

# Bias mitigation

On Friday I gave a talk on cognitive and implicit biases, to a group of employment tribunal judges. The judges were a great audience, far younger, more receptive and more diverse than my own prejudices had led me to expect, and I enjoyed the opportunity to think about the area of cognitive bias, and how some conclusions from that literature might be usefully carried over to the related area of implicit bias.

First off, let's define cognitive bias versus implicit bias. Cognitive bias is a catch all term for systematic flaws in thinking. The phrase is associated with the 'Judgement and decision making' literature which was spearheaded by [Daniel Kahneman](#) and colleagues (and for which he received the Nobel Prize in 2002). Implicit bias, for our purposes, refers to a bias in judgements of other people which is both unduly influenced by social categories such as sex or ethnicity *and* in which the person making this biased judgement is either unaware or unable to control the undue influence.

So from the cognitive bias literature we get a [menagerie of biases](#) such as '[the overconfidence effect](#)', '[confirmation bias](#)', '[anchoring](#)', '[base rate neglect](#)', and on and on. From implicit bias we get findings such as that maths exam papers are marked higher when they carry a male name on the top, job applicants with stereotypically black American names have to send out twice as many CVs, on average, to get an interview or that people sit further away from someone they believe has a mental health condition such as schizophrenia. Importantly all these behaviours are observed in individuals who insist that they are not only *not* sexist/racist/prejudiced but are actively *anti*-sexism/racism/prejudice.

My argument to the judges boiled down to four key points, which I think build on one another:

## 1. Implicit biases are cognitive biases

There is slippage in how we identify cognitive biases compared to how we identify implicit biases. Cognitive biases are defined against a standard of rationality – either we know the correct answer (as in the [Wason selection task](#), for example), or we feel able to define irrelevant factors which shouldn't affect a decision (as in the framing effect found with the '[Asian Disease problem](#)'). Implicit biases use the second, contrastive, standard. Additionally it is unclear whether the thing being violated is a standard of rationality, or a standard of equity. So, for example, it is unjust to allow the sex of a student influence their exam score, but is it irrational? (If you think there is a clear answer to this, either way, then you are more confident of the ultimate definition of rationality than a full century of scholars).

Despite these differences, implicit biases can usefully be thought of as a kind of cognitive bias. They are a habit of thought, which produces systematic errors, and which we may be unaware we are deploying (although elsewhere I have argued that the evidence for the [unconscious nature of these process is over-egged](#)). Once you start to think of implicit biases and cognitive biases as very similar, it buys some important insights.

Specifically:

## **2. Biases are integral to thinking**

Cognitive biases exist for a reason. They are not rogue processes which contaminate what would be otherwise intelligent thought. They are the foundation of intelligent thought. To grasp this, you need to appreciate just how hard principled, consistent thought is. In a world of limited time, information, certainty and intellectual energy cognitive biases arise from necessary short-cuts and assumptions which keep our intellectual show on the road. Time and time again psychologists have looked at specific cognitive biases and found that there is a good reason for people to make that mistake. Sometimes they even find that animals make that mistake, demonstrating that even without the human traits of pride, ideological confusion and general self-consciousness the error persists – suggesting that there are good evolutionary reasons for it to exist.

For an example, take confirmation bias. Although there are risks to preferring to seek information that confirms whatever you already believe, the strategy does provide a way of dealing with complex information, and a starting point (i.e. what you already suspect) which is as good as any other starting point. It doesn't require that you speculate endlessly about what might be true, and in many situations the world (or [other people](#)) is more than likely to put contradictory evidence in front of you without you having to expend effort in seeking it out. Confirmation bias exists because it is an efficient information seeking strategy – certainly more efficient than constantly trying to disprove every aspect of what you believe.

Implicit biases concern social judgement and socially significant behaviours, but they also seem to share a common mechanism. In cognitive terms, implicit biases arise from our tendency towards associative thoughts – we pick up on things which co-occur, and have the tendency to make judgements relying on these associations, even if strict logic does not justify it. The scope of how associations are created and strengthened in our minds is beyond the scope of the post.

For now it is clear that making judgements based on circumstantial evidence is unjustified but practical. An uncontentious example might be you get sick after eating at a particular noodle bar. Maybe it was bad luck, you were going to get sick anyway or it was the sandwich you ate a lunch, but the odds are good you'll avoid the noodle bar in the future. Why chance it, there are plenty of other restaurants? It would be impractical to never make some assumptions, and the assumption-laden (biased!) route offers a practical solution to the riddle of what you should conclude from your food poisoning.

## **3. There is no bias-free individual**

Once you realise that our thinking is built on many fast, assumption-making, processes which may not be perfect – indeed which have systematic tendencies which produce the errors we identify as cognitive bias – you then realise that it would be impossible to have bias-free decision processes. If you want to make good choices today rather than a perfect choices in the distant future, you have to compromise and accept decisions which will have some biases in them. You cannot free yourself of bias, in this sense, and you shouldn't expect to.

This realisation encourages some humility in the face of cognitive bias. We all have biases, and we shouldn't pretend that we don't or hope that we can free ourselves of them.

We can be aware of the biases we are exposed to and likely to harbour within ourselves. We can, with a collective effort, change the content of the biases we foster as a culture. We can try hard to identify situations where bias may play a larger role, or identify particular biases which are latent in our culture or thinking. We can direct our bias mitigation efforts at particularly important decisions, or decisions we think are particularly likely to be prone to bias. But bias-free thinking isn't an option, it is part of who we are.

#### **4. Many effective mitigation strategies will be supra-personal:**

If humility in the face of bias is the first practical reaction to the science of cognitive bias, I'd argue that second is to recognise that bias isn't something you can solve on your own at a personal psychological level. Obviously you have to start by trying your honest best to be clear-headed and reasonable, but all the evidence suggests that biases will persist, that they cannot be cut out of thinking and may even thrive when we think ourselves most objective.

The solution is to embed yourself in groups, procedures and institutions which help counter-act bias. Obviously, to a large extent, the institutions of law have evolved to counter personal biases. It would be an interesting exercise to review how legal cases are conducted from a psychological perspective, interpreting different features as to how they work with or against our cognitive tendencies (so, for example, the adversarial system doesn't get rid of confirmation bias, but it does mean that confirmation bias is given equal and opposite opportunity to work in the minds of the two advocates).

Amongst other kinds of '[ecological control](#)' we might count proper procedure (following the letter of the law, checklists, etc), control of (admissible) information and the systematic collection of feedback (without which you may not ever come to realise that you are making systematically biased decisions).

Slides from my talk here as [Google docs slides](#) and [as PDF](#). Thanks to Robin Scaife for comments on a draft of this post. Cross-posted to the blog of our Leverhulme trust funded project on "[Bias and Blame](#)".

Tom Stafford

22 September 2015

<http://www.tomstafford.staff.shef.ac.uk/?p=342>

