Reasons for Reasons: Insights from a Legal Perspective

Katherine Strandburg
New York University School of Law
“Explanation” is a highly contestable concept
- Transparency?
- Interpretability?
- Justification?
- Causation?

Explanations can be of (at least) two types:
- Descriptive or “account giving”: How did decisionmaker X arrive at outcome Y
- Normative justification: Why is outcome Y right?

When should we demand/require “explanations” of predictive analytics results? What makes an “explanation” sufficient in a given case?
- Answers depend on why we explain
Explanation and Law

- Explanation is a core practice in legal decisionmaking (often associated with “due process”), e.g.
  - Judges generally provide either written or oral explanations of their decisions
  - Administrative rulemaking requires that agencies respond to comments on proposed rules
  - Agency adjudicators must provide reasons for their decisions to facilitate judicial review

- But explanations are not always required:
  - Jury decisions – made by “peers”
  - Legislative enactments – democratic legitimacy
  - Government actions without significant impact or with good reasons not to explain (i.e. investigations)
Why Explain?

A review of theories about explanation in the legal context suggests the following functions:

- Improve Decisionmaking “Accuracy”
- Promote Fair and Unbiased Decisionmaking
- Promote Decision Legitimacy
- Promote Trust in Decisionmakers
- Promote Compliance with Law/Rule
- Respect Individual Dignity and Autonomy

Similar rationales for explanation are common in other decisionmaking contexts
Improving Decisionmaking “Accuracy”

- What does “accuracy” entail from a legal perspective?
  - Correct interpretation of the applicable law
  - Correct application to the particular case
  - Amendment of the law if appropriate
Aspects of “Accuracy”

- **Accurate interpretation**
  - Consistent with rule, statute and/or precedent
  - Acceptable method for explicating ambiguities
  - Analytically sound application of method
  - Appropriately informed by an understanding of the law’s normative goals/policy
Aspects of “Accuracy”

- **Accurate application to a particular case**
  - Relies on accurate and relevant empirical facts
  - Employs an appropriate legal interpretation
  - Analytically sound application of law to facts
Aspects of “Accuracy”

- Amendment of the law in view of:
  - Democratically determined and constitutionally consistent goals/policy
  - Emerging factual scenarios (e.g. changing technology)
  - Changes in social values/views
How Can Explanations Improve Accuracy?

- For both interpretation and application, explanations may:
  - help decisionmakers to catch and avoid errors
  - incentivize more careful analysis
  - provide a basis for disputing decisions and review by higher authorities

- Explanations facilitate critique and debate about:
  - the interpretation of a rule and
  - the policy rationales underlying it

- Explanations may identify normatively troubling outcomes, suggesting a need to **amend** a rule
  - And help to determine how the rule is going wrong
Algorithms, Accuracy and Explanations

- If an automated algorithm:
  - Embodies the “right” rule and
  - All data (facts/evidence) needed to apply the rule are available and accurate

- Then, explanation is not necessary for:
  - Catching and avoiding errors
  - Incentivizing careful analysis
  - Providing a basis for disputing decisions and review by higher authorities in applying the rule to particular cases

- But what if those assumptions are not met?
- And what about the other ways in which explanations can improve accuracy?
Promoting Fair and Unbiased Decisions

- Unfair or biased decisions stem from:
  - Pernicious explicit motivations
  - Implicit or unconscious bias
  - Unanticipated results of applying a rule

- Pernicious explicit motivations
  - Attempts to obfuscate pernicious motivations may result in less persuasive explanations
  - Explanation requirement thus might deter acting on pernicious motivations or help to detect such motivations
  - Of course, this won’t always work
Promoting Fair and Unbiased Decisions

- **Implicit bias**
  - Explanation requirement might help to unmask implicit bias by highlighting weaknesses in decisionmaking rationale
  - Also not guaranteed to work

- **Unintended consequences of correct application of legal rules**
  - Explanations, cumulatively, may highlight biased or unfair outcomes, promoting reform
  - Also may not work
Empirical studies show that “procedural justice” promotes more favorable views of decisionmaking processes

- Explanation is an aspect of procedural justice that can have this effect
- Procedural justice has an evil twin: complacency in the face of substantive injustice!
- E.g. Provide an elaborate hearing, listen to an individual’s arguments, then make an unjust decision
- But explanation-giving is harder for the evil twin, so requiring an explanation deters “faux procedure”
Promoting Compliance

- Explanation clarifies requirements and expectations, makes it easier to meet them
  - For the subject of the decision who will face similar situations in the future
  - Cumulatively, for everyone, especially when explanations are aggregated by some intermediary

- Is gaming the system compliance’s evil twin?
  - Rule of law: citizens ordinarily have the right to know the law and comply (only) with the letter of the law
  - Gaming the system is only possible for decisions made on discretionary grounds, where compliance is not the goal (e.g. targeting investigations)
Promoting Dignity and Autonomy

- Explanations of decisions are inherently valuable because they show respect for the dignity of those affected.
- Explanations enhance autonomy by giving individuals options about whether and how to comply with the law.
- Explanations enhance dignity by treating individuals as democratic citizens rather than subjects.
Explanation and Automated Decisionmaking

- Can/do explanations serve the same purposes for automated decisionmaking?

- Are there substitutes for explanation in the context of automated decisionmaking?