Fitzgerald highlights a challenge not only for medical professionals but for all of us.

**Challenge of Implicit Bias** We cannot avoid habituating reliably vicious motivations. However virtuous our consciously-endorsed goals and deliberative commitments, implicit bias suggests that we will inevitably develop vices which systematically undermine the effectiveness of all the good intentions in the world.

**Discussion: Thesis & Terminology**
- In a group, discuss questions 1–4 (thesis and terminology).
- Extending your answer to question 3, what is the third conception of ‘conscience’ Fitzgerald discusses? How is this similar to and different from (i) the ‘dominant’ conception and (ii) Fitzgerald’s conception?

**Attitudes**

As noted in the course packet, the traditional conception of an attitude in philosophy is rather different from the social psychologist’s sense. We therefore need to be especially careful to ensure that we understand in which sense the term is being used by a given author. Fitzgerald is using ‘attitude’ in the social psychologist’s sense rather than the traditional philosophical one.

What, then, are attitudes in the social psychologist’s sense?

**Attitudes** (in the social psychologist’s sense) are complex, structured evaluations of objects with cognitive, affective and behavioural components which have functional roles in a person’s psychology (Maio and Haddock 2010). Roughly speaking, attitudes are ‘likings’ and ‘dislikings’. What is liked or disliked may be very concrete and particular e.g. my mother’s dislike of a particular scratch on her new floor,

extremely abstract and general e.g. Martin Luther King, Jr.’s love of equality, Elizabeth Fry’s love of prison reform

The content of an attitude towards an object is a function of cognitive, affective and behavioural components associated with that object.

e.g. Fry’s associated belief that even the worst criminals deserve to be treated as human beings and offered opportunities to reform (cognitive);

e.g. Fry’s associated feelings of sympathy for women suffering the misery of conditions in Newgate (affective);

e.g. Genia’s memory of choosing or asking for red clothing (associated past behaviour).

An attitude’s content also includes how much its object is liked or disliked
e.g. the mildness of Bill’s liking for pickled Brussels sprouts, the intensity of Genia’s love of red.

The strength of an attitude is a matter of the strength of the connections between both

- the components which make up the attitude
and

- those components and other elements in the cognitive system, situational features, behaviours and so on.

Within the cognitive system, the strength of the connection between any two elements is a matter of how readily they influence each other when processing information.

e.g. If I have a mild dislike of peas which is strongly connected to the goal of eating vegetables I prefer, on the one hand, and to the presence of bowls of small green balls, on the other, this means that, if there’s a bowl of small green balls on the table, my mild dislike of peas is very likely to influence my cognitive processing by activating my goal of eating vegetables I prefer. So, making some plausible assumptions about the content of my goal, if there is a dish of peas on the table, I am likely to engage in preferred-vegetable-seeking behaviour (e.g. looking around for bowls of orange rings or green leaves) and, if this results in preferred-vegetable-finding, preferred-vegetable-choosing behaviour (e.g. helping myself to carrots and spinach rather than peas) is very likely to follow.

Explicit attitudes are attitudes measured by self-report.

Implicit attitudes are attitudes measured in ways which exclude the possibility of deliberative, conscious control (e.g. by requiring responses which are too fast for people to consciously deliberate).

- Contra Fitzgerald implicit attitudes are not ‘attitudes that are below the level of conscious awareness’ (2014, 25). This may or may not be the case. At any rate, it is controversial among social psychologists.

This does not mean that deliberation cannot influence our implicit attitudes or the effect of those attitudes on cognitive processing. But deliberation cannot directly control the influence of these attitudes on our cognition.

e.g. If you take the Implicit Association Test (IAT), no amount of deliberation as you do the test will change the result because the test requires you to respond too quickly for this to be effective.

But you might be able to influence the results through deliberative efforts before the testing procedure begins — either in the short-term (e.g. by engaging in deliberation as you are about to begin the test) or in the long-term (e.g. by implementing deliberative strategies in the weeks, months or years before the test).

e.g. If you are taking the dark-skin/light-skin IAT, you might think about counter-stereotypical individuals just before the test, put pictures of these people on the walls of the testing room etc. Or you might have implemented a longer-term strategy aimed at achieving your goal of treating people equally which affects whether, and how, your implicit attitudes influence your responses. Either way, conscious deliberation may indirectly influence your non-deliberative responses during the test.

It is important to understand that implicit attitudes are essential to our ability to navigate the world. We just don’t have the time and cognitive resources to deliberate about whether to adopt a positive or negative attitude to everything. Imagine having to consciously consider whether to like or dislike the dog pooh on the pavement every time you come across some when walking, or whether to prefer the left or right shoe when dressing your left foot each time you get ready to leave home.

Moreover, ‘bias’ in the technical sense is not a problem either. There is no fundamental moral objection (I assume) to my explicit bias against peas and towards carrots and spinach. And, while it might be less convenient in some cases, it would not, I assume, be any more problematic, from an ethical perspective, if I was not aware of my dislike of peas despite systematically engaging in pea-avoiding behaviour.

Implicit bias, in this context, refers to the subset of biased implicit attitudes which are forms of prejudice. These attitudes are based on morally problematic implicit associations with characteristics such as race or sex which influence automatic cognition, perhaps despite deliberative abhorrence e.g. race-weapons stereotype.
Evidence suggests that implicit biases affect decision-making and behaviour in a wide-range of social contexts. Consider the ‘shooter task’ developed by Joshua Correll’s laboratory to investigate the effects of racial stereotypes on policing in the US (figure 2). The ‘shooter task’ requires subjects to decide quickly whether to shoot or not in response to images of people carrying either weapons (e.g. gun) or non-weapons (e.g. carton of milk).

- Subjects’ goal is to shoot all and only figures with guns.
- The measure of ‘shooter bias’ is implicit since fast, automatic responses required.
  - Subjects ‘don’t have time to think’.
- Statistically, subjects are more likely to shoot black figures.
- Individuals with stronger race-weapons stereotypes show greater shooter bias.

Discussion: The Argument

- In a group, please discuss question 5. What is Fitzgerald’s argument?
- Clarify the conclusion.
  - The conclusion is the claim that the author is trying to persuade you of.
- What are the premises of this argument?
  - Premises are claims which an argument relies on i.e. the points it starts from.

Goals

Goals share much in common with attitudes. Like attitudes, they may be explicit or implicit and explicit goals, like explicit attitudes, may be automated through repeated activation. Also like attitudes, goals may be more or less accessible to cognitive processing, depending on the strength of the connections between them and other elements in the cognitive processing system. More accessible goals, like stronger attitudes, are more likely to be activated and more likely to influence cognitive processing. Goals, like attitudes, can influence cognition with or without conscious awareness.

Again like attitudes, goals can be very concrete and particular or very abstract and general e.g. being a good parent, treating people fairly, catching the 3 o’clock train to London next Tuesday.

Goals are similarly complex, as well. They link internal and external stimuli with flexible and intelligent behavioural plans e.g. infant’s cry → check nappy, offer food etc.; interviewing black applicant → making good eye contact, treating with respect etc.; watch shows it is 2 o’clock on Tuesday → leave house to catch bus to train station etc.

Goals can be adopted, with or without conscious awareness, without being chosen. But, unlike attitudes, goals can also be chosen — they may be the objects of volition (acts of will).

The precise relationship between attitudes and goals remains largely undetermined. One body of psychological research has investigated attitudes and attitude change (see Maio and Haddock 2010, for an overview). Another, largely distinct, body of research has investigated the automatisation of goals (for a philosophical introduction, see Snow 2006).

Egalitarian Goals

Note that the research Fitzgerald cites does not support her claim that ‘Sheena will only be able to fully embrace the value of equal treatment for women if she also develops awareness of her existing
Dr. Clea F. Rees

The Problem of Implicit Vice

<table>
<thead>
<tr>
<th>Goal</th>
<th>Strength of Generic Regulative Capacity</th>
<th>Will Show Strong Associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appear non-prejudiced</td>
<td>Strong</td>
<td>✔</td>
</tr>
<tr>
<td>Appear non-prejudiced</td>
<td>Weak</td>
<td>✗</td>
</tr>
<tr>
<td>Behave without prejudice</td>
<td>Strong</td>
<td>✔</td>
</tr>
<tr>
<td>Behave without prejudice</td>
<td>Weak</td>
<td>✔</td>
</tr>
</tbody>
</table>

Table 1: Experimental assumptions (slightly simplified) in investigations of IMCP concerning the relationship between (i) reasons for wanting to respond without prejudice on tests of implicit associations, (ii) strength of generic capacities for control over the influence of implicit attitudes on behaviour and (iii) patterns of implicit associations between bad/prejudice and self/prejudice demonstrated in testing.

implicit attitudes towards women and works to change them’ (28).

What empirical work does support is the claim that Sheena needs to develop an implicit commitment to controlling the influence of any implicit biases she has against women. But Sheena might develop this automated goal whether or not she comes to believe that she actually has any such bias. Even if the development and pursuit of this egalitarian commitment is incompatible with the belief that she is free of such bias, it does not follow that the goal requires a positive belief that she does have such bias.

The work which Fitzgerald is referring to concerns what is called Implicit Motivation to Control Prejudice (IMCP). This research does not investigate the effects of automated egalitarian goals on the behavioural manifestation of implicit bias directly. Rather than assessing the existence of implicit egalitarian motivations as such, IMCP measures the strength of implicit associations between:

1. badness and prejudice;
2. self and prejudice.

Test results demonstrating strong associations of both kinds is understood as evidence for a combination of two strong two implicit attitudes:

1. negative evaluation of prejudice;
2. evaluation of self as prejudiced.

The existence of implicit egalitarian goals — that is, implicit motivations to inhibit the influence of implicit bias on behaviour — is then inferred from this combination of implicit attitudes. (The idea, clearly, is that if you think prejudice is bad and you think you are prejudiced, you are likely to be motivated to control the impact of your prejudice on the way you treat people.)

But a strong association between badness and prejudice (1) alone is enough to inhibit the influence of implicit bias, whereas a strong between self and prejudice (2) alone is (unsurprisingly) insufficient (Rees 2016, 205, citing Glaser and Knowles 2008, 170). That is, a negative implicit attitude towards prejudice predicts successful control of implicit bias, even in the absence of a belief that, or evaluation of, the self as prejudiced. (This makes sense since you can be motivated to behave in a non-prejudiced way, even if you don’t think you are prejudiced. You can, for example, be motivated to inhibit the influence of any implicit biases you have without thinking that you actually have any.)

The importance of association (2) is that it allows experimenters to isolate the influence of a goal to control the effects of implicit bias on behaviour by excluding both individuals who may be concerned to inhibit the influence of implicit bias on their behaviour only because they are concerned not to appear prejudiced and those who possess a strong capacity to control the influence of their implicit attitudes on their behaviour generally (table 1).

The criteria for IMCP allow researchers

i. to identify a group of subjects who are concerned to inhibit the influence of their implicit biases because they are concerned about being prejudiced in their behaviour;

ii. to isolate the impact of this specific concern
Key Points
- Goals can be adopted as a result of conscious deliberation.
- Adopted goals can be automatised by repeated activation.
- Automatised egalitarian commitments can reliably shape cognitive processing, inhibiting the influence of implicit bias on cognitive processing.

Figure 3: Selective summary.

on their behaviour.

Individuals high in IMCP

either are relatively unconcerned to avoid appearing prejudiced;
or have a weak generic capacity to control the influence of their implicit attitudes on their behaviour.

Individuals who manifest strong associations between badness and prejudice (1) but weak associations between self and prejudice (2)
might have the goal of not behaving in prejudiced ways (row 3);
or might simply have the goal of appearing not to be prejudiced, combined with a good generic ability to control the effect of their implicit attitudes on their behaviour (row 1).

That is, the results of testing the strength of individuals’ associations between bad and prejudice, on the one hand, and self and prejudice, on the other, do not distinguish between these two groups. That is, the patterns of associations demonstrated in testing will be identical, even though subjects’ actual associations will differ.

The only subjects who can be shown to have a commitment to not manifesting prejudice (as opposed to merely not appearing to manifest it) are those subjects who show the pattern of associations in the fourth row — that is, those who meet the criteria for high IMCP. It is, therefore, this subset of subjects with strong implicit negative attitudes to prejudice who are the focus of the research.

Investigations show that individuals high in IMCP are able to inhibit the influence of implicit biases on cognitive processing without conscious deliberation. That is, the implicit motivation shapes their responses to stimuli such as those used in the IAT automatically. This means that the influence of implicit bias is mitigated or eliminated from even their fast, non-deliberative responses.

For example, subjects high in IMCP show less shooter bias than controls with similarly strong racial stereotypes (figure 2). A strong, automatised commitment to controlling prejudice prevents shooter bias even in subjects with strong race-weapons stereotypes. The automatised egalitarian goal inhibits the activation of the stereotype preconsciously.

This kind of automatic control does not require effortful cognition and is relatively immune from the impact of constraints on the availability of the cognitive resources required for deliberative, conscious control.

Effortful control is a limited resource. Although we can increase our capacity for this kind of control through practice — the analogy here is with strengthening muscles by exercising — the resource remains a limited one for any individual. This means that if you have just completed an especially cognitively demanding task (e.g. solving difficult logic puzzles), you are less able to engage in the kind of cognition required to exercise conscious control over your behaviour.

Implicit motivation, in contrast, appears able to control the influence of implicit bias without requiring or depleting this limited resource.

Discussion: Evaluation

- In a group, please discuss questions 6–8. Is Fitzgerald’s argument a good one?
- Is the argument valid?
  - An argument is valid if, and only if, the conclusion follows logically from the
premises so that the premises could not possibly be true without the conclusion being true as well.

- **Is the argument sound?**

  - An argument is sound if, and only if, two conditions are satisfied. First, the argument must be valid. Second, all of the premises must be true. It follows that the conclusion of a sound argument is true.

## Reflection

- Are implicit biases vices?

- Under what circumstances (if any) are we blameworthy for our implicit biases?

- Under what circumstances (if any) are we blameworthy for the negative influence of these biases on our behaviour?

- Does a properly functioning conscience require awareness of implicit biases?

- Does ‘integrity’? If so, in which sense(s)?

- How should the problem of implicit bias in health care be tackled? What are the dangers of such programmes and how might they best be mitigated or avoided?

- Should awareness of implicit biases and strategies for mitigating, reducing or eliminating them form a part of the ‘character education’ programmes which have recently gained such popularity?

## References


