
  - person → action → consequences
  - virtues → values/norms → utility/consequences
3. Problems in ethics =  
conflicts of virtues/values/utility



Thank You!