# Are we morally responsible for our implicit biases?

## Introduction

This paper argues that we are morally responsible for our implicit biases by suggesting that virtue ethics can rebut Saul's (2013) argument from lack of control. First, I will outline the argument from lack of control before looking at Holroyd's (2012) suggestion that we possess indirect long-range control over our implicit biases. I will then develop Holroyd's (2012) argument by looking at a possible defence from virtue ethics. I will predominantly focus on how acquiring the virtue of courage may enable us to exercise control over our implicit biases after considering the nature of implicit biases. The final section of this paper will respond to two possible objections to my argument.

## Argument from lack of control

Consider a moral agent, Eric. Eric is often late for work as he would rather climb up eight flights of stairs than take the lift due to his claustrophobia. Eric knows that there is no rational basis for his phobia, understanding that getting trapped in the lift is highly unlikely. However, Eric simply cannot control the way closed spaces make him feel. Saul (2013) suggests that the lack of control Eric demonstrates is not merely limited to our fears but that it extends to our beliefs.

Suppose Eric were a recruiter for an employment agency who has never recruited a female mechanic. Eric's failure to recruit female mechanics is completely unrelated to the level of their ability and entirely the result of an implicit bias he possesses. Saul (2013) would argue that Eric ought not to be held responsible for his bias as he has no immediate and direct control over it. Holroyd (2012) summarises Saul's (2013) argument as follows :

> P1: Moral agents cannot be held responsible for cognitive states over which they lack immediate and direct control.
>
> P2: Moral agents lack immediate and direct control over their implicit biases.
>
> C: Moral agents cannot be held responsible for their implicit biases.

This argument can be seen to originate from the nature of an implicit bias. Holroyd (2012) states that implicit biases are commonly perceived to be "below the radar of conscious reflection", beyond the moral agent's control and not revisable in the same sense our reflective beliefs tend to be. This leads

Saul (2013) to suggest that even when a moral agent is aware of their implicit bias, they cannot be held responsible for it as they would not be immediately able to exercise control over it. In fact, psychological research has revealed a rebound effect which occurs once individuals attempt to suppress their biases (Holroyd, 2012).

Suppose Eric's diversity and inclusion officer, Talia, were to make him aware of the fact that he appears to harbour an implicit bias against female mechanics. Under Saul's (2013) view, Talia would not be able to reprimand Eric for his bias or hold him responsible for it. Now suppose Eric's boss, David, was to make him aware of the fact that he arrives at work late almost every day. It would be reasonable to assume that David would hold Eric responsible for his tardiness, even though Eric appears to have no direct control over his claustrophobia. This leads me to suggest that Saul's argument from control is not entirely plausible and that perhaps Talia should hold Eric responsible for his implicit bias regardless of whether he possesses direct control over it.

## Indirect long-range control

Holroyd (2012) and Hieronymi (2008) would likely argue that Eric still ought to be held responsible for his implicit bias even in the absence of direct control. This is because Holroyd (2012) and Hieronymi (2008) both believe that we have long-range indirect control over our implicit biases. One example of exercising long-range indirect control is exercising bladder control; children are not born with direct control over their bladders; they acquire it over time through practice and are expected to do so. Holroyd (2012) argues that long-range indirect control is also essential to morally relevant activities. This idea is going to be explored further in the following paragraphs.

### The cultivation of virtue

Holroyd (2012) claims that the cultivation of virtue stems from long-range control. To better understand Holroyd's (2012) claim, consider the three essential features of a virtue. Foot (2002) describes virtue as:

1. A "disposition" of the will
2. Advantageous to others (and to its possessor at times)
3. "Corrective" of harmful human tendencies

Features 1 and 3 are particularly important to this essay's central argument. Hacker-Wright (2013) highlights the use of the word "disposition" rather than "habit". The word "habit" is indicative of "relatively automatic conduct" (Hacker-Wright, 2013), meaning it is suggestive of the exercise of direct control, whereas "disposition" is not. Foot (2002) further suggests that virtues result from long-range control by claiming that they are acquired through practice and can be cultivated later in life.

Now consider feature 3. Assuming that implicit biases are comparable to destructive human tendencies, I argue that we can correct them through virtue, suggesting that we possess indirect long-range control over our implicit biases. This line of reasoning is not unique to philosophical thought; Iñiguez (2020) states that virtue ethics has influenced both the positive psychology movement and personal and professional development schemes in the business world. This suggests that we are morally responsible for our implicit biases. While we can exercise indirect long-range control over our implicit biases by exercising different virtues, this essay is going to focus on the virtue of courage in particular.

## The virtue of courage

Not all implicit cognitions are necessarily harmful; in many circumstances, implicit cognitions are necessary for making decisions and solving problems. Underwood (1996) suggests that implicit cognitions are shortcuts our brains create to respond to situations quickly. The problem is that some of these shortcuts lead to bias. I argue that fear is the catalyst that turns an implicit cognition into an implicit bias. This argument is informed by Leslie's (2017) claim that:

> "'Ks are F' is true iff some Ks are F, given that being F is a dangerous or harmful property."

To explain, we are more likely to rapidly generalise harmful and dangerous behaviour from individuals to a group, and this tendency is often motivated by fear. Leslie (2017) illustrates this idea by comparing the number of murders an individual must commit to be labelled a murderer to the number of times one has to worry to be considered a worrier. We draw negative conclusions about a group out of fear of some of its members' harmful or violent acts.

Gramlich (2021) suggests that we may have indirect long-range control over our fears. The Little Albert study illustrated that phobias could be seen as CERs (conditioned emotional responses) to objects and situations (Gramlich, 2021). This implies that our fears are 'learned', suggesting that they are under our control. I believe that we would still be morally responsible for our implicit biases even if we were to ignore psychological research, contending that we are not in control of our acquisition of fears, and take a purely philosophical approach instead. This is because we still possess long-range direct control over our cultivation of the virtue of courage. This means that even if we are not responsible for acquiring our implicit biases, we are responsible for their possession.

In Plato's (1926) Laches, the title character eventually describes courage as a "wise perseverance" after quarrelling over the cardinal virtue. I believe that the virtue can be defined in many ways; Ungvarsky

(2021) suggests that courage can be seen as acting with moral strength. As the virtue of courage enables us to act on our moral values rather than our fears or impulses, cultivating it can be seen as an effective way of exercising indirect long-range control over our implicit biases. Through cultivating courage, one can learn to overcome their implicit biases by learning to act with moral strength despite their fears. This suggests that we possess control over our implicit biases, entailing that we are responsible for them.

## Possible objections and responses

### 1. Implicit biases may not stem from fear.

One could argue that my assumption that our implicit biases are motivated by fear is misguided. Consider the example of the job recruiter made above. One might question whether Eric's implicit bias against female mechanics is genuinely motivated by fear. Unlike being a murderer, being a female mechanic hardly appears to be a dangerous or harmful property. I argue that this does not necessarily mean that Eric's implicit bias is not motivated by fear. Leslie (2017) suggests that what leads us to form implicit biases about some groups and not others is our degree of familiarity with them. Leslie (2017) is correct to suggest that we have a higher tendency to form implicit biases about some groups rather than others. For example, while several musicians have been proven to have committed dangerous and harmful acts, most individuals are unlikely to harbour an implicit bias against musicians. I suggest that this is because while we may not personally know any musicians, we possess a level of solidarity with them, which we exhibit when we stream their music. The fact that we are quick to form implicit biases against some groups and not others suggests that our implicit biases are fueled by xenophobia. Therefore, it follows that Eric's implicit bias may be prompted by xenophobia due to his lack of solidarity and familiarity with female mechanics. This leads me to argue that Eric's implicit bias can still be corrected by cultivating the virtue of courage, suggesting that Eric is morally responsible for his implicit bias.

The Spinozan theory of belief fixation offers a strategy for controlling our implicit biases by increasing exposure to stigmatised groups, thus increasing familiarity and knowledge with them, further implying that our biases stem from xenophobia (Spinoza, 1677). Spinoza (1677) also claims that we can avoid acquiring harmful beliefs by avoiding exposure to harmful ideology, suggesting that learning harmful generalisations about stigmatised groups can be avoided. Leslie's (2017) claim that individuals are less likely to accept generalisations about groups they identify or are familiar with increases the strategy's credibility. Suppose Eric were to recognise that he has a professional responsibility to seek treatment for

his claustrophobia to improve his punctuality. His treatment would likely involve increased and prolonged exposure to closed spaces (Colwill, 2021). I believe that Eric's treatment would also involve the cultivation of the virtue of courage as Eric would learn to resist the urge to act on impulse through his increased exposure.

I argue that Eric also has a moral responsibility to correct his implicit bias by exercising the same virtue. By increasing his exposure to female mechanics- be it through looking at more CVs or educating himself- and actively avoiding incorrect generalisations, Eric can increase his familiarity and solidarity with female mechanics to overcome his xenophobia. Over time, this should enable Eric to resist his urge to judge mechanics based on their sex rather than their ability, enabling him to overcome his implicit bias by cultivating virtue. This process is not instantaneous and is indicative of cultivating virtue which entails exercising long-range indirect control. This leads me to reiterate that we have long-range control over our implicit biases, meaning that we are morally responsible for their possession.

## 2. Internalised biases cannot be controlled by exercising courage.

Suppose our job recruiter were a female named Erica who also possessed an implicit bias against female mechanics. With this example, one could argue that Erica's implicit bias is not necessarily motivated by xenophobia, as she harbours an implicit bias against members of a group she identifies with. Erica's example is not an isolated incident; many individuals harbour implicit biases against themselves. This could lead us to contend that courage cannot confound implicit biases as they may not be motivated by fear.

I argue that Erica's implicit bias is still motivated by fear: a fear of non-conformity. To explain, members of marginalised groups may harbour internalised biases due to a fear of going against societal norms. Suppose Erica were to hire a female mechanic who happened to possess any of the negative properties individuals tend to generalise to female mechanics (or female workers in general). This could reflect poorly on Erica. As a female employee who cannot afford to take risks or make mistakes, Erica unknowingly makes what many of her colleagues would view as a safe decision. Erica adheres to our patriarchal society's norms by recruiting male mechanics to work in a male-dominated industry. While this is problematic, it is motivated by a primitive human instinct of self-preservation, suggesting that Erica's implicit bias is motivated by fear. I suggest that different biases can be motivated by different fears or even a myriad of them and do not limit my argument to xenophobia. If anything, I believe that this example further supports my argument by illustrating why virtue ethics offers an appropriate strategy for overcoming implicit biases. To explain, unlike other ethical theories such as deontology,

virtue ethics does not determine the morality of acts by considering their adherence to norms (Foot, 2002). This leads me to maintain that we are responsible for our implicit biases as we can exercise indirect long-range control over them by cultivating virtue.

## Conclusion

To conclude, I have argued in favour of Holroyd's (2012) claim that we are morally responsible for our implicit biases by illustrating how the cultivation of virtue can enable us to exercise indirect long-range control over them. My argument stems from my understanding of the nature of an implicit bias; I believe that our implicit cognitions turn into implicit biases once they are motivated by fear. Setting aside psychological research, which suggests we are responsible for acquiring fears, we are still responsible for our implicit biases. This is because we possess indirect long-range control over our fears by exercising virtue, an act which entails exercising long-range control itself. The virtue of courage, in particular, can enable us to act with moral strength despite our fear, enabling us to exercise control over our implicit biases. I reasoned that our biases could be motivated by a myriad of fears, be it a fear of violence, xenophobia or even a fear of non-conformity. Cultivating courage can effectively combat biases motivated by these fears, as virtue ethics is not contingent on an act's adherence to norms. This means that cultivating virtue is helpful in learning to overcome generalisations regarding stigmatised groups, making it a valuable tool in exercising control over our implicit biases.

## References


1. Colwill, R. M. (2021) 'Implosion (behavior therapy)', Salem Press Encyclopedia of Health. Available at: https://0-search-ebscohost-com.pugwash.lib.warwick.ac.uk/login.aspx?direct=true&db=ers&AN=93872036&site=eds-live (Accessed: 10 May 2022).
2. Foot, P., (2002). Virtues and vices and other essays in moral philosophy. Oxford University Press on Demand.
3. Gramlich, C. A. (2021) 'Fear', Salem Press Encyclopedia of Health. Available at: https://0-search-ebscohost-com.pugwash.lib.warwick.ac.uk/login.aspx?direct=true&db=ers&AN=93871929&site=eds-live (Accessed: 7 May 2022).
4. Hacker-Wright, J., (2013). Philippa Foot's Moral Thought. A&C Black.
5. Hieronymi, P. (2008). Responsibility for believing. Synthese, 161(3), pp.357-373.
6. Holroyd, J. (2012), Responsibility for Implicit Bias. Journal of Social Philosophy, 43: 274-306. Available at: https://0-doi-org.pugwash.lib.warwick.ac.uk/10.1111/j.1467-9833.2012.01565.x (Accessed: 5 May 2022).

7. Iñiguez, S. (2020). Virtue: Philippa Foot/Angelica Kohlmann. In: In an Ideal Business. IE Business Publishing. Palgrave Macmillan, Cham. Available at: https://0-doi-org.pugwash.lib.warwick.ac.uk/10.1007/978-3-030-36379-6_4 (Accessed: 6 May 2022)
8. Leslie, S.J. (2017). The original sin of cognition: Fear, prejudice, and generalisation. The Journal of Philosophy, 114(8), pp.393-421.
9. Plato (1924): Laches, Protagoras, Meno, Euthydemus. Greek with translation by W. R. M. Lamb. Loeb Classical Library 165. Harvard Univ. Press ISBN 978-0674991835 HUP listing
10. Saul, J., (2013). Unconscious influences and women in philosophy. Women in philosophy: What needs to change, pp.39-60.
11. Spinoza, B. (1677). Ethics. Indianapolis, IN, Hackett, 1991.
12. Underwood, G. D. M. (1996). Implicit cognition. Oxford University Press.
13. Ungvarsky, J. (2021) 'Courage', Salem Press Encyclopedia. Available at: https://0-search-ebscohost-com.pugwash.lib.warwick.ac.uk/login.aspx?direct=true&db=ers&AN=137502122&site=eds-live (Accessed: 6 May 2022).