Experiments on bias demonstrate the inability of judges and jurors, and other professionals to shrug off so-called ‘anchoring’ effects, and other common cognitive biases. The goal of ‘Just Cognition’ is to interrogate precisely what these examples teach us about (the veracity of our assumptions about) how bias operates within the judicial sphere.


Judging by Numbers

A group of qualified judges (the ‘participants’) are each given a pair of dice. Unbeknownst to the participants, the dice are loaded: one pair will produce a combined total of 3; the other 9. Participants are asked to roll the dice, and then to indicate the length of sentence that they would hand down for a fictitious shoplifting offence.

Judging by Numbers summarises an experiment run by Englich, Mussweiler and Strack and published in the Personality and Social Psychology Bulletin in 2006. That experiment generated a stark set of results: participants exposed to the higher number offered a sentencing recommendation averaging 8 months; participants
exposed to the lower number offered a sentencing recommendation averaging 5 months.

Though it has become emblematic of the pervasive effects of bias in the professional sphere, *Judging by Numbers* is not a solitary experiment; many others demonstrate the inability of judges and jurors, and other professionals (or non-professionals operating within professional contexts) to shrug off so-called ‘anchoring’ effects, and other common cognitive biases. The goal of ‘Just Cognition’ is to interrogate precisely what these examples teach us about (the veracity of our assumptions about) how bias operates within the judicial sphere.

In their own terms, Edmond and Martire set out to question ‘traditional legal approaches to bias’ (at 633), given a broad and enduring consensus that impartial decision-making is a key aspect of procedural legitimacy. For Edmond and Martire, due process has two limbs. The first is a participative limb: ‘persons whose rights or interests might be affected by a decision should be notified and afforded an opportunity to participate in the process before a decision is made’ (at 640). The second is an impartiality limb: ‘judges are expected (and assumed) to be impartial’ (at 640), where ‘impartial’ means ‘free from bias’. The article is concerned with the second limb, a general proscription of bias. In short, the target of the authors’ analysis is the claim—which makes up a crucial part of many accounts of natural justice, as a concern for the rule of law—that judicial reasoning must be impartial.

The authors’ primary focus in assessing traditional legal approaches to bias is what they term ‘judicial exceptionalism’—the claim that judges are able to avoid, or mitigate the impact of, biases that normally plague human decision-making, therefore contributing to a legal system that is procedurally robust. The authors argue that the empiric claim is not borne out by the evidence, and they make a broader plea for closer engagement with the cognitive scientific literature.

This enquiry is important and timely. In particular, it helps to inform a broader conversation about the appropriate contours and parameters for an interaction between human and algorithmic decision-making in the public sphere. If judicial decision-making is inexorably biased, the appropriate line of enquiry is not ‘does algorithmic decision-making entrench biases?’ but rather ‘does algorithmic decision-making amplify biases?’, where the comparative model is itself imperfect.

What, for the purposes of any such analysis, is ‘bias’? The authors describe bias as ‘the cognitive equivalent of a reflexive knee-jerk’ (at 646). Biases, they say, ‘occur quickly, effortlessly and automatically’ (at 646). Yet, it will be helpful at the outset to draw a general distinction between cognitive biases and what are typically referred to as ‘heuristics’. If ‘bias’ describes a ‘prepondering disposition or propensity’,[2] heuristics are ‘experience-based strategies that reduce complex cognitive tasks to simpler mental operations’.[3] Biases are predispositions; heuristics are shortcuts.

Neither heuristics nor biases are, as the authors note, intrinsically or instrumentally ‘bad’. Suppose that I am driving along a residential road, and a child runs out in front of my car. I perform an emergency stop, and the child runs off, unharmed. I do not have time to weigh all the reasons that apply to me: perhaps there is a car following closely behind me, which may bump the back of my vehicle if I perform the manoeuvre; perhaps I have not secured my seat-belt, and risk a neck injury. Neither reason will be sufficient to negative the case for performing the manoeuvre, but a full consideration of either would have rendered the exercise moot. In this instance, failing to reason in a careful and considered manner—deploying a reasoning-bypassing technique—brought about a successful action.
Not all shortcuts are internal. Lawyers are particularly familiar with one form of external norm, which may help bring about successful decision-making and attendant action. By setting out how we should act in particular circumstances, ousting its subjects’ own processes of practical reasoning, the law may (for some, must aim to) allow them to achieve some goal that could not be achieved, or which could not be achieved so readily, without it. Or particular rules may directly intermediate the decision-making process; the doctrine of judicial notice is one example of such a reasoning crutch, which allows a judge to draw a conclusion without interrogating particular facts that might constitute relevant epistemic reasons. Of course, particularists argue that we may be led into error if we apply external shortcuts (rules or principles) too rigorously, but even they accept the place of so-called ‘rules of thumb’ in guiding action.

Similar claims may be made about bias: a systematic tendency towards caution will help to keep us safe; a systematic tendency towards empathy may help us to form stronger bonds, avoid potentially harmful emotions etc. Thus, biases and (internal or external) heuristics may help to bring about some (individual or social) advantage. Yet, heuristics or biases may also provoke systematic logical errors. In the example with which I began, an anchoring effect leads to undue emphasis on a factor (the number revealed by the dice) that is irrelevant and ought to be disregarded. So-called ‘contextualism’ has been shown to generate conjunction fallacies. In one study, [4] 70% of university students thought that the following conclusion was logically admissible:

\[
\begin{align*}
P1 & \quad \text{All living things need water} \\
P2 & \quad \text{Roses need water} \\
\Delta & \quad \text{Roses are living things}
\end{align*}
\]

Adapt the example to remove contextual cues, and people perform much better:

\[
\begin{align*}
P1 & \quad \text{All animals that belong to the genus Pilarktos are hairy} \\
P2 & \quad \text{Garbles are hairy} \\
\Delta & \quad \text{Therefore, garbles belong to the genus Pilarktos}
\end{align*}
\]

A tendency towards pattern recognition is productive of ‘overdetermination’ (seeing trends that are not supported by the data); overdetermination leads people to invest unwisely, make personal life choices on the basis of horoscopy or astrology. Other fallacies generated by cognitive predispositions are too numerous to list. Similar errors may be productive of racist or misogynistic behaviours, and other forms of social exclusion. Clearly, then, the consequences of heuristics and biases may—often will—be negative. However, some of the cognitive science literature has been too quick to claim that whatever leads us to deviate from ‘normative reasoning’ (reasoning devoid of these errors) is bad for us: normative reasoning is ‘the type of thinking that we would all want to do, if we were aware of our own best interests, in order to achieve our goals’. [5] It bears emphasis, then, that systematic errors will not necessarily frustrate our goals. There may, for instance, be reasons why an individual may be content to make the same systematic errors as other people within a group; these errors may increase feelings of unity or belonging to a particular social
group, or reinforce the strength of a faith from which one draws comfort etc. If there are grounds for moral disapprobation, they are not derived from self-interest.

Systematic deviations from ‘perfect’ reasoning processes—those which are conducted thoroughly, in a manner that is relatively free from the sorts of biases that tend to produce errors of judgment—warrant censure whenever there is a reason to reason with care. To put this another way: we can blame people for failing to reason well whenever the justification for doing so outweighs any countervailing reasons for drawing upon heuristics or biases to bypass those processes. This will be so, inter alia, when they have been given a task that requires them to assess the actions of another individual, by reference to a scheme of judgment that we are committed to upholding, in a manner that has consequences for that other.

Finally, we arrive at the question to which the article is addressed: given reasons (and suitable conditions) to reason without the impact of heuristics and biases, are judges actually capable of doing so? The authors make a good case that the legal literature supports ‘judicial exceptionalism’—a species of what we might term ‘professional exceptionalism’, which assumes that professionals are able (given appropriate conditions for reasoning) to negate the effects of pervasive biases. And they make an equally robust case that the scientific literature supports, or provides part of the case for supporting, the opposite conclusion.

Edmond and Martire argue that we may conclude from examples such as Judging by Numbers that ‘biases might be a problem for decision-makers even when acting with the best intentions’ and “that being experienced or expert in some domain … and even being aware of dangers to cognition … may not be sufficient to counteract insidious effects’ (at 636). A wealth of evidence supports their claim. Indeed, we might go further: bias seems to be a problem for decision-makers even when the setting is formal, the conditions for decision-making robust, and the stakes are high.

When addressing the potential for individual judges to offset the influence of bias, the authors conclude that ‘Many biases (and heuristics) operate automatically or unconsciously such that decision-makers may not be aware of their influence and thus incapable of consciously “overpowering” them’ (at 636). Again, we might go further: there is a vast body of literature indicating that we are systemically incapable of recognising our own biases (so-called ‘blind-spot bias’).

Finally, the authors argue that “[j]udicial decisions may be improperly influenced, even determined, by factors other than (admissible) evidence and, notwithstanding appearances and representations, procedures may not actually be fair’ (at 648). If the authors are correct in their analysis, we might well feel justified in concluding that judicial decisions are inevitably influenced (perhaps decisively so) by irrelevant factors. There are, they suggest, routes to mitigating the impact of bias, but no evidence to support the conclusion that these methods can ever be wholly successful. By contrast, they say, judges ‘have a great deal of experience avoiding the appearance of partisanship and bias’ (at 648, emphasis original).

What should we conclude from all of this? Is judging merely a sham—a hopelessly partisan process clothed in the appearance of justice? The authors stop just short of saying that nothing can be done: ‘It may be’, they say ‘that legal procedures and experience do help judges to resist, to a degree, some of the influences that contaminate the cognition of most other humans’ (at 664). However, they emphasise the need for robust, evidence-based scrutiny of methods for securing judicial impartiality. That is a welcome call, which underscores a broader utility in engaging in meaningful interdisciplinary analysis.
Yet, the authors’ comments about the likely limitations of these procedures and experience raise a further question, which is of both theoretical and practical import: what is the particular value of human-to-human interaction in the decision-making process? Perhaps there is a participative value that is (wholly or partly) independent of the quality of the decisions so made—a value that may not be captured fully by the existing dichotomy between output-oriented accounts of due process, and accounts that perceive adjudication as ‘intrinsically’ valuable where it manifests a rigorous commitment to respect for autonomy.

An effort to articulate that value may require us to disaggregate due process concerns—further even than the two strands (participation and impartiality) identified by Edmond and Martire. And it may reveal tensions between those strands: the reasons for insisting that our legal processes be accurate, consistent and transparent may conflict with the reasons for requiring that they embed some element of human participation—if, as it turns out, human participation entails decision-making that is irremediably ‘poor’.

These questions have immediate conceptual ramifications for the judicial process: whatever participative value we identify will help to inform our conclusions regarding the desirability of e.g. jury trials. But they are of increasing practical significance, as we set out to prescribe the proper interaction between algorithms and human reasoning in the public sphere. Thus, in drawing attention to the lack of empirical support for claims about judicial exceptionalism, this article lays important groundwork for a broader enquiry into when and how we may outsource elements of public decision-making to non-human processes.

References


Published 20.04.2020