



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Epistemic Injustice and Implicit Bias“

verfasst von / submitted by

Sonja Riegler, BA BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien, 2019 / Vienna, 2019

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 066 941

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet: Philosophie

Betreut von / Supervisor:

Univ.-Prof. Dr. Martin Kusch

"Hiermit erkläre ich, die vorgelegte Arbeit selbständig verfasst und ausschließlich die angegebenen Quellen und Hilfsmittel benutzt zu haben. Alle wörtlich oder dem Sinn nach aus anderen Werken entnommenen Textpassagen und Gedankengänge sind durch genaue Angabe der Quelle in Form von Anmerkungen bzw. In-Text-Zitationen ausgewiesen. Dies gilt auch für Quellen aus dem Internet, bei denen zusätzlich URL und Zugriffsdatum angeführt sind. Mir ist bekannt, dass jeder Fall von Plagiat zur Nicht-Bewertung der gesamten Lehrveranstaltung führt und der Studienprogrammleitung gemeldet werden muss. Ferner versichere ich, diese Arbeit nicht bereits andernorts zur Beurteilung vorgelegt zu haben."

Paris, 2019
Sonja Riegler

TABLE OF CONTENTS

| | |
|---|-----------|
| DANKSAGUNG..... | 5 |
| ABSTRACT..... | 7 |
| 1. INTRODUCTION..... | 8 |
| 1.1 Development and Methodology..... | 11 |
| 2. EPISTEMIC INJUSTICE - WHAT DOES IT MEAN TO BE HARMED AS A KNOWER?..... | 13 |
| 2.1 Testimonial Injustice – Power, Knowledge and Credibility..... | 15 |
| 2.1.1 Two Models of Testimony..... | 19 |
| 2.1.2 Fricker’s Perceptual Model of Testimony..... | 22 |
| 2.2 Stereotypes as Heuristics in Cognition..... | 25 |
| 2.2.1 Stereotypes and Credibility Judgements in Fricker..... | 28 |
| 2.2.2 Prejudicial Stereotypes in Fricker’s Account..... | 30 |
| 2.2.3 Credibility Deficit versus Credibility Excess..... | 33 |
| 2.3 Residual Internalizations – An Incomplete Concept?..... | 37 |
| 3. RESIDUAL INTERNALIZATIONS REVISITED..... | 40 |
| 3.1 Implicit Biases and Epistemic Injustice..... | 41 |
| 3.1.1 Implicit Bias and Testimonial Injustice – Similarities and Differences..... | 44 |
| 3.1.2 Stereotype Threat and its Link to Epistemic Injustice..... | 48 |
| 3.2 Belief-Like Versus Associative States of Implicit Bias..... | 52 |
| 3.2.1 Aliefs..... | 54 |
| 3.2.2 Aliefs in the Context of Epistemic Injustice..... | 57 |
| 3.3 The Improvement of Fricker’s Account through Implicit-Bias Research..... | 60 |
| 4. NORMATIVE QUESTIONS RELATED TO TESTIMONIAL INJUSTICE AND IMPLICIT BIAS..... | 61 |
| 4.1 Bias-Related Doubt..... | 61 |
| 4.1.1 Irrelevant Influences on Belief..... | 64 |
| 4.1.2 Bias-Related Doubt, Irrelevant Influences and Credibility..... | 68 |
| 4.2 Epistemic Responsibilities and Obligations..... | 69 |

| | |
|--|------------|
| 4.2.1 Epistemic Responsibilities and Implicit Bias..... | 70 |
| 4.2.2 Medina on Epistemic Responsibility and Culpable Ignorance..... | 74 |
| 4.2.3 Fricker’s Concept of “Epistemic Agent-Regret”..... | 77 |
| 4.3 Remedies against Testimonial Injustice and Implicit Prejudice..... | 79 |
| 4.3.1 Fricker’s Model of the Virtue of “Testimonial Justice”..... | 80 |
| 4.3.1.1 Alcott’s Criticism..... | 83 |
| 4.3.1.2 Virtues, Habituation, and Implicit Bias..... | 84 |
| 4.4 Ameliorated Testimonial Justice and Epistemic Friction..... | 88 |
| 4.4.1 “Meta-Lucidity” as a Virtue of the Oppressed..... | 90 |
| 4.4.2 Structural Remedies against Testimonial Injustice..... | 93 |
| 5. CONCLUSION..... | 98 |
| LITERATURE | 103 |

Danksagung

Ich möchte mich herzlich bei allen Personen, die mich bei der Entstehung dieser Arbeit unterstütz, gestützt und begleitet haben, bedanken.

Für die kompetente fachliche und inhaltliche Betreuung bedanke ich mich bei Herrn Univ.-Prof. Dr. Martin Kusch, dessen kontinuierliche Lektüre und Rückmeldung zu jedem einzelnen Kapitel Denk- und Schreibprozesse in Gang setzten und mir ermöglichten, meinen Stil sowie mein philosophisches Verständnis weiterzuentwickeln. Die Gespräche in der Vorphase und während der Entstehung dieser Arbeit waren äußerst hilfreich und eröffneten mir wichtige Pfade hinsichtlich inhaltlicher und methodischer Aspekte.

In weiterer Instanz bedanke ich mich bei Herrn Univ.-Prof. i.R. Dr. Richard Heinrich, den ich zwei lehrreiche, anspruchsvolle sowie auch humorvolle Jahre als Tutorin unterstützen durfte.

Von ganzem Herzen danke ich meinen beiden Leserinnen: Meiner Mama, Frau Susanne Riegler-Mollang, die mir durch ihre kontinuierliche Lektüre und ihre Verbesserungen und Anmerkungen nicht nur bei dieser Arbeit, sondern mein ganzes Studium unter die Arme gegriffen hat. Vor allem jedoch danke ich ihr für ihr stets offenes Ohr, den immerwährenden Zuspruch, das Vertrauen in mich und meine Fähigkeiten und die Tatsache, mir bereits als Kind gute Bücher gekauft und mich fürs Lesen und Denken begeistert zu haben.

Karoline Paier, die diese Arbeit bereits mit mir besprochen hat, als die in ihr enthaltenen Prämissen noch eher Vermutungen, als klaren Gedanken glichen. Danke für deine ständige, ungebrochene, wertvolle Unterstützung - in freundschaftlicher und philosophischer Hinsicht.

Meinem Papa, Hans Riegler, danke ich von Herzen für die stetige Unterstützung, sowohl in persönlicher, emotionaler als auch finanzieller Hinsicht. Auch ihm bin ich überaus dankbar, mich stets in meiner Studienwahl und meinen Aspirationen bestärkt zu haben und mir Georg Danzer, Thomas Bernhard, das Sams und andere Denkanstöße ans Herz zu legen. Zudem danke ich Claudia Gregorc für ihre stets lieben Worte, Gesten und Taten.

Je remercie Antoine Quintard, qui me soutient, m'encourage, et m'accompagne depuis quatre ans. Tu rends ma vie meilleure.

Ich danke all meinen Freundinnen, die mein Leben bereichern: Livia Gruber, Raphaela Tiefenbacher, Nastasja Ronck, Sabrina Peer, Flora Löffelmann, Julia Hager, Constanze Winkler, uvw., sowie meiner ganzen Familie.

Tristan Bath möchte ich danken, sich für mich durch das Dickicht der englischen Beistrich-Regel Setzung geschlagen zu haben.

Zu guter Letzt danke ich meinen Großeltern, Susanne und Paul Mollang, denen ich diese Arbeit widmen möchte. Ihr Optimismus, ihre Herzlichkeit, Großzügigkeit und Fürsorglichkeit haben mir die schönsten Jahre beschert und tun es noch heute.

Abstract

This master's thesis will be about "epistemic injustice" and more specifically about Miranda Fricker's groundbreaking account of testimonial injustice. After a thorough assessment of Fricker's basic assumptions of testimonial injustice I will propose several ameliorations to Fricker's account, which are all related to a specific aspect of her theory. Both Fricker's model of perceptual testimony and her use of stereotypes as heuristics for judgements give rise to the idea that a) testimonial exchange and consequently b) credibility judgements are not only regulated by *reflective* but also by *affective* cognitive mechanisms. In cases of testimonial injustice for instance, identity prejudices are often not actively endorsed but operate on a more unconscious level, a phenomenon Fricker labels "residual internalizations". In my thesis I will argue that the influence of such passively endorsed prejudices on credibility judgement should be taken seriously. Furthermore, I am going to show that Fricker's development of this issue is inconsistent in its implementation. I will assert that a theory of epistemic injustice should take research on implicit biases into consideration in order to be able to better explain how identity prejudices operate on an unconscious level. Additionally, I will demonstrate that remedies against testimonial injustice are rendered more effective when one includes implicit-bias research into Fricker's account of epistemic injustice.

Zusammenfassung

Die vorliegende Masterarbeit widmet sich dem Thema der „epistemischen Ungerechtigkeit“. Hierbei steht besonders Miranda Frickers Theorie der „Zeugnis-Ungerechtigkeit“ im Vordergrund. In einem ersten Schritt wird Frickers Theorie vorgestellt und einer eingehenden Analyse unterzogen. Des Weiteren werde ich einige Verbesserungen vorschlagen, die alle einen spezifischen Aspekt von Frickers Ausführungen aufgreifen. Sowohl Frickers Auffassung, dass Zeugenschaft auf Perzeption basiert („perceptual testimony“) als auch ihr Verständnis von Stereotypen, die kognitive Hilfen in Entscheidungs- und Denkprozessen darstellen, legen folgende Annahme nahe: Beurteilungen über die Kreditibilität einer Sprecherperson werden nicht nur durch reflexive, sondern auch durch affektive, emotionale Mechanismen generiert. Vorurteile in Bezug auf eine bestimmte Gruppe, die zu Fällen von epistemischer Ungerechtigkeit führen, werden beispielsweise häufig nicht bewusst vertreten, sondern sind Produkt unbewusster Prozesse. Fricker bezeichnet dieses Phänomen als „residuelle Internalisierungen“. In meiner Masterarbeit werde ich das Verhältnis zwischen unbewusst operierenden Vorurteilen und epistemischen Ungerechtigkeiten besonders in den Fokus nehmen und argumentieren, dass Frickers Ausführungen hierzu unzureichend sind. Gleichzeitig schlage ich vor, dass das Heranziehen von Literatur aus der „Implicit-Bias“ Forschung diese Ungenauigkeit tilgen kann. Ziel der Arbeit ist es aufzuzeigen, dass Theorien über epistemische Ungerechtigkeiten davon profitieren können, sich eingehender mit dem Phänomen impliziter Vorurteile auseinanderzusetzen.

1. Introduction

We live in an era, in which women who speak up about sexual harassment still experience insufficient uptake and LGBTQ asylum seekers are deported because authorities do not believe that they are gay and consequently persecuted in their respective countries. When considering these examples, one sadly has to draw the conclusion that not every voice is granted the same degree of agency and power. In other words, when providing testimony about their identity, experiences and perspectives some people are believed and some are not. Some people are epistemically acknowledged whilst others are ignored, downgraded or repudiated.

When giving testimony, members of marginalized groups are particularly likely to be believed less because of negative prejudices regarding the social group to which they belong. They are confronted with distinctively epistemic harm: their testimony will be taken less seriously and thereby, they are excluded from the generation of knowledge. Injustice is done to these marginalized groups by denying them the status of informed knowers.

In the field of philosophy Miranda Fricker developed a theory that is concerned with exactly this kind of injustice, namely “epistemic injustice”. In my thesis I will take a closer look at Fricker’s groundbreaking concept of epistemic injustice, outlined in her book *Epistemic Injustice - Power and the Ethics of Knowing*.¹

Fricker divides cases of epistemic injustice into two categories: “testimonial” and “hermeneutical” injustice. Testimonial injustice affects speakers when they suffer from a credibility deficit of a certain kind: their testimony is believed less than it should be, and this is because of “identity prejudice” concerning the social group to which they belong. Hermeneutical injustice occurs when the disadvantaged in society are denied the hermeneutical resources to make sense of their social experience.² Being denied the status of an informed and credible epistemic agent can have serious ramifications for both the individual and the epistemic community as a whole.

In my thesis I aim to present a thorough analysis of Fricker’s main arguments concerning epistemic injustice. By doing so, I will put special emphasis on testimonial injustice. In my

¹ Miranda Fricker, *Epistemic Injustice. Power & the Ethics of Knowing*. (New York: Oxford University Press, 2007).

² Cf., *ibid.*, 1.

view, Fricker's account constitutes an important philosophical analysis that investigates wrongful ascriptions of credibility and thus provides tools to pin down epistemic imbalances. By doing so, Fricker not only contributes important findings to social epistemology, but also makes it possible to describe and analyze a phenomenon that we encounter in our day-to-day epistemic environments. This is so since epistemic injustice is a "[...] pervasive feature of our social and professional lives."³ Still, my account will not be entirely uncritical. I will identify some problems within Fricker's theory and aim to provide the necessary improvements to remedy these shortcomings.

As stated above, testimonial injustice only occurs when the hearer's judgement is afflicted by prejudices concerning the social group of the speaker. According to Fricker, these prejudices can either be actively endorsed or they can operate on a more unconscious level. The latter case she identifies as "residual internalizations". Residual internalizations describe instances in which prejudicial images from the social imagination persist in a hearer's patterns of credibility judgements although their content is not in line with the hearer's beliefs. Fricker emphasizes that due to unconsciously operating identity prejudices, instances of testimonial injustice are likely to happen on a regular basis.⁴ "Certainly we may sometimes perpetrate testimonial injustice because of our beliefs; but the more philosophically intriguing prospect is that we may very frequently do it in spite of them."⁵

The concept of residual internalizations will become decisive for the further development of my argument. I claim that while it is a crucial notion for Fricker's account her implementation of it lacks sufficient depth. Although Fricker presents a way to reflect upon instances of testimonial injustice that come about by implicit identity prejudices, she does not develop this point to its necessary extent. In the course of her argument it is rarely mentioned again. Moreover, the examples she presents to illustrate instances of testimonial injustice are built on identity prejudices that have been actively endorsed. One of this thesis' core objectives thus is to argue that Fricker's offhand usage of implicit prejudice contradicts her assertion that cases of testimonial injustice happen all the time insofar as they are products

³ Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr, "Introduction to the Routledge Handbook of Epistemic Injustice", in: *The Routledge Handbook of Epistemic Injustice*, ed. Kidd, Ian James, (New York: Routledge, 2017), 1-9, 1.

⁴ Cf., Fricker, *Epistemic Injustice*, 37-38.

⁵ *Ibid.*, 36.

of these prejudice.

The detected neglect concerning implicit prejudice in Fricker's account also becomes evident when we look at the antidote she provides to counteract instances of epistemic injustice. Fricker refers to virtue ethics and proposes, "[...] that we need to develop our reflexivity about identity power and identity prejudice, and we need to create mechanisms by which we can neutralize the impact of prejudice in our credibility judgements."⁶ What Fricker has in mind is a virtue called "testimonial justice": it is meant to block or minimize the influence of prejudicial stereotypes in credibility judgements. This virtue is gained through a reflexive critical awareness.⁷ As stated above, this solution seems inconsistent with Fricker's proposition that acts of testimonial injustice are in many cases products of unconsciously-endorsed identity prejudices. Linda Alcoff has also raised this point: "If identity prejudice operates via a collective imaginary, as she [Fricker] suggests, through associated images and relatively unconscious connotations, can a successful antidote operate entirely as a conscious practice?"⁸

My criticism therefore is twofold. First, I will address the following contradiction: Fricker deems cases in which testimonial injustice results from the implicit prejudice to be philosophically especially intriguing. And yet she primarily focuses on cases in which prejudice is actively endorsed. Second, and relatedly, I will argue that this contradiction also resonates with the way Fricker aims to remedy testimonial justice, namely through critical reflection.

I propose that Fricker should take implicit-bias research into consideration to enable a more in-depth account of how residual prejudices might cloud our credibility judgement. Additionally, I will suggest that correctional measures against epistemic injustice should include considerations regarding the possible influence of implicit bias on our credibility judgements. The core purpose of this thesis is to investigate to what extent features of implicit-bias research could complement and improve Fricker's account of epistemic testimonial injustice.

⁶ Linda Alcoff, "Epistemic Identities", in: *Episteme. Vol. 7. Issue 2.* (2010), 128-137, 132.

⁷ Cf., Fricker, *Epistemic Injustice*, 91.

⁸ Linda Alcoff, "Epistemic Identities", 132.

1.1 Development and Methodology

The second chapter will provide a thorough analysis of how epistemic testimonial injustice arises. In the realm of this assessment, I will first delineate how Fricker explains testimonial exchange based on her perceptual model of testimony. Moreover, I will touch upon practices of stereotyping, since they are a basic element in the genealogy of epistemic injustice. I shall introduce a general account of stereotypes taken from literature in the field of social psychology. In doing so, the thesis briefly explores the question to what extent stereotypes are able to operate as heuristic shortcuts for (credibility) judgements. I will then compare this general account of stereotypes and prejudices with Fricker's model of identity prejudices. These identity prejudices – when negative – represent the engine for the emergence of epistemic injustice. Using work by José Medina, I shall show that positive identity prejudices also lead to testimonial injustice. The second chapter will conclude with an introduction to Fricker's notion of residual internalizations.

In the third chapter, I will focus on implicit-bias research and its utility for Fricker's account. It will be shown that theories of epistemic injustice could greatly benefit from insights acquired in implicit-bias research. I argue that these findings allow reflecting upon the causes and consequences of epistemic injustice and coming up with possible remedies against it in a more thorough manner. At the outset, the notion of implicit bias will be introduced by mainly focusing on work by Jennifer Saul and Michael Brownstein. In a further step I will introduce Tamar Gendler's concept of "alief" and argue that the integration of this term can also be beneficial for a discussion about epistemic injustice.

Chapter four will pose some normative questions related to the subject. First, I shall ask whether the emerging data on implicit bias suggests that we have to become skeptics about our beliefs. If this is the case, how should this question be treated in the realm of Fricker's account on epistemic injustice? I will continue with an investigation of epistemic obligations, responsibility and culpability by referring to Medina, Fricker and an article by Natalia Washington and Daniel Kelly. The questions addressed in this section will be the following: does acting responsibly require that we are always fully aware of our beliefs and their formation? To which degree can we be held responsible for our implicit biases when they influence our judgements in a manner that is not easily traceable? My aim is to show that it is possible to be regarded accountable for one's implicit biases, even when one is not aware

of the impact they have on our judgements. Fricker's concept of "epistemic agent-regret" will help me to formulate this point.

Deliberations about epistemic obligations and responsibilities will lead me, finally, to discussing possible remedies against epistemic injustice. I will present, and to some extent critically assess, Fricker's virtue-theoretical perspective on testimonial justice and its respective training. My objective is to come up with an ameliorated form of testimonial justice that (a) considers the possibility of implicit prejudice, (b) focuses on both perpetrators and victims of testimonial injustice and, (c) emphasizes that testimonial justice should not only be regarded as an individual, but also a structural virtue.

I support Fricker's proposition that critical self-awareness is decisive to successfully counteract testimonial injustice. And yet I will also stress the importance of habituation when it comes to the development of virtues. This will be underlined by drawing on Linda Zagzebski's thoughts regarding habit and virtue. Emphasizing the role of habit in the development of virtues allows me to address the question as to how we could correct for implicit prejudices. I will show that the habituation of egalitarian motives could help to block the influence by implicit bias.

Moreover, we should not only ask what perpetrators of testimonial injustice can do to improve epistemic relations. Rather, we also need to include the victims' perspective and ask whether there is a way for them to actively defend against acts of testimonial discrimination.⁹ I suggest that Medina's virtue of "meta-lucidity" can be seen as a way of reattributing agency to people who are downgraded as epistemic protagonists. Finally, in drawing on Elizabeth Anderson's paper "Testimonial Justice as a Virtue of Social Institutions"¹⁰, I will conclude that testimonial justice should not only be regarded as an individual, but also as a structural virtue.

In conclusion, the objective of this thesis is to point out that implicit prejudice is one core factor for the emergence of testimonial injustice. I argue that Fricker's theory would benefit from taking into consideration research on implicit bias in order to clarify the influence of implicit prejudice on our credibility judgements.

⁹ Martin Kusch, Review: "Miranda Fricker: Epistemic Injustice: Power and the Ethics of Knowing", in: *Mind*, Volume 118, Issue 469, (2009), 170–174.

¹⁰ Elizabeth Anderson, "Epistemic Justice as a Virtue of Social Institutions", in: *Social Epistemology*. 26:2, (2012), 163-173.

2. Epistemic Injustice - What does it mean to be harmed as a knower?

Following Fricker's idea that there is a distinctively epistemic kind of injustice, the following chapter is going to explore how the phenomenon of epistemic injustice arises, how it operates and what harm it causes both to an epistemic community and to the affected individual. I will first delineate Fricker's account of the two forms of epistemic injustice: testimonial and hermeneutical injustice.¹¹ This thesis will then mainly focus on testimonial injustice.

Epistemic injustice describes a special form of discriminatory action in which individuals are not taken seriously in their status as epistemic agents able to acquire knowledge, justified belief and understanding.¹² According to Fricker, it is natural to first intuitively associate epistemic injustice with distributive unfairness regarding epistemic goods like education and information. Social agents are understood as being equally entitled to access various commodities, and only some of them are of an epistemic nature.¹³ Fricker asserts that there is nothing distinctively epistemic about this association as "[...] it seems largely incidental that the good in question can be characterized as an epistemic good."¹⁴

Therefore, Fricker aims to define two forms of epistemic injustice that are intrinsically and unmistakably epistemic. In her 2017 paper for the *Routledge Handbook of Epistemic Injustice* she retrospectively defines this objective as follows: "My chief purpose in invoking the label was to delineate a distinctive class of wrongs, namely those in which someone is ingeniously downgraded and/or disadvantaged in respect of their status as an epistemic subject."¹⁵ She does so by coining the terms of testimonial and hermeneutical injustice. Instances of testimonial injustice occur when a specific form of prejudice prompts a hearer to attribute a speaker with a reduced status of credibility. Hermeneutical injustice arises at a prior stage and characterizes instances where the powerless in society have less access to interpretative resources and are therefore at a disadvantage when it comes to making sense

¹¹ Cf., *ibid.*, 1.

¹² Cf. Katherine Puddifoot, "Epistemic Discrimination", in: *The Routledge Handbook of the Ethics of Discrimination*, ed. Kaspar Lippert-Rasmussen, (New York: Routledge, 2017), 54-67, 57.

¹³ Cf. Fricker: *Epistemic Injustice*, 1.

¹⁴ *Ibid.*

¹⁵ Fricker, "Evolving Concepts of Epistemic Injustice", in: *The Routledge Handbook of Epistemic Injustice*, ed. Kidd, Ian James, (New York: Routledge, 2017), 53-61, 53.

of their social experience.¹⁶ In her book review of Fricker's *Epistemic Injustice – Power and the Ethics of Knowing*, Linda Alcoff distinguishes the two instances of epistemic injustice as follows: “[...] Whereas testimonial injustice wrongly responds to speech, hermeneutical injustice preempts speaking. Testimonial and hermeneutic injustice are in this way distinct on her [Fricker's] view, though they can be mutually supportive.”¹⁷

Fricker presents us with the following examples in order to illustrate the two kinds of epistemic injustice: “An example of the first might be that the police do not believe you because you are black; an example of the second might be that you suffer sexual harassment in a culture that still lacks that critical concept.”¹⁸ As we can draw from these two examples, epistemic injustice especially concerns members of stigmatized groups due to social prejudicial stereotypes others apply to them. By being associated with these negative stereotypes, members of marginalized groups are excluded from positions of power. It thus becomes harder for them to accomplish an improvement in their epistemic situation, as they are being constantly denied status as informed knowers. This exclusion from the pooling of knowledge leads to a hybridized form of harm that is both epistemic and ethical in nature. For Fricker, being wronged in one's capacity as a subject of knowledge, means to be wronged in a capacity that is essential to human value.¹⁹ For that reason it is her overarching aim to shed light on the ethical aspects of our most basic epistemic practices: “[...] conveying knowledge to others by telling them, and making sense of our own social experiences.”²⁰

Moreover, instances of epistemic injustice can be performed intentionally or unintentionally by unconsciously endorsing the prejudices that lead to the discriminatory act against the speaker. I will argue that although Fricker asserts that cases of testimonial injustice in which identity prejudices are not actively endorsed are more intriguing from a philosophical perspective, she does not develop this argument to its necessary extent. This leads to certain inconsistencies in her account.

In what follows I will propose improvements necessary to save Fricker's theory. However, it is crucial to begin with a thorough investigation of her arguments regarding testimonial

¹⁶ Cf. Fricker, *Epistemic Injustice*, 1.

¹⁷ Cf. Linda Alcoff, “Epistemic Identities”, 129.

¹⁸ Fricker: *Epistemic Injustice*, 1.

¹⁹ Cf. *Ibid.*, 5.

²⁰ Cf. *Ibid.*, 1.

injustice in order to point out the specific shortfalls within her account that this thesis wants to highlight.

2.1 Testimonial Injustice – Power, Knowledge and Credibility

“What I know is this: they didn’t believe me. Twelve people and none of them took me at my word.”

Tayari Jones (*An American Marriage*)

Fricker commences her analysis on testimonial injustice with an example. In Anthony Minghella’s screenplay for *The Talented Mr. Ripley*, Marge Sherwood, the young fiancée of Dickie Greenleaf, has good reason to believe that Dickie’s recent disappearance is not in fact the result of a tragic accident. Rather, she is convinced that his best friend, Tom Ripley, has murdered Dickie. When she confronts Dickie’s father with her suspicion, Greenleaf repudiates her assumption with the following sentence: “Marge, there is female intuition, and then there are facts.”²¹ For Greenleaf, Sherwood reacts too emotionally to his son’s disappearance and therefore he judges her incapable of objectively assessing the situation.²² This example from *The Talented Mr. Ripley* graphically illustrates what Fricker has in mind when she refers to instances of testimonial injustice.

Before we can focus on the mechanisms and precise implications of testimonial injustice, we first need to take a closer look at the broader background. Let us do so by returning to the fictional example Fricker presents us with. Greenleaf’s silencing of Sherwood clearly involves a demonstration of social power - gender power - to be more precise.²³ But how is social power related to Fricker’s notion of testimonial injustice? First, it is crucial to understand that as social agents we have the capacity to influence how things work out in the social world. According to Fricker, this social power can operate both *actively* and *passively*.²⁴

²¹ Minghella, Anthony: *The Talented Mr. Ripley* – Based on Patricia Highsmiths Novel. (London: Methuen, 2000), 130.

²² Cf. Fricker, *Epistemic Injustice*, 9.

²³ Cf., *ibid.*, 9.

²⁴ Cf., *ibid.*

In order to illustrate this idea, she gives the example of a traffic warden, who has the *active* social power to impose a fine when she sees a car that has been parked incorrectly. However, this power also operates *passively*, whenever the warden's ability to impose a fine affects a person's parking behavior.²⁵ Consequently, power exists even when it is not actively executed. This fact contradicts Foucault's famous claim that "[...] power exists only when it is put into action."²⁶ Moreover, Fricker distinguishes between *agential* and *structural* power. In the latter variation of power, there is no particular agent exercising it but it is rather "[...] so thoroughly dispersed through the social system that we should think of it as lacking a subject."²⁷ However, Fricker claims that the differentiation between agential and structural power should not be made too strongly. Even in agential operations of power, it is already a structural phenomenon, for power always depends upon coordination with other social agents and social contexts.²⁸ This is an important aspect of social power that we ought to keep in mind.

What other distinctive features, besides being an agential and structural phenomenon that is executed either actively or passively, does social power have? One classical response to this question is that power has the capacity to impede someone's objective interests.²⁹ In Fricker's view this conception of power seems both too negative and narrow. What she proposes instead is a more neutral formulation, namely that social power has the capacity to effect social control.³⁰ The working conception of social power Fricker provides on that account is the following: "*a practically socially situated capacity to control other's action, where this capacity may be exercised (actively or passively) by particular social agents, or alternatively, it may operate purely structurally.*"³¹

One could concede that this conception of social power allows for the interpretation that the execution of power is not necessarily bad for anyone. However, as Fricker puts it, "[...] wherever power is at work, we should be ready to ask who or what is controlling whom, and

²⁵ Cf., *ibid.*

²⁶ Michel Foucault, "How is Power Exercised?", trans. Sawyer, Leslie from Afterword in Dreyfus, H.L. and Rabinow, P: *Michel Foucault: Beyond Structuralism and Hermeneutics*. (Hemel Hempstead: Harvester Press, 1982), 219.

²⁷ Cf., Fricker, *Epistemic Injustice*, 10-11.

²⁸ Cf., *ibid.*

²⁹ Cf., *ibid.*, 12.

³⁰ Cf., *ibid.*, 13.

³¹ *Ibid.*, 13. (Original quote in italics as well.)

why.”³² This seems legitimate, especially when we consider in which context Fricker employs her considerations about social power - namely the communicative imbalance that arises through the exploitation of being in a privileged position of this kind of power. This becomes even more evident when we continue assessing Fricker’s text. One further assumption she makes in order to define what she has in mind when talking about a specific kind of power that regulates social interaction is the fact that power also draws on a certain imaginative social co-ordination.³³

There can be operations of power which are dependent upon agents having shared conceptions of social identity – conceptions alive in the collective social imagination that govern, for instance, what it is or means to be a woman or a man, or what it is or means to be gay or straight, young or old and so on. Whenever there is an operation of power that depends in some significant degree upon such shared imaginative conceptions of social identity, then *identity power* is at work.³⁴

It is crucial to understand that identity power plays an important role in defining what Fricker means when she talks about epistemic injustice and, more specifically, testimonial injustice.

Keeping in mind the definition of identity power Fricker provided us with, let us again take a closer look at her opening example. The mechanisms Greenleaf used to silence Sherwood are strongly connected to identity power. The kind of identity at stake in this example is gender. Gender identity power occurs when, for instance, “[...] a man (possibly unintended) makes use of his identity as a man to influence a woman’s actions [...]”.³⁵

Consider again how Herbert Greenleaf patronizes Marge Sherwood by *actively* dismissing her suspicion. He silences her by exercising gender identity power, invoking “women’s intuition” as the supposed opposite to (male) “reason”. However, one could also argue that the manner in which Herbert Greenleaf silences Marge Sherwood could be understood as a *passive* execution of identity power. As Fricker states, Marge “[...] might already be silenced by the mere fact that he is a man and she a woman.”³⁶ The story is set in the 1950s, so it is likely that both an active and passive execution of gender power plays a role in the injustice done to Marge, as stereotypical gender roles were more directly governing people’s interpersonal relations. The more important aspect, however, is not the question as to whether

³² Ibid., 14.

³³ Cf., *ibid.*

³⁴ Ibid.

³⁵ Ibid.

³⁶ Ibid., 15.

the kind of injustice Marge is confronted with was caused by a direct discriminatory action or by mere social hierarchy. Rather, what should interest us is the question as to what facilitates this hierarchy? For Fricker, the answer lies in the imaginative social co-ordination that regulates identity power.³⁷ In other words, we all share collective conceptions of social identities mediated through stereotypes. In Sherwood's case, the negative stereotype Greenleaf applied to her is that women are intuitive and therefore less inclined to rational thinking. Moreover, Sherwood is immersed in an intense state of grief, a fact that according to Greenleaf renders it impossible for her to think or act rationally.

Furthermore, Fricker asserts, "[...] that the operation of identity power does not require that either party consciously accept the stereotype as truthful."³⁸ Even if Sherwood were aware of the stereotype applied to silence her, it would still be likely that she would nevertheless be silenced by it, as Fricker outlines in the following quote:

The conceptions of different social identities that are activated in operations of identity power need not be held at the level of belief in either subject or object, for the primary *modus operandi* of identity power is at the level of the collective social imagination. Consequently, it can control our actions even despite our beliefs.³⁹

This aspect of social identity that is not consciously, reflectively endorsed will become crucial in the course of Fricker's argument. We will encounter it both in Fricker's concept of "residual internalizations", which will be presented in the final passage of this chapter and in the proposed ameliorations to Fricker's account, namely implicit-bias research.

For now, let us return to the question of how Fricker construes testimonial injustice through drawing on identity power and stereotypes. What we can infer so far is that Fricker regards identity power to be an integral part of testimonial exchange "[...] because of the need for hearers to use social stereotypes as heuristics in their spontaneous assessments of their interlocutor's credibility."⁴⁰ This use of stereotypes might be entirely innocuous, or it can be harmful, depending on the stereotype. Yet, if we assume that the stereotype has prejudicial content that concerns the speaker then this leads to two consequences, both of

³⁷ Cf., *ibid.*

³⁸ *Ibid.*

This aspect about stereotypes will also become crucial in the next chapter that is concerned with implicit-bias research.

³⁹ *Ibid.*

⁴⁰ *Ibid.*, 16-17.

which are central for defining what testimonial injustice is. First, prejudicial content will cause an epistemic dysfunction in the testimonial exchange. Owing to the prejudicial stereotype the hearer will attribute less credibility to the speaker than the speaker deserves. Second, the reduced credibility judgement isn't just epistemically, but also ethically bad, as the speaker "[...] is wrongfully undermined in her capacity as a knower [...]", meaning in a capacity that is central to human value.⁴¹ What we are presented with at this point is a first definition of what Fricker calls testimonial injustice: it is the kind of injustice in which someone is specifically wronged in her status as a knower due to a prejudice concerning the social group she belongs to.

In what follows, I will present a general account of stereotypes drawn from the social psychology literature, in order to gain a better understanding of how social stereotypes function as cognitive aids and heuristics. I shall then compare this general account with the kind of negative prejudice Fricker construes, namely "identity prejudice".⁴² Furthermore, we have to clarify what type of prejudicial dysfunction leads to testimonial injustice. According to Fricker, this kind of dysfunction is credibility deficit.⁴³

Before we continue our analysis of Fricker's account of testimonial injustice, it is crucial to gain a better understanding of the way testimonial exchange is regulated in her theory. Let us therefore briefly investigate her perceptual model of testimony, in order to comprehend how it differs from standard models of testimony and why it is important for her account of epistemic injustice.

2.1.1 Two Models of Testimony

The phenomenon Fricker calls "testimonial injustice" is not solely confined to testimonial exchange, but allows for a view in which testimony is construed more broadly. However, the basic harm caused by testimonial injustice is to undermine the speaker in her capacity *qua* knower. That is why cases of telling, where telling is understood as the conveying of knowledge, are at the center of Fricker's analysis.⁴⁴ To better understand this concept, let us

⁴¹ Cf., *ibid.*, 17.

⁴² Cf., *ibid.*, 27.

⁴³ Cf., *ibid.*, 17.

⁴⁴ Cf., *ibid.*, 60.

therefore briefly revisit how standard models in philosophy treat the epistemology of testimony and then compare these models to Fricker's account of testimony, which differs from the classical literature.⁴⁵

Epistemologies of testimony are divided into two broad categories: inferential and non-inferential accounts.⁴⁶ Evidently, as Fricker concedes, there is room for more diverse views, however, "[...] a key motivation for any will be the author's inclination *vis-à-vis* inferentialist and non-inferentialist pictures of the obligations upon a hearer if she is to gain knowledge from her interlocutor."⁴⁷

Inferentialists argue that in order for a hearer to gain knowledge *p* from a speaker, they must "[...] in some way (perhaps very swiftly, perhaps even unconsciously) rehearse an argument whose conclusion is *p*."⁴⁸ The hearer therefore makes an inference that somehow confirms that what she has heard is right, because the speaker seems reliable, sincere and trustworthy. Alternatively, non-inferentialists assert that our testimonial exchange with others is based on a more spontaneous reception.⁴⁹ Non-inferentialists thus argue "[...] for some sort of default of credulity or acceptance of what others tell us."⁵⁰ Referring to Thomas Reid⁵¹, Fricker claims that we make two default assumptions: we generally regard speakers as honest, and hearers as motivated to trust speakers.⁵²

Drawing from this short introduction, it seems as if we have to decide which epistemology of testimony, inferentialism or non-inferentialism, we want to endorse:

One story presents the hearer as gaining knowledge only if she rehearses an appropriate inference. The other story seems to present the hearer as gaining knowledge by way of one or another default of uncritical receptivity such that he is entitled to accept what she is told without exercising any critical capacity.⁵³

Which of the two accounts should we consult when we want to analyze testimony? According

⁴⁵ In doing so, the thesis will primarily refer to Fricker's presentation of the debate.

⁴⁶ Cf., *ibid.*

⁴⁷ *Ibid.*, 61.

⁴⁸ *Ibid.*

⁴⁹ Cf., *ibid.*, 62.

⁵⁰ *Ibid.*

⁵¹ Thomas Reid was the first to argue for a default of accepting what we are told owing to the twin principles of *credulity* and *veracity*; see Reid, Thomas: *An Inquiry into the Human Mind on the Principle of Common Sense.*, 1764.

⁵² Cf., Fricker, *Epistemic Injustice*, 62.

⁵³ *Ibid.*

to Fricker, “[...] the shortcoming of each is the allure of the other.”⁵⁴ Let us therefore take a closer look at the benefits and drawbacks of both the inferentialism and non-inferentialism Fricker presents in order to grasp how she designs her perceptual model of testimony.

Inferentialism is designed to ease doubts about our justifications for accepting what others tell us. Following Fricker’s analysis of inferentialist accounts, this will mostly be done by inductive arguments, “[...] for instance, an argument about the individual speaker’s past reliability on these matters, or about the general reliability of people like that about things like this.”⁵⁵ Consider a situation where someone new to a town asks a person whom she knows to be a local resident for directions to the nearest train station. Inferentialists would argue that the woman’s knowledge that the speaker she addresses has lived in the town for many years now serves her as a premise in an inference, the conclusion of which prompts the woman to regard the speaker as a reliable source of information. Critiques of inferentialism, however, claim that the woman’s reasoning for believing what the local has told her comes about in a more spontaneous manner. In other words, non-inferentialists say that we can rationally trust speakers without making an inference. Fricker also endorses this view. According to her, the picture inferentialists draw of testimonial exchange “[...] does not match our everyday phenomenology of informal testimonial exchange, which presents learning something by being told as distinctly *un-laborious* and spontaneous.”⁵⁶ Representatives of inferentialist accounts respond to this criticism by saying that the hearer will assess the argument she has been presented with *readily* and *easily*.⁵⁷ But, as Fricker argues, the more inferentialists insist that the inference is being executed in an intellectually *un-laborious* manner, the more their model becomes obsolete.⁵⁸

For Fricker non-inferential accounts therefore seem better equipped to describe how our every-day testimonial exchanges function:

In the absence of cues for doubt, we surely accept most of what we are told without going in for any active critical assessment, and so our experience as hearers can seem to be that we are trustful unless and until some prompt for doubt is picked up on.⁵⁹

⁵⁴ Ibid.

⁵⁵ Ibid., 63.

⁵⁶ Ibid.

⁵⁷ Cf., *ibid.*

⁵⁸ Cf., *ibid.*

⁵⁹ Ibid., 64.

When we assess our example about the woman asking a local for directions from a non-inferentialist point of view, advocates of this account would argue that the woman unreflectively accepts what the man is telling her, without making an inference based on his status as a local resident. According to non-inferentialists, this unreflectiveness is underlined by the fact that if the woman does have reason to doubt what the local is telling her, she could experience an intellectual “shift of gear”, “[...] out of that unreflective mode and into a reflective, more effortful mode of active critical assessment.”⁶⁰ For instance, she knows that the train station is north of where she is at the moment, but the local points to a direction in the south, which is inconsistent with her knowledge and triggers a cue for doubt. This inconsistency will therefore lead to a critical assessment of the provided testimony. And yet Fricker claims that there is a problem that arises in non-inferential accounts: they present the hearer’s critical faculties as being in snooze-mode when she finds herself in a situation of testimonial exchange.⁶¹ More specifically, Fricker’s criticism concerns the intellectual shift of gear from an uncritical to a critical reception that non-inferentialist accounts want to put forward. In her opinion, the more accurate characterization would be one that presents it “[...] as a shift from unreflective to reflective modes on the part of the hearer, where *either* mode is one in which the hearer may give a critical reception to her interlocutor’s word.”⁶² Therefore, what Fricker wants to achieve with the account of testimony she develops, is a model that is critical yet unreflective and non-inferential.

2.1.2 Fricker’s Perceptual Model of Testimony

Fricker claims that what is needed is a rational sensitivity that enables the hearer to critically assess what she has been told without the need to make an active inference.⁶³ She describes this rational sensitivity as a certain sort of social perception. But what kind of perceptual capacity would this be? The answer can be retrieved from the following quote:

In order for the hearer to, so to speak, see his interlocutors in epistemic colour, the

⁶⁰ Ibid.

⁶¹ Cf., *ibid.*, 65.

According to Fricker, this problem applies to the two non-inferentialist default accounts she presents: Thomas Reid’s twin-principle account of credulity and veracity and Thomas Burge’s account of an a priori justified principle of acceptance.

⁶² Ibid., 66.

⁶³ Cf., *ibid.*, 70.

perceptual capacity would have to be informed by a background ‘theory’ (body of generalizations) not simply of human competences and motivations *per se*, but more specifically, a socially situated ‘theory’ of the competences and motivations of this or that social type in this or that context. [...] That the hearer must trade in social types in this way was why, in the previous chapter, we found stereotypes to be a proper part – indeed, an essential part – of credibility judgements.⁶⁴

To further strengthen her point that the hearer *perceives* her interlocutor in a way that is epistemically charged and that this perception is strongly mediated through social stereotypes, Fricker draws an analogy to virtue ethics. More specifically, she refers to a neo-Aristotelian account of virtue ethics, in which moral cognitivism plays an important role. The form of moral cognitivism Fricker emphasizes advances the idea of moral perception.⁶⁵ Analogously, with her perceptual model of testimony, Fricker presents us with an account that clarifies how testimony ideally should be regulated and how one could conduct responsible hearing. In some way then, her whole book can be understood as a description of what happens when one fails to be a virtuous, responsible hearer.

Let us take a closer look at the analogy Fricker draws between the virtuous agent’s moral perceptual capacity and the virtuous hearer’s testimonial capacity in order to better understand what happens in cases of divergent social perception. The analogy she constructs depends on five points that are closely related.

First, “[...] in the testimonial sphere and in the moral sphere, the model for judgement is perceptual, and so non-inferential.”⁶⁶ Due to a proper moral upbringing and socialization, the moral virtuous agent “[...] sees the world in moral colour.”⁶⁷ Her moral judgements thus come about in a spontaneous and unreflective manner. Accordingly, in the testimonial case, Fricker makes the parallel suggestion that the virtuous hearer’s perception is epistemically enriched. The virtuous hearer has the capacity to spontaneously perceive the speaker as more or less trustworthy, relying on certain background assumptions that are related to different social types in different social contexts.⁶⁸

Secondly, good judgement in the moral and the epistemic spheres is uncodifiable. Neither the morally virtuous agent, nor the epistemically virtuous hearer uses theoretical

⁶⁴ Ibid., 71.

⁶⁵ Cf., *ibid.*

⁶⁶ Ibid., 72.

⁶⁷ Ibid.

⁶⁸ Cf., *ibid.*

generalizations to structure perception beforehand. “She does not apply a *theory*, nor does she *apply* a theory – it isn’t a theory, and she does not apply it.”⁶⁹ Rather, she just sees her interlocutor in a certain light that is morally and epistemically charged. In her book review, Alcoff puts it like this: “Fricker resists the idea that the theories operative in perception can be rendered via a set of rules, or an algorithm; rather, it is like she says, a “sensitivity” that is too complex with too many variable contexts for coding into rules.”⁷⁰

The third and fourth parallel Fricker draws between morally virtuous and epistemically virtuous perception is that both are intrinsically motivating and intrinsically reason-giving.⁷¹ In the moral sphere this means that my moral perception of something I perceive as good or bad or unjust calls for action. Additionally, these moral perceptions are justificationaly charged since the motivation to react to something perceived as, e.g. unjust, is rational.⁷² In this respect we can also see a direct parallel with the testimonial case. The virtuous hearer’s perception of her interlocutor as, for instance, trustworthy provides both the motivation and the justification to accept what the interlocutor is saying.⁷³

Finally, Fricker asserts that in both the testimonial and the moral sphere, “[...] judgement typically contains an emotional aspect that is a proper part of the cognition.”⁷⁴ This aspect will become crucial in the further development of the thesis. In the moral case, the emotional aspect of the cognition is connected to the motivational one. To illustrate this, she draws on Martha Nussbaum’s interpretation of a specific aspect in Aristotelian virtue ethics, depicted in the following quote: “Aristotle holds that the truly good person will not only act well but also feel the appropriate emotions about what he or she chooses.”⁷⁵ Again, Fricker draws the parallel to the testimonial case. She asserts that if a hearer perceives the interlocutor as trustworthy, this attitude towards her is not only a purely intellectual one, but, as Fricker puts it, “[...] such an attitude contains a *feeling* of trust.”⁷⁶ This feeling of trust becomes crucial

⁶⁹ Ibid., 73.

⁷⁰ Alcoff, “Epistemic Identities”, 130.

⁷¹ Cf., Fricker, *Epistemic Injustice*, 73.

⁷² Cf., ibid., 77.

⁷³ Cf., ibid.

⁷⁴ Ibid., 72.

⁷⁵ Martha Nussbaum, “The Discernment of Perception: An Aristotelian Conception of Private and Public Rationality.” in: *Love’s Knowledge: Essays on Philosophy and Literature*. (Oxford: Oxford University Press, 1990) 54-105, 74.

⁷⁶ Fricker, *Epistemic Injustice*, 79.

when the hearer judges the speaker's sincerity because picking up on attitudes like sincerity or trustworthiness implies some kind of emotional response on the part of the hearer.⁷⁷ "Thus, the hearer, like the moral subject, makes judgements that (at least typically) have an emotional aspect."⁷⁸ Fricker makes this point even stronger by affirming that the feeling of trust "[...] is a sophisticated emotional radar for detecting trustworthiness in speakers."⁷⁹

In conclusion, the five points of parallel between the virtuous moral and the virtuous testimonial subject all pave the way for the non-inferential, perceptual conception of testimony Fricker wants to propose. These parallels are able to explain how the phenomenology of testimony can be unreflective and spontaneous yet at the same time critical.⁸⁰ Testimonial sensibility therefore is potentially rational, even when it is unreflective. However, as stated above, Fricker's model of testimony is designed to describe ideal testimonial exchanges, whereas the remaining focus of her book concentrates on instances where virtuous hearing fails.

Therefore, we ought to continue our investigation by concentrating on cases where testimonial sensibility and virtuous hearing are badly trained. Following Fricker, such bad training is caused especially by the influence of prejudicial stereotypes.⁸¹ In the following section I will briefly investigate to what extent stereotypes operate as heuristic aids in human cognition, and more specifically in credibility judgements.

2.2 Stereotypes as Heuristics in Cognition

As we have discussed above, testimonial injustice occurs when the hearer's judgement of a speaker's credibility is corrupted by prejudice. According to Fricker, prejudice can enter credibility judgement in a number of ways, the most important of which is social stereotypes.⁸² And yet it is crucial to understand that in general, Fricker construes stereotypes as a neutral, broad concept and a necessary part of cognition. This implies that stereotypes are not automatically bad or unreliable; rather that they are "[...] a proper part of the hearer's

⁷⁷ Cf., *ibid.*

⁷⁸ *Ibid.*

⁷⁹ *Ibid.*, 80.

⁸⁰ Cf., *ibid.*

⁸¹ Cf., *ibid.*, 81.

⁸² Cf., *ibid.*, 30.

rational resources in the making of credibility judgements.”⁸³ Fricker is not the first to endorse this view. A large body of literature in social psychology supports her assumption that stereotypes play a role in cognition. Let us therefore take a closer look at how stereotypes can work as heuristics in cognition and credibility judgements in order to understand where Fricker’s understanding of stereotypes derives from. After this short excursion into social psychology I will then introduce the notion of identity prejudice, a concept that plays a vital role for the emergence of epistemic injustice in the course of Fricker’s account.

The term “stereotype” was first coined by political journalist Walter Lippmann. Lippmann suggests that stereotypes have an image-like function which means that they work like pictures in our heads: “For the most part, we do not first see, and then define; we define first and then see.”⁸⁴ Thus, the way we perceive the world is strongly influenced by stereotypical images. This view was adapted by social psychology later on. Furthermore, it is not simply that we perceive the world through stereotypical images, but rather that categorization and stereotyping are useful tools to navigate through complex and overwhelming environments.⁸⁵

In social psychology, this kind of approach is summarized under the strand of cognitive models of stereotyping. According to David Hamilton and Jeffery Sherman, the cognitive approach “[...] views stereotypes as belief systems or cognitive structures that can guide information processing, and it examines [...] how their influence on information processing affects perceptions of and interactions with members of stereotyped groups.”⁸⁶ The underlying assumption of this approach is that social perception is complicated, and that “[...] the social perceiver cannot possibly process all of the social information to which he or she is exposed.”⁸⁷ Therefore, stereotypes are used to simplify, encode and classify social situations. Obviously, this applies especially in the evaluation of social groups. Imagine a New York cab driver, an elementary school teacher and a classical musician. Likely, the stereotypical image of a New York cab driver would be that of a man of immigrant decent,

⁸³ Ibid.

⁸⁴ Walter Lippmann, *Public Opinion*. (New York: Harcourt, Brace, 1922), 81.

⁸⁵ Tamar Szabó Gendler: “On the Epistemic Costs of Implicit Bias”, in: *Philos Stud* N°156. (2011), 33-63, 38.

⁸⁶ David Hamilton, Jeffrey Sherman, “Stereotypes”, in, *Handbook of Social Cognition*. (2nd ed. Vol.2), eds., R.S. Wyer Jr. & T.K. Srull, (Hillsdale, NJ: Erlbaum, 1994), 1-68, 2.

⁸⁷ Charles Stangor, Thomas Ford, “Accuracy and Expectancy-Confirming Processing in Orientations and the Development of Stereotypes and Prejudice. In: *European Review of Social Psychology*. Vol. 3:1. (1992), 57-89, 63.

the elementary school teacher will likely be imagined as female and the classical musician as a white man rather than, for instance, a black woman. Of course, we know that there are New York cab drivers that are not of immigrant decent, men who are elementary school teachers and black women who are distinguished classical musicians. And yet the tendency to categorize according to what we regard as typical is persistent, even if we wish not to think or act according to stereotypical images. Doing so, however, does not necessarily imply a bad ethical attitude towards a social group. People are likely to associate certain social groups with certain attributes and certain activities simply because doing so makes it easier to navigate through complex social environments.

A further characteristic of stereotypes is that these associations happen automatically and spontaneously.⁸⁸ Think back to how Fricker describes testimonial exchange as being unreflective and non-inferential, yet critical. In the context of an assessment of social stereotypes, this means that, if the stereotype is accurate and informed by experience, then it helps us to make sense of social environments in a spontaneous, yet critical way.⁸⁹ “In short, rapid, automatic generalization on the basis of categories is fundamental to how we make sense of the world.”⁹⁰ How we process and decode our environment however, is highly dependent on the way we select and arrange the information we are confronted with.

Cognitive approaches to social stereotypes therefore emphasize the role that informational selectivity plays in the formation of stereotypes. The way we assess the world and process information through drawing on stereotypes is clearly a form of selectivity, but this selectivity equally applies to how the stereotype is formed in the first place. Various authors in social psychology literature put forward the idea that perceivers actively seek out information while ignoring other information, which leads to a constant confirmation and perpetuation of the stereotype.⁹¹ This means that stereotype-congruent information is attended to while stereotype-incongruent information is likely ignored. “Under appropriate conditions, stereotype-consistent information may receive more attention than and be better recalled than inconsistent information.”⁹² According to Hamilton and Sherman, this selectivity process

⁸⁸ Cf., Hamilton, Sherman, “Stereotypes”. 40.

⁸⁹ Presupposed that the stereotype is reliable.

⁹⁰ Gendler, “On the Epistemic Costs of Implicit Bias”, 39.

⁹¹ Cf., Stangor, Ford, “Accuracy and Expectancy-confirming Processing in Orientations and the Development of Stereotypes and Prejudice, 63-64.

⁹² Hamilton, Sherman, “Stereotypes”, 37.

demonstrates that one of the major functions of a stereotype is its self-maintenance.⁹³ This aspect will become crucial later on in this thesis when we investigate possibilities to detect and change stereotypical prejudices in order to prevent instances of testimonial injustice.

Therefore, automaticity and informational selectivity are important aspects within the cognitive approach to stereotypes in social psychology. Fricker also includes these two issues into her investigation of the role of stereotypes in credibility judgements. What is decisive for Fricker's train of thought, and consequently for this thesis, is that automaticity and informational selectivity also play an important role in the formation of prejudices. As mentioned above, categorizing a social group in terms of a stereotype about that group does not necessarily imply a bad ethical attitude towards its members. Yet, if the stereotype also contains a bad affective investment besides its cognitive function, it is likely to become prejudiced. In social psychology, prejudices are defined as powerful emotional attitudes consisting of three components: an affective-emotional component, a cognitive component (mediated through the stereotype) and a behavioral component, which likely leads to discriminatory action, if the prejudice is a negative one.⁹⁴ Hence, it is important to distinguish stereotypes and prejudices, even if the former are likely responsible for the creation of the latter.

To summarize, whereas stereotypes help us to make sense of social environments through providing generalizations⁹⁵, prejudices comprise a negative affective, automatized attitude regarding the generalization. To further clarify how stereotypes and prejudices overlap and differ in their functioning let us now turn back to Fricker, who provides an analysis of the intersection of these two notions.

2.2.1 Stereotypes and Credibility Judgements in Fricker

Similarly to classical accounts in social psychology, Fricker defines stereotypes "[...] as

⁹³ Cf., *ibid.*

⁹⁴ Cf., Elliott Aronson, Timothy Wilson, Robin M. Akert, *Social Psychology*, (Hoboken, NJ: Pearson Education Inc., 2013, 8th ed.), 362.

⁹⁵ Evidently, one has to distinguish between reliable, accurate stereotypes and unreliable, inaccurate stereotypes. We will do so in the next section when we discuss Fricker's use and definition of prejudicial stereotypes.

widely held associations between a given social group and one or more attributes.”⁹⁶ As we can derive from this definition, the conception of stereotypes Fricker proposes is fairly broad. More specifically, it is broad in regard to three aspects. First, the concept allows stereotypes to be both reliable and unreliable generalizations concerning a social group.⁹⁷ Second, stereotypes may not only be held as beliefs, but they also have a more affective dimension, which renders them harder to detect. In some passages, in reference to Lippmann’s conception of stereotypes, she compares them to images, an aspect that will become crucial at a later point in the thesis when we consider the influence of implicit bias and aliefs on credibility judgement. Third, Fricker’s conception of stereotypes allows an interpretation in which stereotypes can have a positive or negative valence.⁹⁸ Consider, for instance, the way Greenleaf dismissed Sherwood’s testimony by referring to the stereotype that women are intuitive. In this context, the valence of the stereotype clearly is a negative one, because it is prejudiced and derogatory. However, in contexts where being “intuitive” is regarded as something positive, the stereotype can be understood as something positive.⁹⁹

What we therefore know about Fricker’s broad view of stereotypes so far is that they are associations between social groups and certain attributes that can be reliable or unreliable. Furthermore, stereotypes can be held as beliefs, but can also operate on a more affective dimension. Finally, stereotypes have positive or negative valence, depending on how they are used in specific contexts. Fricker draws the conclusion that “[...] if stereotypes are widely held associations between a group and an attribute, then stereotyping entails a cognitive commitment to some empirical generalization about a social group [...]”¹⁰⁰ These generalizations play an important role in categorizing complex social environments and, in consequence, also influence credibility judgements. This aspect of stereotypes within Fricker’s account is in line with the cognitive approach in social psychology we discussed above. Hearers are constantly confronted with the task of measuring how likely it is that the testimony they are confronted with is true. According to Fricker, such a measurement necessarily relies on generalizations, mediated through stereotypes.¹⁰¹ To illustrate this, she

⁹⁶ Fricker, *Epistemic Injustice*, 30. (Original quote in italics as well.)

⁹⁷ Cf., *ibid.*

⁹⁸ Cf., *ibid.*, 30-31.

⁹⁹ Cf., *ibid.*, 31.

¹⁰⁰ *Ibid.*, 31.

¹⁰¹ Cf., *ibid.*, 32.

gives the example of the dependable family doctor:

In so far as the association crystallized in this stereotype means that it embodies an empirically reliable generalization about family doctors, it is epistemically desirable that the stereotype should help shape the credibility judgements we make when such doctors give us medical advice. Much of everyday testimony requires the hearer to engage in a social categorization of speakers, and that is how stereotypes oil the wheels of testimonial exchange.¹⁰²

Consequently, stereotypical generalizations are useful tools when it comes to making sense of what others tell us. However, as much as stereotypes can function as heuristic aids in credibility judgements, they are not immune to containing unreliable empirical generalizations about social groups. According to Fricker, such stereotypes are prejudiced. Consider someone who estimates male doctors to be more capable than female doctors, even when the two groups have the same level of education, expertise and medical skillfulness. In this case, the judgement clearly comprises a prejudice against female doctors. In what follows, we will investigate how Fricker defines prejudice, namely through her concept of identity prejudice.

2.2.2 Prejudicial Stereotypes in Fricker's Account

First, it is crucial to understand that prejudicial stereotypes are likely to concern stigmatized and historically powerless groups, such as women, black people or working-class people. Often, the attributes associated to them are "[...] inversely related to competence or sincerity or both: over-emotionally, illogicality, inferior intelligence, evolutionary inferiority, incontinence, lack of 'breeding', lack of moral fibre [...] etc."¹⁰³ If these associations are false, the stereotype in question embodies an unreliable empirical generalization. However, this is not a sufficient factor to render the stereotype prejudicial, as the generalization might amount to a non-culpable mistake.¹⁰⁴ Such non-culpable mistakes might be the result of epistemic bad luck, for example, a lack of information or misleading evidence. Prejudices should always be conceived as something epistemically culpable, since their basic structure is that of a *pre-judgement* without the proper assessment of the available evidence.¹⁰⁵

¹⁰² Ibid.

¹⁰³ Ibid., 32.

¹⁰⁴ Cf., *ibid.*

¹⁰⁵ Cf., *ibid.*, 33.

However, neither is the classification of prejudice as a pre-judgement sufficient to entirely define what prejudices are. Prejudice, as stereotype, is a broader notion:

It is broader in two respects. First, while prejudice is most certainly an idea of a judgement formed or maintained in a manner resistant to the evidence, and where this resistance is caused by some kind of motivation on the part of the subject, this permits motivations that are not ethically bad. [...] Second, prejudice is not always *against* someone or something, for there can be prejudice *in favour*.¹⁰⁶

To further illustrate the first aspect, where prejudice is caused by a motivation that is resistant to counterevidence, and yet not ethically bad, Fricker imagines a panel of referees at a scientific journal who are prejudiced against a certain scientific method. The referees form their judgement thusly, as they are “[...] insufficiently sensitive to the benefits of the new scientific method owing to a deep-seated feeling of loyalty to methodological orthodoxy, or perhaps they feel threatened by intellectual innovation.”¹⁰⁷ Evidently, these are not eligible motivations, but neither are they ethically bad, following Fricker.¹⁰⁸

This example of a panel of scientific journal referees also resonates with the second aspect, which claims that prejudice does not necessarily have to be negative. The panel might be prejudiced in favor of a specific scientific method, leading them to be easily over-impressed when a proposal of the sort is submitted.¹⁰⁹ Therefore, prejudices, as stereotypes, can have positive valence.

In the course of her account, Fricker primarily focuses on negative identity prejudice, which is generated by ethically bad motivations and resistant to counter-evidence. In the following section of the thesis I will consider a criticism formulated by José Medina who claims that positive identity prejudice can similarly distort credibility judgement and trigger instances of epistemic injustice. Therefore, Fricker’s exclusive focus on negative identity prejudice will be challenged. Before we present such a criticism, we first need to clarify more specifically how Fricker defines her notion of negative identity-prejudicial stereotype. According to her, it is:

A widely held disparaging association between a social group and one or more attributes, where this association embodies a generalization that displays some (typically, epistemically culpable) resistance to counter-evidence owing to an ethically

¹⁰⁶ Ibid., 35.

¹⁰⁷ Ibid.

¹⁰⁸ Cf., *ibid.*

¹⁰⁹ Cf., *ibid.*

*bad affective investment.*¹¹⁰

At this point we have reached a crucial aspect of Fricker's theory for negative identity prejudice according to her is "[...] at work in systematic testimonial injustice."¹¹¹ Let us highlight this with an example Fricker presents. It will illustrate how negative identity prejudice operates and how it furthermore leads to epistemic injustice.

The example Fricker gives us is from Harper Lee's *To Kill a Mockingbird*. Situated in the year 1935 in Alabama, a young black man, Tom Robinson, is charged with raping a white girl, Mayella Ewell. According to Fricker, "[...] it is obvious to the reader, and to any relatively unprejudiced person in the courtroom, that Robinson is entirely innocent."¹¹² Atticus Finch, the counsel for defense in charge of the case, has proven that Robinson could not have done what he is accused of, namely beating and raping Ewell, as the cuts and bruises she suffered from that day were clearly produced by a blow of the left arm. However, Robinson's left arm is disabled. Hence, it could not have been him who inflicted the beatings.¹¹³ Nonetheless, this evidence is not enough to invalidate the accusations against Robinson. As Fricker puts it:

[...] The trial proceedings enact what is in one sense a straightforward struggle between the power of evidence and the power of racial prejudice, with the all-white jury's judgement ultimately succumbing to the latter.¹¹⁴

Although counter-evidence is provided, the jury is incapable of revising their prejudice. The negative identity prejudice at work in this example is that all black men lie, that all black men are immoral beings who cannot be trusted around women.¹¹⁵ This negative image of black men, which is projected on Robinson results in an instance of epistemic testimonial injustice committed against him. Even though he tells the truth by asserting that he is innocent, nobody in the jury believes him since he is affected by an identity-prejudicial credibility deficit.¹¹⁶

¹¹⁰ Fricker, *Epistemic Injustice*, 35.
(Original quote in italics as well.)

¹¹¹ Ibid.

¹¹² Ibid., 23.

¹¹³ Cf., *ibid.*

¹¹⁴ Ibid.

¹¹⁵ Cf., *ibid.*, 25.

¹¹⁶ Cf., *ibid.*, 28.

This represents the central case of testimonial injustice. Robinson is not judged as trustworthy because of the identity prejudice concerning the social group of black men he belongs to. He is thereby prevented from conveying knowledge and hence devalued as a rational subject. As Fricker puts it: “To be wronged in one’s capacity as a knower is to be wronged in a capacity essential to human value.”¹¹⁷ Moreover, the testimonial injustice done to Robinson is persistent and systematic since it is the outcome of racial prejudice and a racist ideology. But not only is Robinson wronged and harmed as an individual hearer, but the testimonial injustice done to him also induces a dysfunction to the epistemic system as a whole.¹¹⁸ This aspect will be discussed more thoroughly in the next section, when we briefly investigate Medina’s holistic approach to epistemic injustice.

In summary, the example of *How to Kill a Mockingbird* illustrates how the concept of negative identity prejudice constitutes testimonial injustice. It is likely that the majority of instances of testimonial injustice derive from the conscious or unconscious implementation of negative identity prejudices that lead to credibility deficit. But what about cases where one is attributed more credibility than one would actually deserve, owing to a positive identity prejudice? In what follows, I will present an account that takes the negative ramifications of such examples into consideration.

2.2.3 Credibility Deficit versus Credibility Excess

In his book *The Epistemology of Resistance*, Medina presents an account that integrates the notion of granting someone with an excessive amount of credibility into an investigation of epistemic injustice. Before elaborating on why one should consider credibility excess as a trigger for epistemic injustice, one first needs to clarify why Fricker rules out the possibility of credibility excess as creating the same kind of epistemic harm as credibility deficit does: “[...] the primary characterization of testimonial injustice [...] remains such that it is a matter of credibility deficit and not credibility excess.”¹¹⁹ Although Fricker acknowledges that credibility excess can contribute to testimonial injustice, she does not accredit it with the same kind of harm that credibility deficit generates. Fricker insists that in cases of credibility

¹¹⁷ Ibid., 44.

¹¹⁸ Cf., *ibid.*, 43.

¹¹⁹ Ibid., 21.

excess, only a cumulative effect of these inadequate attributions of credibility leads to an epistemic harm, since none of these excessive attributions wrongs the subject in an immediate manner.¹²⁰

At this point, Medina raises his first point of criticism. He asserts that one should not only focus on the individual moments of testimonial exchange but rather, a proper analysis of the exchange should look into what happens before and after it.¹²¹ What he is proposing is a more holistic approach to epistemic injustice:

The harms that excessive attributions of credibility can inflict will indeed not be perceived in an immediate and direct way, given the holistic aspects of injustice. Epistemic injustices have robust temporal and social dimensions, which involve complex histories and chains of social interactions that go beyond particular pairs and clusters of subjects. Because epistemic injustices are a holistic matter, their analysis too must be holistic.¹²²

A significant and valuable contribution of Medina's account is the fact that his holistic approach to epistemic injustice allows an analysis that not only considers the epistemic harms done to the speaker, but also to the interlocutor. According to Medina, Fricker claims that credibility excess produces no immediate harm to its recipient, meaning the speaker. "[...] But the epistemic harms that excessive attributions of credibility can do go well beyond the speaker being epistemically appraised in a wrongful way from a second-person perspective."¹²³

What Medina's analysis therefore achieves is a perspective that includes the harms that are produced by attributing credibility excess to both speakers and hearers. Medina outlines that credibility judgements in general have effects on everybody involved in the epistemic interaction due to their *interactive* nature. He therefore construes credibility judgements as *comparative* and *contrastive*.¹²⁴ This means that credibility does not only apply to subjects individually, but that this credibility attribution also says something about other subjects in a social network. He illustrates this by referring to the effects of credibility excess:

So it should not be surprising that, in the case of excessive attributions of credibility, the disproportionate epistemic trust given to the speaker affects everybody involved in the interaction and not just the speaker, for it affects the very dynamic that unfolds in the

¹²⁰ Cf., *ibid.*

¹²¹ Cf. José Medina, *The Epistemology of Resistance. Gender and Racial Oppression, Epistemic Injustice and Resistant Imaginations*, (New York: Oxford University Press, 2013), 59.

¹²² *Ibid.*, 60.

¹²³ *Ibid.*

¹²⁴ Cf., *ibid.*, 61.

interaction. By assigning a level of credibility that is not *proportionate* to the epistemic credentials shown by the speaker, the excessive attribution does a disservice to everybody involved: to the speaker by letting him get away with things; and to everybody else by leaving out of the interaction a crucial aspect of the process of knowledge acquisition: namely, opposing critical resistance and not giving credibility or epistemic authority that has not been earned.¹²⁵

By giving epistemic authority to a speaker who does not deserve this degree of credibility, the hearer produces an epistemic harm by perpetuating a power structure that put the speaker in his privileged position in the first place. The harm produced in these instances does not only affect the speaker. In my view, it primarily harms the hearer since he or she attributes authority to the speaker and thereby risks of being overly gullible.

Credibility excesses are unjust because they involve an undeserved treatment of epistemic subjects, which indirectly affects others who are also unfairly treated as enjoying comparatively less epistemic trust.¹²⁶

To further strengthen his point, Medina revisits the example from *How to Kill a Mockingbird* that Fricker presents. According to Medina, Fricker analyzes the epistemic injustice done to Robinson only from a perspective where credibility deficit is assigned to him by “[...] the prosecutor, by most members of the white audience, and especially – ultimately and fatally – by the all-white jury.”¹²⁷ But following Medina this alone does not produce the harm that finally leads to the dismissal of Robinson’s testimony.

For Medina, credibility excess plays an equally fatal role. The credibility excess at play can, for instance, be located in the differential treatment of the witnesses. As Medina asserts, “[...] it is noteworthy that Fricker does not analyze Atticus Finch’s interrogation of Mayella Ewell and how she is perceived by the jury [...]”¹²⁸ Therefore the epistemic injustice done to Robinson is not only produced by the credibility deficit assigned to him after the interrogation, but also by the credibility excess his accusers are attributed with because they are members of the privileged racial group.¹²⁹ However, as Medina demonstrates, such credibility differentials did already exist before the defendant and his accusers were confronted in court. The epistemic authority ascribed to the prosecutor is the product of a

¹²⁵ Ibid.

¹²⁶ Ibid., 63.

¹²⁷ Ibid., 65.

¹²⁸ Ibid.

¹²⁹ Cf., *ibid.*, 66.

society that does not allow a perspective in which the black man is telling the truth and thus is innocent. The all-white jury misplaces their trust by being epistemically lazy and closed-minded, all of which are effects of an attribution of epistemic authority to the white prosecutor and the white accuser, Mayella Ewell. Through this wrongful attribution severe epistemic harm is certainly done to Robinson, but it also bestowed to the white-jury as a whole. The latter can be understood as a self-inflicted harm, as the jury is not capable of seeing beyond racial prejudices. As Medina puts it: “This laziness becomes an epistemic obstacle in the pursuit of knowledge that can easily lead to epistemic injustices.”¹³⁰ The epistemic injustice done to Robinson therefore has to be understood as being part of a bigger discriminatory picture, rooted in the social imaginary and the recourse to (bad) epistemic habits.¹³¹

To summarize, the innovative aspect of Medina’s approach lies in the fact that he is able to consider the negative epistemic ramifications arising in instances of epistemic injustice that not only concerns the speaker but also the hearer. He does so by drawing on the plausible ascription of credibility excess. To this extent, he provides a broader concept of epistemic injustice than Fricker, who puts her primary focus on the hearer when it comes to detecting the epistemic harms effectuated by epistemic injustice.¹³²

At a later point, this thesis will resume this point of criticism when it comes to remedies against epistemic injustice. Fricker’s concept of the virtue of “testimonial justice”, which is accomplished by critical reflection, equally only considers the hearer and leaves the speaker in a rather passive position. However, to say that Fricker does not construe epistemic injustice as a holistic phenomenon embedded in a broader system of social relations of power is too strong a criticism. This quotation, for instance, demonstrates that she at least considers the harm done not only to a speaker but also to the hearer in instances of testimonial injustice:

¹³⁰ Ibid., 68.

¹³¹ Cf., *ibid.*, 69.

¹³² It needs to be conceded that Fricker has another reason to neglect considerations about the ramifications that arise through an excessive ascription of credibility in her account. According to Medina, Fricker does not draw a link between credibility excesses and credibility deficits in her analysis because she does not want to rely on a distributive conception of justice as such a conception, in her opinion, does not apply to epistemic goods, such as credibility. Medina reacts to this objection in saying that even if credibility does not have a distributive nature, we still have to acknowledge that there is an intimate relation between credibility excesses and credibility deficits and that this relation becomes especially apparent in situations of oppression. This thesis supports this claim. (Cf., *ibid.* 62.)

“There is of course a purely epistemic harm done when prejudicial stereotypes distort credibility judgements: knowledge that would be passed on to a hearer is not received.”¹³³ Fricker thus favors a perspective where, not granting someone with the due credibility they would deserve harms not only the speaker, but the hearer as well. In the case of the latter, this harm is, of course, self-inflicted. Moreover, as shown above, Fricker is well aware that power as a structural phenomenon regulates our everyday interaction with the world and others. Fricker’s elaborate focus on stereotypes and prejudicial stereotypes, which can be detected throughout her book, demonstrates that she is a firm believer in the power of the social imaginary and the influence of epistemic habits.

A particularly crucial notion for my thesis is concerned with an aspect of Fricker’s book, which she describes as “residual internalizations”. These internalizations make it likely that our credibility judgements are influenced by beliefs or belief-like states that we do not consciously endorse. In Medina’s terms, those internalizations could be translated into epistemic habits.

2.3 Residual Internalizations – An Incomplete Concept?

In order to introduce the phenomenon of residual internalizations in Fricker’s account and the criticism related to it that this thesis wants to put forward, let us briefly recapitulate how Fricker describes the generation of epistemic testimonial injustice.

Testimonial injustice arises through ascribing credibility deficit to a speaker due to the mediation of negative identity prejudices. Referring back to Fricker’s perceptual model of testimony, this means that the hearer *perceives* the speaker as more or less trustworthy and this perception is strongly influenced by (prejudicial) stereotypes.¹³⁴ As we have seen in the section above, Medina expands this definition by integrating the possibility of excessive ascriptions of credibility as being equally harmful to the speaker, but furthermore, that this harm also concerns the hearer who made the credibility attribution. Therefore, positive and negative identity prejudices can distort a hearer’s credibility judgements, or, as Fricker puts it: “[...] *it distorts the hearer’s perception of the speaker.*”¹³⁵ In the example of *How to Kill*

¹³³ Ibid., 43.

¹³⁴ Cf., Fricker, *Epistemic Injustice*, 36.

¹³⁵ Ibid.

a Mockingbird, for instance, the jury's judgement is distorted by prejudicial racial stereotypes to such a degree that they can no longer perceive Tom Robinson differently than a lying black man. In this example, as in the example of the *Talented Mr. Ripley* the prejudicial stereotype is doxastically mediated, which means that it is actively endorsed and that the subject is conscious of the content of the prejudicial stereotype. And yet Fricker states that her prior interest lies in cases of epistemic injustice, where the prejudice is not actively endorsed but operates on a more unconscious level. The following quote illustrates this:

But our focus will be chiefly on the operation of prejudice at the non-doxastic level; for concentrating on beliefs would lead us to underestimate the incidence of testimonial injustice. I believe that the right vision of epistemic relations is such that testimonial injustice goes on much of the time, and while it may be hard enough to police one's beliefs for prejudice, it is significantly harder reliably to filter out the prejudicial stereotypes that inform one's social perceptions directly, without doxastic mediation.¹³⁶

At the beginning of her book, Fricker refers to a quotation by Judith Shklar who criticizes that the history of philosophy taught us to think about justice as the norm and in turn injustice should be understood as the aberration of this norm. "This moral model of justice does not ignore injustice but it does tend to reduce it to a prelude to or a rejection and breakdown of justice as if injustice were a surprising abnormality."¹³⁷ Fricker cites Shklar in order to voice her conviction that testimonial injustice should not be understood as an occasional occurrence, but rather "[...] various degrees of testimonial injustice happen all the time."¹³⁸ I suggest that we take Fricker seriously when she cites Shklar.

In order to illustrate how it is possible that we all are likely perpetrators of epistemic injustice, Fricker introduces her concept of residual internalizations.

Residual internalizations constitute instances in which prejudicial images from the social imagination persist in a hearer's patterns of credibility judgement, even though their content conflicts with the hearer's remaining beliefs.¹³⁹ Fricker's aim is to point out that because of these unconsciously operating identity prejudices, instances of testimonial injustice are likely to happen on a regular basis, as the hearer might not even be aware of the fact that he or she is actually committing an act of testimonial injustice. In another passage, Fricker

¹³⁶ Ibid.

¹³⁷ Judith Shklar, *The Faces of Injustice*. (New Haven, London: Yale University Press, 1990), 17.

¹³⁸ Fricker, *Epistemic Injustice*, 37.

¹³⁹ Cf., *ibid.*

claims that she is mainly interested in the question of how prejudices operate on a non-doxastic level. “Certainly we may sometimes perpetrate testimonial injustice because of our beliefs; but the more philosophically intriguing prospect is that we may very frequently do it in spite of them.”¹⁴⁰

The concept of residual internalizations will become decisive for the further development of this thesis’ main argument. In what follows I will claim that the notion of residual internalizations is crucial for Fricker’s account. However, her implementation is lacking in depth. Although Fricker provides the groundwork for reflecting upon instances of testimonial injustice that come about unconsciously by non-actively endorsed residuals of identity prejudices, she does not develop this point to its fullest extent. In the course of her argument it is rarely mentioned again. Moreover, the examples she presents to illustrate instances of testimonial injustice are built on identity prejudices that have been actively endorsed. In the case of *The Talented Mr. Ripley* we are confronted with an example where sexist prejudices that finally lead to the instance of testimonial injustice are a firm component of the social imagination. To judge a woman as less credible than a man was nothing unusual at the time in which the novel and movie are set in, rather representing the common practice of the era. Therefore, however morally wrong and blameworthy we might judge Herbert Greenleaf’s action from our current perspective, we should not be surprised that he is enacting the sexist prejudices of his time.

Similarly, in *How to Kill a Mockingbird*, the members of the jury commit a testimonial injustice against Tom Robinson because they are actively endorsing the common racial prejudice predominant at the time, namely that black men are not trustworthy. This raises the question as to why Fricker made these examples so prominent. The identity prejudice in both examples was actively endorsed. This is not consistent with her assertion that the more philosophically intriguing question concerns cases where prejudices are endorsed in spite of our remaining beliefs or operate on an unconscious level. One of the thesis’ core objectives is to argue that this contradicts Fricker’s statement cited above, which announced that “[...] various degrees of testimonial injustice happen all the time.”¹⁴¹ Although, I am in full agreement with Fricker that prejudices operating on a non-doxastic level are of bigger

¹⁴⁰ Ibid., 36.

¹⁴¹ Ibid., 37.

interest for a philosophical investigation, as their detection is much harder to effectuate, I will argue that Fricker is not consistent in the execution of this argument. In the following chapter the thesis will therefore introduce the notions of “implicit bias” and “aliefs” in order to propose an improvement to Fricker’s concept of residual internalizations.

3. Residual Internalizations Revisited

In order to shed further light on the phenomenon of residual internalizations, let us consider the following example. Imagine a woman who consults two doctors, one male, and the other female, on a specific medical issue. Although the two doctors were trained in the same medical field, they come to different conclusions regarding their patient’s health issue and thus provide two different diagnoses. Even though the patient does not explicitly support any kind of normative sexist behavior or thinking, and explicitly endorses feminist beliefs, she is nonetheless inclined to intuitively trust the diagnosis of the male over that of the female doctor. Let us moreover assume that the patient grew up in a household in which stereotypical gender norms played an important role. Her father was the so-called “head” of the household, whereas her mother never had a say on sociopolitical matters. Even though the woman does not consciously endorse stereotypical gender-specific attitudes, traditional gender norms still linger in her unconsciousness and incline her to regard the male doctor as more competent than the female doctor.

Fricker would describe such a case as a residual internalization, because “[...] cognitive commitments held in our imaginations retain their impact on how we perceive the social world even after any correlative beliefs have faded away.”¹⁴² As already stated above, Fricker regards instances of testimonial injustice in which residual prejudices persist in the subject’s unconsciousness as philosophically more puzzling than cases in which the prejudice is actively endorsed. What makes them especially difficult – philosophically as well as practically – is the fact that they are harder to detect than cases with consciously approved prejudices. “Residual prejudice [...] is the sort of prejudice that will bring about the most surreptitious and psychologically subtle forms of testimonial injustice.”¹⁴³ For that reason, it is likely that someone becomes a perpetrator of testimonial injustice despite his or

¹⁴² Fricker, *Epistemic Injustice*, 37.

¹⁴³ *Ibid.*, 39.

her conscious beliefs. Thus, instances of testimonial injustice happen more regularly than we might think.¹⁴⁴ However, as I have already suggested, the way in which Fricker implements her concept of residual internalizations lacks depth.

This criticism can be supported by three observations. First, throughout Fricker's account, the notion of residual internalization is hardly mentioned again. Second, as already has been established, the two main examples Fricker provides to illustrate cases of testimonial injustice present instances where the identity prejudices were actively endorsed. Finally, the remedy that Fricker proposes against testimonial injustice is the virtue of testimonial justice, which is accomplished through a critical awareness on the part of the individual. This, however, seems to neglect, or at least downgrade, cases where identity prejudices are unconsciously endorsed. I aim to show that this antidote against testimonial injustice is inconsistent with Fricker's claim that instances of epistemic injustice happen all the time, as they are the product of unconsciously endorsed identity prejudices. Alcoff also raises this criticism by asking: "Can volitional epistemic practices correct for non-volitional prejudices?"¹⁴⁵ To answer this question I suggest taking a look at research regarding implicit biases.

3.1 Implicit Biases and Epistemic Injustice

This section provides a brief presentation of the metaphysical, epistemological and ethical questions related to implicit-bias research. Subsequently I shall investigate the connection between implicit bias and Fricker's account of epistemic injustice.

Drawing on Jennifer Saul and Michael Brownstein's work on implicit biases in philosophy, the phenomenon can be described as follows:

"Implicit bias" is a term of art referring to evaluations of social groups that are largely outside of conscious awareness or control. These evaluations are typically thought to involve associations between social groups and concepts or roles such as "violent", "lazy", "nurturing", "assertive", "scientist", and so on. Such associations result at least in part from common stereotypes found in contemporary liberal societies about members of these groups.¹⁴⁶

¹⁴⁴ Cf., *ibid.*

¹⁴⁵ Alcoff, "Epistemic Identities", 128.

In the last chapter of the thesis we will see how Fricker responds to this criticism.

¹⁴⁶ Jennifer Saul, Michael Brownstein, *Implicit Bias and Philosophy. Metaphysics and Epistemology. Volume 1*, (Oxford: Oxford University Press, 2016) 1-2.

Implicit bias therefore picks out the tendency of people to unconsciously foster stereotypical associations and attitudes in spite of their conscious beliefs. Furthermore, these biases impact social behavior. For instance, implicit racial biases influence the evaluation of résumés. Studies have shown that people tend to prefer résumés with stereotypically white names and downgrade those of people with stereotypically black names.¹⁴⁷ Another example is the so-called “shooter bias” test, a computer simulation with the help of which it has been shown that people are more likely to shoot an unarmed black person than an unarmed white person.¹⁴⁸ If we consider the incidences that led to the “Black Lives Matter” movement in the U.S., this example is sadly not only confirmed in the computer simulated tests, but the same biased mechanism has caused numerous real-life tragedies.

In order to illustrate how implicit biases function, let us take a look at the following fictional example, constructed by Eric Schwitzgebel. Juliet is a white professor at an American philosophy department. She has the firm belief that there is no cognitive difference between black students and white students, and thus argues for equality of intelligence between them. This belief overlaps with her beliefs regarding other subjects, which on the whole are liberal ones. Yet, something about Juliet’s unreflective behavior of certain individuals indicates a racial bias:¹⁴⁹

When she gazes out on class the first day of each term, she can’t help but think that some students look brighter than others – and to her, the black students never look bright. When a black student makes an insightful comment or submits an excellent essay, she feels more surprise than she would were a white or Asian student to do so, even though her black students make insightful comments and submit excellent essays at the same rate than others do. This bias affects her grading and the way she guides class discussion.¹⁵⁰

How is it possible that Juliet still is surprised when a black student submits a good essay or makes a clever remark in class, when she seemingly sees everybody equally gifted regardless of their race? This question concerns metaphysical implications about implicit biases. What is their structure and “[...] how do they fit into the architecture of the mind?”¹⁵¹

¹⁴⁷ Cf., *ibid.*

¹⁴⁸ Cf., *ibid.*

¹⁴⁹ Cf. Eric Schwitzgebel, “Acting Contrary to Our Professed Beliefs, or the Gulf Between Occurrent Judgement and Dispositional Belief”, in: *Pacific Philosophical Quarterly* 91(4), (2010), 531-532.

¹⁵⁰ *Ibid.*, 532.

¹⁵¹ Saul, Brownstein, *Implicit Bias and Philosophy*, 2.

First, we need to distinguish between implicit and explicit biases. Obviously, we might think of implicit bias as unconscious, whereas explicit bias is conscious. However, in claiming this, we should give a more thorough explanation of what we mean by “conscious”. What makes Juliet’s bias implicit and thus unconscious is that it is not introspectable: “[...] She cannot report straight off that she possesses it, as she can report her explicit views, and she becomes aware of it only through observing or being informed of its effects on her behavior.”¹⁵² Thus Juliet does not endorse the implicit bias in her conscious decision-making whereas explicit bias “[...] would be a bias that is endorsed in conscious deliberation.”¹⁵³ How does implicit bias occur in Juliet’s unconscious cognition about black students that leads her to be surprised when such a student attracts her attention through academic achievements?

To answer this question, one first needs to acknowledge that much of what we do is performed without conscious deliberation. Just think of someone playing the piano. Most of the related movement is likely to be executed in a spontaneous and unreflective manner.¹⁵⁴ However, this unreflective behavior is intelligent, “[...] in the sense of being responsive to our beliefs and desires, and we would naturally explain it in belief-desire terms.”¹⁵⁵ For instance, we would expect a piano player to play a sonata roughly as the written music suggests it. Nevertheless, she might include spontaneous variations of the sonata or rather know the piece of music by heart and does not need to think about which chord she will play next. “That is, unreflective behavior (or much of it, at any rate) appears to be the product of practical reasoning, rationally responsive to the agent’s beliefs and desires.”¹⁵⁶ But does this mean that the mental state of Juliet’s implicit bias is that of a belief?¹⁵⁷

In the field of implicit-bias research, philosophers can be divided into roughly two camps

¹⁵² Keith Frankish, “Playing Double. Implicit Bias, Dual Levels and Self-Control”, in: *Implicit Bias and Philosophy. Volume 1. Metaphysics and Epistemology*, ed. Saul, Brownstein, (Oxford: Oxford University Press, 2016), 23-46, 25.

¹⁵³ Ibid., 26.

¹⁵⁴ Cf., ibid., 26.

¹⁵⁵ Ibid.

¹⁵⁶ Ibid.

¹⁵⁷ Schwitzgebel, from whom we drew the example of Juliet and her implicit attitudes towards black students, argues that Juliet’s implicit bias has a belief-like structure. He calls this a “contradictory belief”. “Contradictory belief holds that an agent’s implicit and explicit attitudes both reflect what she believes, and that these different sets of beliefs may be causally responsible for different behaviors in different contexts. (<https://plato.stanford.edu/entries/implicit-bias/#DoxMod>).

concerning this question. One camp argues that implicit biases are *sui generis* associative states, whereas the other camp defines them as beliefs or belief-like states.¹⁵⁸ I will not take a stance on either interpretation. Rather, I will present accounts from both camps and try to decide which one best fits Fricker's concepts of residual internalization and identity prejudice.

In addition to highlighting metaphysical questions about the structure of implicit biases, this thesis also wants to shed light on the idea that implicit bias poses interesting epistemological and ethical challenges. In terms of the epistemological aspect, one could ask if we need to become skeptics about our beliefs and judgements when we assume that it is likely that we are influenced by implicit biases. If so, how should this question be treated in the realm of Fricker's account of epistemic injustice? I will address this issue in the next chapter by drawing on Saul's concept of "bias-related doubt."

Regarding ethical deliberations on implicit biases, one might ask if individuals are morally responsible for their implicit biases and the effects they have on their behavior.¹⁵⁹ This aspect will also become decisive in the next chapter since I will apply the question of ethical and epistemic responsibilities concerning our implicit biases to Fricker's account of testimonial injustice. Fricker's overarching aim is to examine the ethical aspects of our fundamental epistemic practices, namely conveying knowledge to others and decoding our own social experience.¹⁶⁰ That is why I believe it is crucial to ask whether we can responsibly conduct credibility judgements even when it is likely that some of these judgements are informed by implicit biases.

In order to facilitate the improvement of Fricker's account of epistemic injustice through drawing on implicit-bias research, the next section will clarify how the two notions intersect and differ from one another.

3.1.1 Implicit Bias and Testimonial Injustice – Similarities and Differences

A first point of similarity between testimonial injustice and implicit bias is that they concern first and foremost members of marginalized groups. As it was shown throughout my first

¹⁵⁸ Cf., <https://plato.stanford.edu/entries/implicit-bias/>

¹⁵⁹ Cf. Saul, Brownstein, *Implicit Bias and Philosophy*, 2.

¹⁶⁰ Cf. Fricker, *Epistemic Injustice*, 1.

chapter, epistemic injustice according to Fricker describes the ways in which members of marginalized groups are downgraded in their capacity as knowers owing to identity prejudices. Similarly, implicit-bias research focuses its attention on how implicit stereotypical images about members of marginalized groups influence an individual's unconscious evaluation of that group.¹⁶¹ Remember the studies mentioned above that showed how implicit biases affect the ways résumés are evaluated, which demonstrated a tendency of preferring résumés from people with “white-sounding names”. A further example of this is illustrated in a test designed by psychologists, called the “Implicit Association Test” (“IAT”). The participating subjects are asked to arrange words and pictures into categories as fast as possible while trying to make as few errors as possible. It was shown that the tested subjects are likely to sort stereotypical-consistent images faster than stereotypical-inconsistent images, even if they do not consciously endorse the content of the stereotype. For instance, one recent review of the test demonstrated that white participants more easily associated black faces with negative terms (e.g. war, bad, violent, etc.) and white faces with positive ones (peace, good, etc.).¹⁶²

These findings suggest that implicit biases are probably traceable to instances of epistemic injustice. One merely has to cast one's thoughts back to how Fricker constructs her notion of residual internalizations. She describes them by pointing out “[...] the influence of prejudicial images from the social imagination that persist in a hearer's patterns of judgement even where their content *conflicts* with the content of her belief.”¹⁶³ The notion of residual internalizations in Fricker's account thus features striking similarities with implicit biases.

This thesis is not the first paper to draw a connection between implicit-bias research and epistemic injustice. In her 2016 paper “Fault and No-Fault Responsibility for Implicit Prejudice – A Space for Epistemic Agent Regret”¹⁶⁴, Fricker examines to what extent agents

¹⁶¹ Cf. Jennifer Saul, “Implicit Bias, Stereotype Threat and Epistemic Injustice”, in: *The Routledge Handbook of Epistemic Injustice*, ed. Kidd, Ian James, (London, New York: Routledge, 2017), 234-243, 234.

¹⁶² <https://plato.stanford.edu/entries/implicit-bias/>

¹⁶³ Fricker, *Epistemic Injustice*, 37.

¹⁶⁴ Miranda Fricker, “Fault and No-Fault Responsibility for Implicit Prejudice – A Space for Epistemic Agent Regret”, in: *The Epistemic Life of Groups. Essays in the Epistemology of Collectives*, ed. Michael S. Brady, Miranda Fricker, (Oxford: Oxford University Press, 2016), 33-51.

should be held responsible for their implicit biases and simultaneously compares implicit bias to implicit prejudice. Drawing on Jules Holroyd's deliberations on implicit bias, Fricker asserts that implicit bias incorporates implicit prejudicial thinking. As demonstrated above, prejudicial attitudes arise due to a motivated or unconscious maladjustment to evidence. For Fricker, this maladjustment can be explained by drawing on implicit bias since they are automatically activated and concern negative stereotypical associations.¹⁶⁵ However, she concedes that automatic association need not necessarily lead to prejudice, but it will do so when one fails to "[...] properly gear one's attitudes to the evidence [...]"¹⁶⁶ which is the case in most automatic associations of negative traits with stigmatized groups.¹⁶⁷

This resistance to counter-evidence, which leads to the prevalence of implicit, prejudices "[...] seriously comprises our conception of ourselves as cognitively authentic, or even epistemically responsible."¹⁶⁸ Still, Fricker does not apply implicit bias to the totality of her theory of epistemic injustice, but only inquires to what extent we should be held responsible for our biases.¹⁶⁹

In a contribution for the *Routledge Handbook of Epistemic Injustice*, Saul provides a more detailed account of how to link implicit-bias research to epistemic injustice. But my thesis makes a stronger claim than Saul does in her paper. Whereas Saul only hints at the similarities and differences between implicit-bias research and epistemic injustice, I propose that findings in implicit-bias research serve as a direct antidote to the inconsistency regarding residual internalizations arising in Fricker's account.

One point of similarity that Saul presents concerns Fricker's notion of negative identity-prejudicial stereotypes leading to credibility deficits.

[...] Fricker's discussion of how credibility deficits actually function is in fact a very nice fit with discussions of implicit bias. First, her crucial notion of a negative identity-prejudicial stereotype does not seem to be one that requires consciousness.¹⁷⁰

I support this observation. Let us recall how Fricker constructed her notion of negative identity prejudices by defining them as associations between a social group and specific

¹⁶⁵ Cf., *ibid.*, 39.

¹⁶⁶ *Ibid.*, 38.

¹⁶⁷ *Ibid.*

¹⁶⁸ *Ibid.*, 37.

¹⁶⁹ I will draw on Fricker's deliberations concerning this aspect in the next chapter.

¹⁷⁰ Saul, "Implicit Bias, Stereotype Threat and Epistemic Injustice", 235.

attributes, where this generalization might come about by an *affective* investment.¹⁷¹ Thus, according to Fricker, it is possible that such prejudices are not *reflectively* but *affectively* endorsed. This, however, does not automatically render them implicit. Yet, as stated above, Fricker also considers the possibility (and likelihood) of implicit identity prejudices that are at play in instances of testimonial injustice. She does so by formulating the notion of residual internalizations. Thus, to draw a connection between implicit biases and testimonial injustice seems legitimate, considering how implicit biases and Fricker's notion of negative identity prejudice and residual internalization overlap in their functioning. Nevertheless, implicit bias and testimonial injustice should not be understood as describing the exact same phenomenon.

Saul identifies four ways in which implicit bias theory and testimonial injustice differ from one another. First, testimonial injustice only exists when it is directed at someone, whereas implicit biases are associations “[...] that may or may not ever manifest in this way.”¹⁷² This brings us to the second point of difference Saul presents. Whereas implicit bias is a psychological notion regarding the state of mind of an individual, testimonial injustice is an interactive notion that requires a speaker, an audience and an occasion in which the testimonial injustice is executed.¹⁷³ As Saul states: “Implicit bias can give us a small bit of this – it can be part of why a speaker might perpetrate an epistemic injustice. But it will not ever give us the whole of this.”¹⁷⁴

A third and crucial point of difference Saul emphasizes concerns credibility. She concedes that there are clearly implicit biases that are related to credibility. Think of the example of the professor, Juliet, who implicitly attributed black students with less intelligence and thus less credibility. However, according to Saul, not all implicit biases are related to credibility.¹⁷⁵ To illustrate this, she provides the example of associating black people with violence, which, in her view, is not a matter of credibility.¹⁷⁶

A final point of divergence Saul presents is the question as to whether or not all implicit biases are negative. Remember that Fricker construes identity prejudices that lead to

¹⁷¹ Cf. Fricker, *Epistemic Injustice*, 35.

¹⁷² Saul, “Implicit Bias, Stereotype Threat and Epistemic Injustice”, 235.

¹⁷³ Cf., *ibid.*

¹⁷⁴ *Ibid.*, 236.

¹⁷⁵ Cf., *ibid.*

¹⁷⁶ Cf., *ibid.*

instances of testimonial injustice as primarily negative.¹⁷⁷ It seems to me that the difference may not be that decisive. Remember Medina, who, in contrast to what Fricker says, argues that positive identity prejudices are just as harmful as negative ones, owing to the *comparative* and *contrastive* structure of credibility judgements.¹⁷⁸ Consequently, positive identity prejudice and positive implicit bias are just as likely to lead to instances of testimonial injustice as negative ones, for an excessive attribution of credibility can be harmful both for the speaker and the interlocutor.

Through drawing on similarities and differences between implicit bias and testimonial injustice, the thesis aims at showing that the two notions should not be understood as interchangeable. However, implicit-bias research seems well elaborated when it comes to describing instances of testimonial injustice where the identity prejudice at stake operates on an unconscious level. In the following section the thesis will provide further justification for this claim by discussing the phenomenon of stereotype threat.

3.1.2 Stereotype Threat and its Link to Epistemic Injustice

In what follows, I will explore the phenomenon of stereotype threat and its relation to Fricker's account of epistemic injustice. I will argue that some of the (implicit) prejudices at work in instances of stereotype threat are exactly those implicit identity prejudices that give rise to testimonial injustice. Moreover, the two notions overlap when it comes to their negative psychological, practical and epistemic ramifications. This observation also justifies the claim that some instances of stereotype threat and testimonial injustice are produced by the same implicit prejudicial mechanisms.

Stereotype threat describes a psychological phenomenon that is triggered by a specific stereotype concerning a social group. As such, it occurs when "[...] members of a group that are negatively stereotyped at some particular task care about doing well at it, and are reminded of the negative stereotype of their group."¹⁷⁹ In the previous chapter we discussed the example of Juliet, the university professor, whose implicit biases concerned the intellectual abilities of black students. These implicit biases might affect the way Juliet

¹⁷⁷ Cf., *ibid.*

¹⁷⁸ Cf. Medina, *The Epistemology of Resistance*, 61.

¹⁷⁹ Saul, "Implicit Bias, Stereotype Threat and Epistemic Injustice", 236.

interacts with those students. She might act surprised when they do well in class or, prefer to let white students explain difficult theoretical issues in presentations. If this happens on a regular basis, the black students will eventually become aware of Juliet's implicit biases towards them. In the 'best' case, this will induce them to prove her wrong and make an effort to impress her intellectually. However, in the worst case, Juliet's black students will suffer from stereotype threat. This will subsequently have a negative effect on their academic performance.

One negative ramification of stereotype threat is that it hinders performance.¹⁸⁰ Apart from underperformance, stereotype threat has further severe practical and psychological effects, namely "psychological disengagement" and "domain avoidance".¹⁸¹ Let us illustrate these effects by applying them to our previous example. Imagine three black students in Juliet's class, who become aware of her implicit attitudes towards them.¹⁸² The first student might care about doing well in her class in order to escape the stereotype. However, because he is so eager to prove Juliet wrong he might be excessively nervous on the day of the exam and consequently fail. This would be an example of underperformance. The second student who becomes aware of the stereotype concerning her group drops the class because she gets intimidated. This is a case of domain avoidance. The third student continues with the class but always sits in the back and tries not to attract attention. This would be an example of psychological disengagement.¹⁸³ All three effects of stereotype threat are "[...] part of individuals' reactions to threats of devaluation that are triggered by the presence and possible salience of a negative stereotype."¹⁸⁴

In addition to these practical and psychological effects, stereotype threat also likely leads to self-doubt, representing a notably severe epistemic harm.¹⁸⁵ By focusing on the negative effects of self-doubt arising through stereotype threat, we get a glimpse of how "[...] deeply

¹⁸⁰ Cf., Stacey Goguen, "Stereotype Threat, Epistemic Injustice and Rationality", in: *Implicit Bias and Philosophy. Volume 1. Metaphysics and Epistemology*, ed. Saul, Brownstein, (Oxford: Oxford University Press, 2016) 216-237, 216.

¹⁸¹ Cf., *ibid.*, 218.

¹⁸² It needs to be conceded that in order for the black students to suffer from stereotype threat, they do not even have to be aware of Juliet's implicit attitudes towards them. Rather, it suffices that they are aware of the negative prejudices concerning the intellectual abilities of black persons predominant in society to be afflicted by stereotype threat.

¹⁸³ For a similar example see: Goguen, "Stereotype Threat, Epistemic Injustice and Rationality", 221.

¹⁸⁴ Cf., *ibid.*

¹⁸⁵ Cf. *ibid.*, 223.

stereotype threat can affect our epistemic lives and our very sense of ourselves.”¹⁸⁶

Analyzing the negative epistemic effects that arise in situations of stereotype threat is germane to the purposes of this thesis. In the following I will draw an analogy with the harms arising in instances of testimonial injustice that Fricker presents, in order to point out that they are almost identical to the harms generated by stereotype threat. This will serve as a further justification for taking implicit-bias research as a fruitful extension of Fricker’s account.

Taking a closer look at Fricker’s text, we encounter two categories of harm that she distinguishes in instances of testimonial injustice: primary and secondary.¹⁸⁷ Primary harms could be thought of as ethical harms, as they downgrade a speaker *qua* knower, a competence central to human value:

The capacity to give knowledge to others is one side of that many-sided capacity so significant in human beings: namely, the capacity for reason. [...] When someone suffers a testimonial injustice, they are degraded *qua* knower, and they are symbolically degraded *qua* human.¹⁸⁸

Thus, the primary harm of an instance of testimonial injustice not only degrades the concerned person in her capacity as a transmitter of knowledge, but also harms her in her very humanity. At a later point of the thesis we will see how the same harm occurs in situations of stereotype threat. Taking a closer look at the secondary forms of harm that Fricker distinguishes, the similarity to cases of stereotype threat becomes even more evident. Elaborating on the secondary aspects of harms, Fricker distinguishes them into a *practical* and an *epistemic* dimension of harm.¹⁸⁹

Let us start by focusing on the practical harms. Fricker defines practical harms by providing examples. She describes an instance where a woman from Egypt knew that her suggestions regarding policies would not receive approval in professional meetings. Disadvantaged by the prejudicial attitudes towards her testimony as a woman, she would thus ask a sympathetic male colleague to put her ideas forward in the meetings. Because she, as a woman, had so much experience of not being taken seriously when she came up with an

¹⁸⁶ Ibid.

¹⁸⁷ Cf., Fricker, *Epistemic Injustice*, 44.

¹⁸⁸ Ibid.

¹⁸⁹ Cf., *ibid.*, 46.

idea, she gave in to the situation and let her male colleague earn the credit. Getting the policies implemented was more important to her than earning credit and admiration.¹⁹⁰ One could argue that the woman was not completely threatened by the stereotype, as she still cognitively participated in the meetings. Since she was aware that her motion would receive no endorsement, she ceased to actively participate in the meetings.

I want to propose that the woman's behavior in this situation is similar to behavior that arises in stereotype threat related situations. To some extent, the woman disengaged and avoided situations in which she could become ignored or even downgraded on account of her gender identity and the prejudices related to it.

Furthermore, it is likely that her male colleagues did not only ignore her testimony, but never really asked her to present her ideas in the first place. In Fricker's account, we are confronted with a possible explanation for such behavior, namely the idea of *pre-emptive* testimonial injustice.¹⁹¹ Pre-emptive testimonial injustice arises when "[...] hearer prejudice does its work in advance of a potential informational exchange: it pre-empts any such exchange."¹⁹² This means that it is likely that the woman's testimony in professional meetings was never even solicited. In other words she was silenced before she even had the chance to speak. For that reason, she asked her male colleague to put forward her ideas, and disengaged.

This example shows us that practical harms for the individual are rather similar in situations of testimonial injustice and in situations of stereotype threat. For instance, because of the woman's disengagement, it is likely that she did not receive the promotion her male colleague might have received, since she came across as passive and incompetent. Disengagement and avoidance in situations of stereotype threat might lead to similar practical harms. One example could be that women are less likely to strive for high profile positions, as they themselves do not feel competent enough to apply for such positions.

Let us now turn to the secondary epistemic effects that Fricker characterizes in instances of testimonial injustice. Here, she uses a term that we have already encountered in the discussion about stereotype threat, namely self-doubt.

¹⁹⁰ Cf., *ibid.*, 47.

¹⁹¹ Cf., *ibid.*, 130.

¹⁹² *Ibid.*

The second category of secondary harm caused by testimonial injustice is (more purely) epistemic harm: the recipient of a one-off testimonial injustice may lose confidence in his belief, or in his justification for it, so that he ceases to satisfy the conditions for knowledge; or alternatively, someone with a background experience of persistent testimonial injustice may lose confidence in her general intellectual abilities to such an extent that she genuinely is hindered in her educational or other intellectual development.¹⁹³

This similarity between the harms arising from testimonial injustice and the harms arising from instances of stereotype threat is striking. In cases of stereotype threat, being afflicted by self-doubt will likely censor speaking because the affected individual becomes self-conscious and withdraws from communicative situations. By being associated with the prejudicial stereotype, the individual affected is both excluded from the pooling of knowledge, and excludes herself by her actions, once aware of the stereotype. Thus, she is not taken seriously as a rational knower and also loses confidence in her intellectual abilities. As we have encountered above, these are exactly the kind of ethical and epistemic harms Fricker describes in cases of testimonial injustice.

Of course, one needs to point out that cases of implicit bias and stereotype threat are not necessarily related to cases of testimonial injustice. Whereas implicit bias and stereotype threat are notions that stem from the social psychology literature, testimonial injustice is a concept originating in social epistemology and political philosophy.¹⁹⁴ However, this thesis makes a strong claim for the likelihood of the same implicit biases playing a decisive part in many cases of testimonial injustice and stereotype threat. As we have seen, stereotype threat impairs performance by becoming aware of the (implicit) stereotypes concerning one's social group. It is thus likely that some of these stereotypes are precisely the same stereotypical identity prejudices that lead to epistemic injustice. This becomes especially evident when we look at the ethical and epistemic harms arising in the two notions.

3.2 Belief-Like Versus Associative States of Implicit Bias

In the previous section I have hinted at the concept of pre-emptive testimonial injustice that Fricker introduces in her account. Fricker asserts that pre-emptive forms of testimonial

¹⁹³ Ibid., 47-48.

¹⁹⁴ Cf., Saul, "Implicit Bias, Stereotype Threat and Epistemic Injustice", 236.

injustice may be “[...] especially hard to detect from the outside, for it is by definition literally passed over in silence.”¹⁹⁵ While I think that this assertion is perfectly legitimate, I want to propose another reason why pre-emptive testimonial injustice is hard to detect, namely the likelihood of its relation to implicit bias.

Think back to the example of Juliet, the university professor. Juliet’s implicit bias against black students influences the ways she interacts with them. Let us assume that she discusses a difficult topic in class and automatically primarily addresses her white students because her implicit biases prompt her to think that her black students are not up to the task. In the discussion, Juliet thus only solicits the opinion of white students, which leads to a pre-emptive silencing of her black students. Thus, through drawing on the likelihood of implicit biases that cloud Juliet’s judgements of her black students, it becomes clear why such pre-emptive cases of testimonial injustice are harder to detect.

I suggest that if Juliet became aware of her implicit biases towards black students, then she would also decrease the possibility of perpetrating (pre-emptive) testimonial injustices to them. This raises the following challenging question: how could Juliet become aware of her implicit attitudes against black students that consequently influence how she interacts with them? One possibility would be that an outsider observes Juliet’s divergent behavior towards her black and white students and points it out to her. This, of course, would help Juliet decode her implicit biases against black students. Yet, a more desirable and, in my view, effective way would be for Juliet herself to become aware of her implicit attitudes. One way of doing so would be to reflectively, cognitively realize that her behavior is influenced by implicit biases. This is in line with the way Fricker proposes to counteract instances of testimonial injustice. Fricker puts forward the idea of a virtue of testimonial justice coming about through a distinctively *reflexive* critical awareness which the prejudiced individual should perform. However, this suggestion presents a severe challenge for instances of testimonial injustice in which the prejudice operates on an unconscious level, as is the case with implicit biases. In other words, how should explicit cognition regulate implicit attitudes?¹⁹⁶

¹⁹⁵ Cf., Fricker, *Epistemic Injustice*, 131.

¹⁹⁶ I will discuss this criticism in more detail in the next chapter when I present remedies against testimonial injustice.

What makes the detection of implicit biases even more challenging is the fact that they do not only operate in a belief-like manner. Rather, as already mentioned when we discussed the metaphysical status of implicit biases, the psychological and philosophical literature also describes them as “associative states”.¹⁹⁷

3.2.1 Aliefs

In this section I will focus on a concept introduced to philosophical research on implicit attitudes by Tamar Szabó Gendler, namely “aliefs”. My aim is to present an alternative to belief-based accounts of implicit biases. After elaborating on the structure and functioning of aliefs, the thesis will return to Fricker’s text in order to investigate if certain parallels between testimonial injustice and aliefs can be drawn.

Alief is a neologism with the connotations of “belief” in mind. And yet alief as a concept is designed to illustrate exactly the opposite, namely mental states that do not have belief-like structure. More specifically, Gendler defines aliefs as follows:

To have an alief is, to a reasonable approximation, to have an innate or habitual propensity to respond to an apparent stimulus in a particular way. It is to be in a mental state [...] that is *associative*, *automatic* and *arational*. As a class, aliefs are states that we share with non-human animals; they are developmentally and conceptually *antecedent* to other cognitive attitudes that the creature may go on to develop. Typically, they are also *affect-laden* and *action-generating*.¹⁹⁸

Aliefs thus are inherent or habitual associative mental states that occur as a response to a perceptual stimulus. Let us think this through by drawing on examples Gendler provides to illustrate aliefs.

I am standing on a transparent walkway traversing over the Grand Canyon. While I *believe* that the glass construction is perfectly safe, since I have the empirical evidence of thousands of people walking across it every year I nonetheless *alieve* a form of danger to be present. This activation of a feeling of danger may influence my physical behavior. I will turn my eyes downward and try to measure the distance between the walkway I am standing

¹⁹⁷ <https://plato.stanford.edu/entries/implicit-bias/#Met>

¹⁹⁸ Tamar Szabó Gendler, *Intuition, Imagination & Philosophical Methodology*, (Oxford: Oxford University Press, 2010), 288.

on and the Colorado River. This may cause a feeling of vertigo and my starting to tremble.¹⁹⁹ “Eventually, despite my uncompromised belief in the safety and stability of the structure on which I am standing, [...] I may find myself in a powerful – though reflectively disavowed – contrary inclination to remove myself from it [the walkway].²⁰⁰ This example demonstrates how aliefs function. They trigger action in an unconscious, associative manner.

Aliiefs are at play in all sorts of situations. For instance, they are the reason why we are reluctant to drink lemonade out of sterilized bedpans, or throw darts at a picture of a loved one.²⁰¹ What happens in cases of aliefs is that our reflective beliefs suggest one type of response to a stimulus, whereas “[...] our implicit associations and habitual patterns render occurrent another sort of response routine.” These examples are unsettling since they suggest that certain mental states are automatically activated. And yet they need not be a cause for concern from an ethical point of view. Aliefs can become harmful, however, when their content contains stereotypical racist or sexist associations. For instance, even if my beliefs are totally anti-racist on a conscious level, I could still alieve danger when I find myself alone in a neighborhood where only people of color live.²⁰²

What happens in this example and in the examples given above is that some sense of *should* is violated.²⁰³ When walking on the glass construction I *should* feel safe. It *should* not disgust me to drink lemonade out of a sterilized bedpan and, most importantly, I *should* feel safe in a black neighborhood, even more so if my conscious beliefs do not entail racial prejudices like: “black persons are criminals”.

Aliiefs can therefore be described as triggering belief-discordant behavior. Gendler explains this by contrasting aliefs with other mental states such as belief and imagination in terms of their relation to acceptance of propositional attitudes.²⁰⁴ “Unlike belief or pretense or imagination or supposition, alief does not involve acceptance.”²⁰⁵ Consider someone who actively endorses stereotypical racist attitudes and someone who strongly objects to

¹⁹⁹ Cf., *ibid.*

²⁰⁰ *Ibid.*, 41-42.

²⁰¹ Cf., Gendler, *Intuition, Imagination & Philosophical Methodology*, 286.

²⁰² Gendler calls such cases “aversive racism”. (Cf., Gendler, “On the Epistemic Costs of Implicit Bias”, 43.)

²⁰³ Cf., Gendler, *Intuition, Imagination & Philosophical Methodology*, 283.

²⁰⁴ Cf., *ibid.*, 270.

²⁰⁵ *Ibid.*, 268.

stereotypical racist attitudes, but knows that they exist and that many people hold them to be true. In an all-black neighborhood, the first person will feel threatened as a result of their beliefs about black people. However, the second person might feel threatened, not because they believe in the stereotype, but rather due to their awareness of the stereotype. They do not accept it, but still it triggers the alief: “Black person. Danger. Cross the street!” “At its core, alief involves the activation of an associative chain – and this is something that can happen regardless of the attitude that one bears to the content activating the associations.”²⁰⁶ In other words, merely knowing what others believe about a social group can lead to stereotype-concordant behavior that contradicts the individual’s conscious beliefs.

Awareness of certain stereotypes is therefore enough to trigger associative chains, leading the affected individual to act in ways discordant with their beliefs. This claim can also be applied to instances of stereotype threat. As I have demonstrated above, stereotype threat impairs performance in many ways; for example, it leads to avoidance, disengagement and loss of self-confidence. Gendler provides a further perspective on stereotype threat by linking it to her notion of aliefs. She argues that stereotype threat impairs performance because it temporarily interferes with the accessibility of knowledge. To quote an illustration, someone affected by stereotype threat may be unable to recall the date of the French Revolution, although the person memorized this date a long time ago and never had any problems remembering it. Furthermore, subjects may temporarily lose confidence in their true beliefs. They may, for instance, triple-check if 11×11 is 121, even though this kind of multiplication never has posed any difficulty before.²⁰⁷ Gendler argues that the activation of self-referential cultural stereotypes does not necessarily imply that the affected subject holds the stereotype to be true. Rather, it is enough that the subject is aware of the existence of the stereotype in order to be affected by it. Consequently, the alief of a victim of stereotype threat could, for instance, have the content: “‘Female’ applies to me and ‘female’ is associated with poor math performance; (anxiously) better make sure that I’m doing these math problems correctly; double-check, double-check, double-check.”²⁰⁸

²⁰⁶ Ibid., 270.

I think this fact is a really important insight regarding implicit attitudes and their connection to instances of epistemic injustice. The thesis will elaborate on this in the next section.

²⁰⁷ Cf., *ibid.*

²⁰⁸ Ibid., 51.

At this point one should mention that the concept of aliefs is confronted with some criticism within implicit-bias research.²⁰⁹ Even Gendler herself asserts that her deliberation on aliefs is work in progress and therefore risks being fuzzy.²¹⁰ Still, I think that some of the insights Gendler's concept of aliefs grants us are useful tools when assessing Fricker's account of epistemic injustice. The next section will demonstrate how the concept of alief could improve Fricker's theory.

3.2.2 Aliefs in the Context of Epistemic Injustice

In the course of Fricker's argumentation there are a number of implications that indicate kinship with Gendler's concept of aliefs. In the following section I will summarize them under three aspects that demonstrate to what extent aliefs could serve as an explanatory improvement to certain mechanisms at stake in instances of epistemic injustice. While the first aspect concerns the metaphysical status of identity prejudices in Fricker's account, the second and third aspects investigate how aliefs and identity prejudices coincide in their manner of activation.

As mentioned in the second chapter of the thesis, Fricker asserts that instances of epistemic injustice arise in situations where identity power is exercised. This exercise of identity power relies on an *imaginative* social co-ordination. In other words, agents have shared conceptions of social identities – “[...] conceptions alive in the collective social imagination that govern, for instance, what it is or means to be a woman or a man, or what it is or means to be gay or straight, young or old, and so on.”²¹¹ Moreover, Fricker claims that conceptions of social identities are not necessarily held at the level of belief, “[...] for the primary *modus operandi* of identity power is at the level of the collective social imagination.”²¹² Conceptions of social identity can thus influence our actions despite our beliefs. The metaphysical status of stereotypes and identity prejudices is not captured only through describing them as beliefs. Rather, Fricker's definition of identity prejudice suggests

²⁰⁹ For a further discussion of criticism about aliefs see: <https://plato.stanford.edu/entries/implicit-bias/#SuiGenSta>

²¹⁰ Cf., Gendler, *Intuition, Imagination & Philosophical Methodology*, 263.

²¹¹ Fricker, *Epistemic Injustice*, 14.

²¹² *Ibid.*, 15.

that she understands them to be associative states.

This can also be inferred from the manner in which Fricker describes negative identity prejudice, which is the primary factor for the emergence of testimonial injustice. Remember how Fricker explicitly describes identity prejudice as “[...] *an association between a social group and one or more attributes, where this association embodies a generalization that displays some [...] resistance to counter-evidence owing to an ethically bad affective investment.*”²¹³ Thus, it seems that from Fricker’s point of view not only implicit, but also explicit prejudices manifest an associative structure. Obviously, the latter are actively endorsed by the individual and thus are held at the level of belief at some point. But what originally initiates them is an association. Implicit identity prejudices, or “residual internalizations” as Fricker labels them, arise through associations that the subject creates without conscious control. The same can be said for aliefs. As we have seen, aliefs are associative mental states that arise through a perceptual stimulus and lead to belief-discordant behavior. Aliefs can therefore illustrate how associations that do not come about consciously cloud our credibility judgements in instances of testimonial injustice.

The second aspect aims to highlight another reason why one can characterize identity prejudices as associative and related to aliefs, namely the image-like character of identity prejudices. By referencing Lippmann, Fricker describes social stereotypes as images that express an association between a social group and one or more attributes.²¹⁴ This image-like character of stereotypes also applies to prejudices. Moreover, Fricker claims that prejudicial images may impact our judgement in a manner that escapes our conscious control. “This is most starkly illustrated when the influence of prejudicial images from the social imagination persist in a hearer’s patterns of judgement even where their content conflicts with the content of her beliefs.”²¹⁵

Aliefs also display a deep connection with images. Think back to the case where one believes that it is perfectly safe to drink lemonade from a sterilized bedpan, but alieves that the bedpan is unclean and thus refuses to drink from it. The reason why we are reluctant to drink from bedpans, even if they are sterilized, is that we have a certain image of bedpans

²¹³ Ibid., 35. (Original quote in italics as well.)

²¹⁴ Cf., *ibid.*, 37.

²¹⁵ Ibid.

and how they are actually used. The perceptual stimulus of seeing a bedpan thus triggers the image of its actual use and further activates associations leading to an alief with the content: “Bedpans. Yuck. Do not use as a drinking device.”

Finally, the third aspect concerns the mode of activation of residual identity prejudices and aliefs. As I have shown above, Gendler argues that aliefs are activated through stereotypical associative chains.²¹⁶ This activation is already triggered by mere awareness of the stereotype: “Moreover, awareness of these stereotypes is sufficient to give rise to the relevant associative chains.” If I know that stereotypical images consider black people to be, for instance, “[...] lazy, ignorant, musical, unreliable, loud, aggressive, low in intelligence [...] then – at least some of the time – those associations will be triggered by my thoughts about interactions with members of such groups.”²¹⁷

This thesis argues that the activation of a stereotypical associative chain through simple awareness of the stereotype can also be discovered in Fricker’s account. For instance, by elaborating on her concept of residual internalizations, she distinguishes diachronic from synchronic ways in which prejudicial residue may subsist in a person’s consciousness even when it conflicts with the person’s remaining beliefs.²¹⁸ The diachronic case is exemplified by cases such as the one presented in the thesis. A woman consults two doctors, a man and a woman, and tends to intuitively put more weight on the male doctor’s diagnosis. Though the woman’s beliefs have moved on, residues of her upbringing, which has been gender normative, still persist in her unconsciousness and thus influence her judgements. The synchronic case can be exemplified by a committed anti-racist “[...] whose patterns of social judgement none the less betray a residue from racist elements that are contained in the collective social imagination.”²¹⁹ In the examples given to illustrate the diachronic and the synchronic case of residual internalizations of prejudices, both subjects do not endorse the stereotypes that influence their judgement. Rather, the simple awareness of the stereotypes suffices to activate them.

²¹⁶ Cf., Gendler, “On the Epistemic Costs of Implicit Bias”, 43.

²¹⁷ Ibid.

²¹⁸ Cf., Fricker, *Epistemic Injustice*, 39.

²¹⁹ Ibid.

3.3 The Improvement of Fricker's Account through Implicit-Bias Research

I propose drawing on implicit-bias research because of a neglect arising in Fricker's theory regarding unconscious prejudice. I argue that implicit-bias research is the best scientific explanation available to remedy this neglect.²²⁰ Throughout this chapter it was shown how implicit-bias research could be used to complement Fricker's account concerning the role of implicit residual prejudices in instances of testimonial injustice.

Implicit-bias research addresses metaphysical, epistemic and ethical challenges that are of interest for Fricker's theory of testimonial injustice as well. For instance, by drawing on Gendler's concept of aliefs this thesis demonstrated that Fricker's description of identity prejudices intersects with aliefs in terms of their associative, image-like structure and their mode of activation.

Moreover, similarities between implicit-bias research and epistemic injustice are striking when we look at the negative practical and epistemic ramifications of stereotype threat and testimonial injustice. In other words, implicit biases that lead to cases of stereotype threat are likely the same biases that lead hearers to perpetrate testimonial injustices against speakers.

Additionally, the "Implicit Association Test" introduced above measures implicit attitudes in individuals. Findings in these tests demonstrate how likely it is that we all have implicit biases that affect our judgements and, more importantly for the purposes of this thesis, our credibility judgements.

Still, if implicit prejudice impacts a hearer's perception of members of marginalized groups, then there is hope that the hearer's conscious beliefs about these groups should be able to serve as a corrective force for the implicit prejudicial images. Becoming aware of our implicit biases is an important step towards an improvement in how we conduct and regulate epistemic exchange.

²²⁰ In some way, this argument can be framed as a methodological naturalist argument. To deepen our understanding of how implicit residual prejudices function within Fricker's account, one should look at the best scientific theories available to explain implicit attitudes. These are provided by psychological research on implicit biases. The "IAT", for instance, presents a possibility to measure implicit preferences regarding in-groups. (Cf., Saul, Brownstein, *Implicit Bias and Philosophy*, Volume 1, 6.) These implicit preferences help to explain why hearers implicitly perceive certain speakers as less credible than others.

4. Normative Questions Related to Testimonial Injustice and Implicit Bias

The aim of this chapter is to explore normative questions arising in implicit-bias research. Moreover, I will argue that challenges regarding implicit bias also are relevant for Fricker's account of epistemic injustice. Influence by implicit bias is unsettling, especially when we realize how pervasive implicit prejudice really is: we have seen how likely implicit biases affect our basic communicative relations, our perceptions, and consequently, our credibility judgements. But how much responsibility and blame should we bear for an unconscious phenomenon that is automatically activated and not easily detectable via introspection? Can we regulate implicit prejudice and its influence on our credibility judgements and if so, will this render antidotes against testimonial injustice more successful? Addressing these questions is the task of this chapter.

4.1 Bias-Related Doubt

In her 2012 paper "Skepticism and Implicit Bias" Saul explores the idea that "[...] what we know about implicit bias gives rise to something *akin to* a new form of skepticism."²²¹ She calls this phenomenon "bias-related doubt" and argues that in light of our knowledge that implicit bias is likely to influence our judgements, we have good reason to believe that we cannot fully trust our knowledge-seeking faculties.²²²

This thesis argues that this observation is highly relevant for Fricker's account of testimonial injustice. How can we trust our credibility judgements when it is likely that our perceptions and our reasoning are affected by the influence of implicit bias? Before I can comment on this question, it is crucial to gain a better understanding of what bias-related doubt is. In order to do so, I will first investigate what exactly sparks Saul's assumption that implicit bias gives rise to a form of skepticism. Secondly, I will investigate the moral and epistemic consequences of bias-related doubt by comparing the phenomenon to traditional skeptical scenarios.

Saul asserts that bias-related doubt is both stronger and weaker than traditional forms of skepticism. Since implicit bias does not affect all of our judgements, we can safely say that

²²¹ Saul, "Skepticism and Implicit Bias", in: *Disputatio*, Vol. V, No. 37 (2012), 243-263, 243.

²²² Cf., *ibid.*

we are not mistaken about everything in the external world. This renders bias-related doubt weaker than traditional skepticism. However, findings in implicit-bias research suggest that we are mistaken about a great number of things. Furthermore, the kind of doubt triggered by implicit bias poses threats to our knowledge-seeking faculties that demand action:²²³

With traditional skepticism, we feel perfectly fine about setting aside the doubts we have felt when we leave the philosophy seminar room. But with bias-related doubt, we don't feel fine about this at all. We feel the need to do something to improve our epistemic situation.²²⁴

The phenomenon of bias-related doubt is thus stronger than traditional skepticism because it is more concrete. It should worry us that many people are likely to prefer CVs with typically white-sounding names to CVs of applicants with black- or Arab-sounding names. It should worry us that prestige bias suggests that institutional affiliation “[...] has a dramatic effect on the judgements made by reviewers [...]”²²⁵ regarding the evaluation of academic papers. It should worry us that the persons making such judgements are unaware of what role social identity and prestige play when forming their opinion of candidates and papers. In other words, our judgements are likely to be influenced by factors that are irrelevant to their formation.

As demonstrated above, implicit bias operates on an unconscious level of our cognitive processing. Furthermore, studies of shooter-bias that I have introduced at an earlier point in this thesis suggest that implicit bias already influences the most basic level of our cognition, namely our perceptions. This shows “[...] that implicit bias is getting to us even before we get to the point of reflecting upon the world – it affects our very perceptions of that world, again in worrying ways.”²²⁶ This stings and has a severe impact on our epistemic self-esteem.

Let us reflect upon some consequences of the ways in which implicit bias affects our judgements. Saul first presents consequences that have obvious moral and political implications. Making unconscious inaccurate judgements about who is the best candidate for a job or about which paper to submit is not only disturbing to ourselves. Rather, “[...] it means that we are being dramatically unfair in our judgements, even though we are doing so

²²³ Cf., *ibid.*, 243.

²²⁴ *Ibid.*, 244.

²²⁵ *Ibid.*, 245.

²²⁶ *Ibid.*, 246.

unintentionally.”²²⁷ Moreover, making these inaccurate judgements ensures that unfair treatment is prolonged and consequently the stereotype that originally caused and created our implicit bias will be maintained.²²⁸

Saul’s primary focus lies in the epistemological aspects of these unfair situations. She illustrates this by drawing on the field of philosophy: “When we misjudge a paper’s quality, we’re making a mistake about the quality of an argument.”²²⁹ Furthermore, the evaluation of the paper is influenced by something completely irrelevant to the paper’s quality, namely by the knowledge of the social identity of its author. In addition to that, the irrelevant influence operates on a level that is unconscious and thus inaccessible to inspection and rationality. Saul consequently infers that in philosophy we may accept arguments that we should not accept and we may not reject arguments that we should reject. “Many of our philosophical beliefs – those beliefs we take to have been arrived through the most careful exercise of reason – are likely to be wrong.”²³⁰

The same can be said about the sciences. This, however, is no news to us. We are aware of the probability that most of what is accepted as science nowadays will likely be proven false within the coming centuries. However, as Saul puts it, “[...] my claim is not that we’re likely to be accepting some falsehoods, or even a lot of falsehoods. That’s not unsettling. My claim is that we’re likely to be *making errors*.”²³¹ These errors are of a very specific sort: we let the social identity of a person making an argument or giving testimony affect our evaluation of the argument or testimony.²³² “It is part of our self-understanding as rational enquirers that we will make certain sorts of mistakes. But not this sort of mistake. These mistakes are ones in which something that we actively think *should not* affect us does.”²³³

Bias-related doubt therefore is worrying, and, as Saul asserts, it is more worrying than traditional skeptical scenarios. We cannot leave the problems of implicit bias behind as we leave behind worries about brains in vats and barn façades. Outside of constructed thought

²²⁷ Ibid.

²²⁸ Cf., *ibid.*

²²⁹ *Ibid.*, 247.

²³⁰ *Ibid.*

²³¹ *Ibid.*

²³² Cf. *Ibid.*

We will reflect upon bias-related doubt and its relation to credibility judgements at the end of this section.

²³³ *Ibid.*

experiments we do not really doubt the existence of the external world. However, not really believing in the scenarios presented by traditional skepticism does not mean that we act irresponsibly. “Bias-related doubt is different from that, though. It is very much a real doubt.”²³⁴

In order to provide further insight into what exactly makes bias-related doubt so pressing and worrying, Saul compares it with “live skeptical scenarios”. Referring to Bryan Frances’ work on live skeptical scenarios, she characterizes traditional skeptical arguments “[...] as relying on facts that cannot be ruled out.”²³⁵ However, just because they cannot be ruled out does not mean that they are “live”. This means that they are so implausible that we cannot take them seriously. Frances argues that there are skeptical scenarios that are not like this. He provides live skeptical settings that contain “[...] compelling scientific and philosophical reasons to think that the hypotheses are actually true.”²³⁶ And yet Saul claims that bias-related doubt is still stronger and gives us more reason to worry than live skeptical scenarios. Her claim is “[...] not just that the hypothesis is live – that sensible and knowledgeable people might endorse it on the basis of good reason. Instead, it’s that *we all have very good reason to believe that it is true.*”²³⁷

Bias-related doubt thus makes a stronger claim than just saying that a hypothesis is “live”, since it compels us to say that the hypothesis is true. This calls for action and poses a real challenge to an image of ourselves that we aim to maintain, namely the image of critical enquirers.

4.1.1 Irrelevant Influences on Belief

In the following I will briefly discuss an account that is not concerned with cases of skepticism, but still is similar to Saul’s concept of bias-related doubt. In her 2018 paper “Irrelevant Influences” Katia Vavova examines how factors such as upbringing, socialization, university affiliation or one’s birthplace, influence beliefs on specific issues.

²³⁴ Ibid., 252.

²³⁵ Ibid., 254.

²³⁶ Ibid.

Frances refers to things like eliminativism about belief or error theory about colour in order to give examples for such live hypotheses.

²³⁷ Ibid., 255. (Original quote in italics as well.)

She calls such factors “irrelevant influences” on belief and exemplifies them by presenting the following scenario. In 1961 G.A. Cohen chose Oxford over Harvard as his preferred graduate school. After having finished his degree, he realized that philosophers of his generation, who studied with him at Oxford, tend to accept the analytic/synthetic distinction, whereas students, who studied at Harvard, despite having been presented with the same arguments, tend to reject it. “This lead Cohen to worry that in some sense of ‘because’ and in some sense of ‘Oxford’, he accepts the analytic/synthetic distinction because he studied at Oxford.”²³⁸ Cohen thinks that this is disturbing since the mere fact of having studied in Oxford is no reason to believe that the distinction is correct.²³⁹

Vavova proceeds by defining irrelevant influences on belief as follows: “[A]n irrelevant influence for me with respect to my belief that *p* is one that (a) has influenced my belief that *p* and (b) does not bear on the truth of *p*.”²⁴⁰ But what is so unsettling about irrelevant influences on belief? The answer seems obvious. It is desirable to believe what we believe because of evidence and sound arguments and not because of where we were born, or what school we attended.²⁴¹

In some way then, irrelevant influences create the same kind of worry implicit biases do, namely that our beliefs or belief-like states are not reliably formed. Furthermore, the definition of irrelevant influences also seems in line with the way implicit biases influence our beliefs and judgements. They (a) implicitly influence my belief or belief-like state that *p* and (b) do not bear on the truth of *p*.²⁴² Let us assume that *p* equals “women are bad at mathematics.” Implicit bias therefore influences my belief that women are bad in mathematics even if (b) this does not bear on the truth of my belief, because, for instance, many women nowadays are gifted mathematicians. I could still implicitly believe that women are bad mathematicians even if I had the necessary evidence to prove that women are just as good in mathematics as men are. At an earlier point in the thesis I have pointed out that implicit bias to some extent is resistant to counter-evidence. Thus, cases in which my beliefs

²³⁸ Katia Vavova, “Irrelevant Influences”, in: *Philosophy and Phenomenological Research*, Vol. XCVI, No. 1 (2018), 134-152, 134.

²³⁹ Cf., *ibid.*

²⁴⁰ *Ibid.*

²⁴¹ Cf., *ibid.*

²⁴² We have discussed at an earlier point that the metaphysical status of an implicit bias is not necessarily the one of a belief. Implicit bias can also be understood as an associative state.

are influenced by implicit biases are worrisome because these beliefs are resistant to counter-evidence and thus not reliably formed.

A similar line of argument can be traced in Vavova's paper. "The worry about irrelevant influences is thus not that someone with your evidence believes otherwise. Nor is it exactly that an alternate you would have believed otherwise. It is that your actual beliefs may not be reliably formed."²⁴³ However, Vavova claims that this does not mean that we have to become skeptics about our beliefs. As she puts it "[...] this is not a worry about disagreement or skepticism. It's a worry about irrationality or error."²⁴⁴

Contrarily, according to Saul, knowledge about implicit bias can lead us to become skeptics about our beliefs. However, as stated above, bias-related doubt is stronger than traditional skepticism as the threat that it poses to the formation of our judgements and beliefs is real and not merely a constructed scenario. Thus, in cases where implicit bias has informed our judgements, these judgements are prone to errors of a specific sort. As Saul explains, the error concerns the fact that we are likely to let the social identity of a person affect the way we evaluate the person's argument, application, testimony etc.²⁴⁵ Comparing the effects of implicit biases to Vavova's account of irrelevant influences on belief thus seems sound. Implicit bias does not only suggest that we should become skeptics about our beliefs, but that many of our beliefs are actually wrong.

Both Vavova and Saul claim that the evidence of error should compel us to change our behavior and the way in which we let implicit bias and other irrelevant factors influence our judgements.²⁴⁶ When it comes to implicit biases this seems to be a particularly difficult task since their implicit, unconscious structure makes them especially hard to detect. Saul presents several ways through which we can try to reduce our implicit biases, such as active blocking and thinking of counter-stereotypical examples.²⁴⁷

Vavova's account also provides a way of distinguishing innocuous from problematic

²⁴³ Ibid., 142.

²⁴⁴ Cf., *ibid.*

²⁴⁵ Cf., Saul, "Skepticism and Implicit Bias", 249.

²⁴⁶ Cf., Saul, *Ibid.*, 255-260, Cf., Vavova, "Irrelevant Influences", 135.

²⁴⁷ Cf., *ibid.*, 257.

We will discuss remedies against implicit bias more thoroughly in the last section of the thesis, since I will argue that a sustainable antidote against testimonial injustice can only be achieved when we consider methods of reducing the influence of implicit bias.

cases of irrelevant belief influences. At this point, one might ask as to whether or not the method she presents can be applied to cases of implicit bias. Vavova proposes to evaluate the epistemic situation that led to the error by coming up with her “good independent reason principle” (GIRP).²⁴⁸ The GIRP goes as follows: “To the extent that you have good independent reason to think that you are mistaken with respect to p , you must revise your confidence in p accordingly – insofar as you can.”²⁴⁹ In what follows, I will ask as to whether or not implicit bias can be analyzed according to this principle.

First, Vavova makes it clear that the GIRP does not apply to skeptical situations of doubt because evidence of irrelevant influence indicates an ordinary sort of error. This sort of error occurs because we are fallible creatures, but not because we generally doubt all our knowledge-seeking faculties, as traditional skepticism does.²⁵⁰ Therefore, one could argue that the GIRP cannot be applied to Saul’s concept of bias-related doubt. And yet, as we have seen, bias-related doubt is *not* like traditional skepticism. It is stronger because the evidence of error arising in bias-related doubt scenarios is real. Bias-related doubt thus points to a kind of error that could be corrected by applying the GIRP.

The second way in which the GIRP might not apply to cases of implicit bias is harder to dismiss. Vavova suggests, “[...] that there may even be cases in which you cannot correct at all because you cannot know how you’ve been influenced [...]”.²⁵¹ This seems to be the case in an implicit bias-kind of influence. Since implicit bias cannot be traced through mere introspection, it likely influences our beliefs without our knowledge. Yet, evidence about the influence of implicit bias exists. The IAT, shooter-bias test and other studies have demonstrated that we are very likely to have implicit biases. In some cases, this should count as a “[...] good independent reason to think that we are mistaken with respect to p ,” which means that the GIRP would apply. Subsequently, we can conclude that Vavova’s GIRP principle can be used in instances of bias-related doubt.

Finally, it all boils down to the question of how it is possible to trace and counteract for our biases. I will elaborate on this at a later point when considering antidotes against implicit prejudice and testimonial injustice.

²⁴⁸ Cf., Vavova, “Irrelevant Influences”, 145.

²⁴⁹ Ibid.

²⁵⁰ Cf., *ibid.*

²⁵¹ Ibid., 145.

4.1.2 Bias-Related Doubt, Irrelevant Influences and Credibility

How do phenomena such as bias-related doubt and irrelevant influences apply to Fricker's account of testimonial injustice? At an earlier point in this thesis I criticized that Fricker is not sufficiently committed to deliberations about possible implicit factors that lead to instances of testimonial wrongdoing.

Fricker mainly bases her concept of credibility judgements on the mediation of social stereotypes. As we have seen, this is not problematic when the stereotype is reliable. However, as her theory of testimonial injustice suggests, in a lot of instances these stereotypes are not reliable because they are prejudiced. Even more so, they are implicitly prejudiced, or as Fricker puts it, they operate on a non-doxastic level.²⁵²

I believe that the right vision of epistemic relations is such that testimonial injustice goes on much of the time, and while it may be hard enough to police one's beliefs for prejudice, it is significantly harder to filter out the prejudicial stereotypes that inform one's social perceptions directly, without doxastic mediation.²⁵³

I want to hold on to Fricker's claim and therefore argue that it should be crucial that her account includes a method to distinguish irrelevant but innocuous from epistemically problematic influences on beliefs and credibility judgements, such as implicit bias.²⁵⁴

Furthermore, Fricker's theory contains numerous aspects that suggest linking it to implicit-bias research. Her model of testimony for instance, emphasizes the role that perception plays in testimonial exchange. Hearers *perceive* their interlocutor in a way that is epistemically charged. Moreover, an ideal perception for Fricker is one that is also virtuous. In other words, the virtuous hearer has the capacity to spontaneously perceive the speaker as more or less trustworthy, relying on certain background assumptions that are related to different social types in different social contexts.²⁵⁵ This perception, however, is not always virtuous. Rather, because it relies on background assumptions related to social types, such as stereotypes, the perception is prone to be influenced by implicit bias.

This is exactly what Saul demonstrates. Implicit bias affects us even before we

²⁵² Fricker, *Epistemic Injustice*, 36.

²⁵³ Ibid.

²⁵⁴ Fricker is not completely oblivious to this. In a later section I will draw on an account by Fricker that distinguishes between culpable and non-culpable implicit prejudices.

²⁵⁵ Cf., *ibid.*, 72.

consciously try to make sense of the world because it “[...] affects our very perceptions of the world [...] in worrying ways.”²⁵⁶ Bias-related doubt therefore suggests that we are likely making errors when evaluating testimony. Although Saul asserts that this kind of doubt does not only concern questions of credibility, I argue that it is nevertheless crucial to investigate the impact of implicit bias on our credibility judgements and consider if we have to become skeptics about them.

Residual identity prejudices should be understood as irrelevant influences on our judgements that give rise to reason for doubt. Even more so, they suggest that an actual error has occurred in our judgements. This should be taken seriously. Furthermore, just because implicit biases operate on an unconscious level and are hard to detect does not mean that implicit biases’ probable influence is a “carte blanche” to free us from responsibility for our actions. To the contrary, it should compel us to be even more vigilante when we assess the testimony of marginalized social groups.

4.2 Epistemic Responsibilities and Obligations

“[I]n the story of one’s life there is an authority exercised by what one has done, and not merely by what one has intentionally done.”²⁵⁷

Bernard Williams (*Shame and Necessity*)

In the last section it was shown that implicit bias likely leads to error in our judgements. The unsettling results gathered in implicit-bias research should motivate us to interrogate if our beliefs and credibility judgements are reliably formed. One might ask if we can be held responsible for implicit discriminatory and prejudicial behavior when such behavior occurs unintentionally and inadvertently. In other words, does acting responsibly imply that we are always fully aware of our beliefs and their formation? Moreover, does the influence of implicit bias not only create an epistemic failure, but also an ethical, political one?

This section will show how agents can be conceptualized as responsible for their actions, even when the beliefs that set their behavior in motion are informed by implicit bias. I will

²⁵⁶ Saul, “Skepticism and Implicit Bias”, 245.

²⁵⁷ Bernard Williams, *Shame and Necessity*. (Berkeley: University of California Press. 1993), 69.

start by presenting an account by Natalia Washington and Daniel Kelly that clarifies how we can be held responsible for our implicit biases despite their automatic activation and the difficulty of their detection.

In the following I argue that knowledge about implicit bias forces us to cautiously assess how our beliefs and judgements are formed. This will be underlined by an example by Medina who illustrates how ignorance can lead to culpable epistemic error. However, knowledge about implicit bias is not equally accessible to everyone. Rather than holding everyone accountable for ignorance and implicit biases to the same degree, we need to come up with a comparative system of responsibility that takes into consideration what kind of knowledge is available to whom.

Finally, I will draw on an account by Fricker that considers the impact of implicit prejudice on our judgements. Interestingly, Fricker provides a way of thinking about epistemic responsibilities for non-culpable implicit-bias influence, such as epistemic bad luck. In reference to Bernard Williams, Fricker calls this concept “epistemic agent-regret”. Considerations about responsible agency in testimonial exchange will eventually clear the way for thinking about possible antidotes for testimonial injustice.

4.2.1 Epistemic Responsibilities and Implicit Bias

A basic assumption to be made when talking about epistemic responsibilities is that people are responsible for their behavior in the majority of cases. And yet people should not be held morally responsible for all their behaviors. For instance, there are certain exculpating conditions that could free agents of their responsibility in some situations. Drawing on a paper by Natalia Washington and Daniel Kelly, I will present such exculpating conditions and clarify why implicit bias should not always be understood as falling under these conditions.

Consider two variations of the following scenario. “Cate eats a batch of cookies that her roommate made especially for tomorrow’s bake sale.”²⁵⁸ In the first variation, Cate is fully

²⁵⁸ Natalia Washington, Daniel Kelly, “Who’s Responsible for This? Moral Responsibility, Externalism, and Knowledge about Implicit Bias”, in: *Implicit Bias and Philosophy. Volume 2. Moral Responsibility, Structural Injustice and Ethics*, ed. Saul, Brownstein, (Oxford: Oxford University Press, 2016), 11-37, 15.

aware that the cookies are meant for the bake sale, yet she eats them anyway, simply because she is hungry. In a second variation, let us assume that Cate eats the cookies that her roommate made for the bake sale while she is sleepwalking.²⁵⁹ In this case, Cate did not know what she was doing and also had no control over her actions. She thus should be exculpated from blame, since people are not generally held responsible for their behavior when they are sleepwalking.

Therefore, the kind of behavior portrayed in the second scenario illustrates two kinds of exculpating conditions that Washington and Kelly put forward: knowledge and control. They label them the “knowledge condition” and the “control condition”.²⁶⁰ Yet, exculpating conditions also have an exception clause. To understand this, consider a third variation of our previous example. “Cate, while in a somnambulant daze, eats a batch of the cookies that her roommate made especially for tomorrow’s bake sale, but she was in that somnambulant daze because she had taken a hefty dose of Ambien.”²⁶¹ Moreover, Cate has a long history of sleepwalking and binge eating whenever she takes Ambien. According to Washington and Kelly, this case illustrates an exception clause to exculpating conditions.

Somnambulant Cate is unaware of and unable to consciously control what she is doing, but, like a drunk driver, she is responsible for having put herself in that compromised condition, and she is blameworthy for what she does once she inhabits it – in this case, especially since she has a well-known history of such Ambien-induced destructive behavior.²⁶²

We can therefore infer that agents are exculpated for doing X if (a) they do not know that they do X and if b) they do not have control over doing X. These exculpating conditions, however, do not apply when the agent is responsible for having been unaware of doing X or when he or she is responsible for lacking control.²⁶³

Washington and Kelly proceed by considering if implicit bias could be understood as such an exception clause to exculpating conditions. As I have investigated above, implicit biases are not easily detected via introspection. Think back to Juliet, the university professor, who was completely unaware of the fact that she implicitly regarded her black students as

²⁵⁹ Cf., *ibid.*

²⁶⁰ Cf., *ibid.*, 16.

²⁶¹ *Ibid.*

²⁶² *Ibid.*

²⁶³ Cf., *ibid.*, 16-17.

less capable than her white students. Furthermore, implicit biases do not only work implicitly, but are activated automatically, a fact that Washington and Kelly label the “recalcitrance” of implicit biases.²⁶⁴ These two characteristics about implicit bias align neatly with the exculpating conditions to situations of blameworthiness indicated above, namely knowledge and control. The existence and influence of implicit bias is not easily detectable, which means that an individual can be implicitly biased without *knowing* it. In addition to this, implicit biases are recalcitrant which means that it is hard to *control* their activation.²⁶⁵

This alignment could lead to the following epistemic worry. “Since implicit biases are opaque to introspection and can operate outside of conscious awareness, a person should be exculpated, and not blamed or held responsible for behaviors that manifest them.”²⁶⁶ Washington and Kelly dismiss this worry by arguing that the possibility of knowledge about implicit biases has important implications for responsibility and blame. They provide a thought experiment called “The Hiring Committee”, which illustrates how, in some cases, implicit bias could satisfy the exception clause to exculpating conditions. Consider a hiring committee consisting of three different people with the following psychological profiles.

First, there is the “earnest, explicit racist”. The explicit racist has implicit racist biases which are in line with her explicit racist biases. “Though she does not know about her implicit biases, if made aware, she would take pride in the fact that these instinctive evaluative tendencies run in tandem with her more reflective judgements, and that both express her considered values.”²⁶⁷ Furthermore, there is the “old-school egalitarian” (circa 1980). The old-school egalitarian is explicitly anti-racist but she also harbors implicit racist attitudes of which she is unaware. However, “[...] not only is she unaware that she herself is implicitly biased, she has never heard of implicit biases at all.”²⁶⁸ Finally, there is the “new egalitarian” (circa 2014). The new egalitarian is genuinely anti-racist but also harbors implicit racist biases of which he is not aware. “Like many others in 2014, however, the new egalitarian is vaguely aware of the phenomenon of implicit bias, but has not looked into the

²⁶⁴ Cf., *ibid.*, 18.

²⁶⁵ Cf., *ibid.*, 19.

²⁶⁶ *Ibid.*

²⁶⁷ *Ibid.*, 19.

²⁶⁸ *Ibid.*, 20.

matter very much, and so does not know any details.”²⁶⁹ If he became aware of the influence of implicit biases on his decisions, he would acknowledge that something had gone wrong.²⁷⁰

This hiring committee has the task of going through a set of résumés and deciding which candidates to invite for an interview. They all put a lot of time and effort into this task. But as we have seen, all three of them are implicitly biased. “As a result, all of those selected to be interviewed turn out to be white; the committee overwhelmingly favored résumés from candidates with “white-sounding” names [...]”²⁷¹ even though all candidates were equally qualified. Therefore, we could say that because their judgement was influenced by implicit bias, something in their decision-making process went wrong. At this point we might ask if every member of the committee is equally responsible and blameworthy for this.

Of course, the earnest explicit racist is responsible for the committee’s choice, and thus deserves blame. Her degree of blameworthiness, however, is the least interesting for our purposes since her implicit and explicit biases would have led to the same results. “A much more interesting matter lies in what we think is an important difference between the old and new egalitarians.”²⁷² Washington and Kelly hold the new egalitarian to be more responsible for the racially biased outcome, than the old-school egalitarian and therefore more blame should be directed to the new egalitarian. Their argument demonstrates how implicit bias can be understood as an exception clause to the exculpating conditions of knowledge and control. Although neither the old nor the new egalitarian knew that they have implicit biases, “[...] the new egalitarian *could* have and *ought* to have known about this, and *could* have and *ought* to have taken appropriate steps to nullify or counteract their influence on the decision process.”²⁷³ Therefore, knowledge about implicit bias is relevant for responsibility. If such knowledge is available in one’s epistemic environment, one becomes blameworthy for not double-checking if one might be influenced by implicit attitudes.

While I agree with this view I think we should add more nuance to it. In my opinion, the old-school egalitarian who did not know anything about implicit biases and therefore could be described as epistemically ignorant out of bad luck should not be let off the hook that

²⁶⁹ Ibid.

²⁷⁰ Cf., *ibid.*

²⁷¹ Ibid.

²⁷² Ibid., 21.

²⁷³ Ibid.

easily. We will see why this is the case when we discuss Fricker's considerations about the blameworthiness for implicit bias. But before we turn to Fricker, let us briefly discuss how Medina construes the relation of social knowledge, self-knowledge and accountability for epistemic ignorance.

4.2.2 Medina on Epistemic Responsibility and Culpable Ignorance

Although Medina does not explicitly draw on implicit biases, I will argue that his notion of ignorance and particularly that of culpable ignorance can be analyzed in a similar way to implicit bias. Referring to George Sher's criticism of the "searchlight view"²⁷⁴, Medina claims that we are not only responsible for the things we actively choose to do, but also for the things we do unconsciously, as well as for the things we should have known, but didn't.²⁷⁵ To illustrate this, Medina presents an example that is structurally similar to the new egalitarian's position on the hiring committee. He reports a real-life incident that occurred at his own academic community, Vanderbilt University.

After a fraternity party in which a pig head had been roasted and eaten, an intoxicated frat boy walked across the street with the pig's head and left it at the doorsteps of the Ben Schulman Center for Jewish Life. [...] The incident happened during the Jewish High Holy Days that begin with Rosh Hashanah and end with Yom Kippur, and many thought that "someone was sending Vanderbilt's Jewish community a chilling message during the holiest days of the year." (The Nashville Scene, October 20, 2005, p. 1)²⁷⁶

After an investigation had been launched, the student who had dropped the head came forward and confessed. He apologized but also excused himself by appealing to his ignorance; he had no knowledge about the significance of pig's parts and their connection to the oppression of Jewish people, nor did he know that the building where he had dropped the head was a Jewish cultural center. He only knew that the building housed a vegetarian cafeteria and thought that leaving the pig's head there would be "[...] a funny joke about vegetarianism."²⁷⁷

First, the boy's excuse that he was making a joke about vegetarianism is in fact not a valid

²⁷⁴ "As George Sher (2009) has argued, the epistemic condition of responsible agency is misunderstood if we think that an agent's responsibility extends only as far as "the searchlight" of his or her consciousness." (Cf., Medina, *The Epistemology of Resistance*, 123.)

²⁷⁵ Cf., *ibid.*, 123.

²⁷⁶ *Ibid.*, 135.

²⁷⁷ *Ibid.*

excuse, but already problematic.²⁷⁸ Second, Medina asserts that the epistemic failure, and thus the blameworthiness, in the pig head-dropping incident not only lie with the individual student, but with the whole epistemic community.²⁷⁹ The student's appeal to ignorance was accepted by the University administration, a fact that is just as problematic. In an official statement, the board declared that although the incident portrayed "bad taste", the student should not be understood as bigot or racist.²⁸⁰ In my view, the alleged exculpating factors, namely the boy's ignorance and lack of knowledge, are not really relevant to the situation. In other words, the exculpating condition of lack of knowledge (and possibly control, since the boy was intoxicated) does not apply, since the kind of ignorance the boy appealed to *should* not exist. As Medina puts it "[...] even if we believe the perpetrator's own account of his epistemic situation, there is still a failure in epistemic responsibility with ethical and political consequences."²⁸¹

I argue that the epistemic failure at stake in this situation is similar to the epistemic failure of the new egalitarian. Both of them *should* have known better. One cannot simply plead ignorance, refer to a lack of knowledge and deny responsibility for one's actions. As stated above, the student tried to exculpate himself by claiming he did not know that the building where he had dropped the pig's head was a Jewish cultural center. This is a sign of his culpable ignorance. For instance, not knowing where the cultural center of a historically oppressed minority is located on campus is already an epistemic, but also an ethical failure, since it unmasks a tendency to be unfamiliar with social contexts on campus other than the boy's own (assumably mostly white, Christian) fraternity. Medina concedes that exactly how much detailed knowledge is required to be a responsible agent in the Vanderbilt community is a tricky question. "But it is clear that not knowing *anything at all* about this history of exclusion and symbolic traces it left behind is unacceptable, and it constitutes a blameworthy lack of epistemic responsibility [...]."²⁸²

²⁷⁸ Vegetarianism is a choice many people make in order to live and eat more earth-conscious. Leaving a dead animal's head at the door of a vegetarian cafeteria thus constitutes a way of disrespecting and ridiculing this choice.

²⁷⁹ Cf., *ibid.*, 137.

²⁸⁰ Cf., *ibid.*

Medina asserts that bodies of ignorance maintain themselves through the 'division of cognitive laziness.'

²⁸¹ *Ibid.*, 136.

²⁸² *Ibid.*, 139.

In Medina's view two cognitive minima are violated in the incident portrayed above that are necessary for being a responsible agent in a social-environment, namely minimum self-knowledge and minimum empirical knowledge of the social world.²⁸³ What he means by this is that "[...] responsible agency requires that one be minimally knowledgeable about one's mind and one's life, about the social world and the particular others with whom one interacts, and about the empirical realities one encounters."²⁸⁴ I argue that this is exactly what made the new egalitarian more blameworthy than the old egalitarian, namely that the former had more empirical knowledge and thus should have known about the possibility of implicit bias clouding his judgement. Furthermore, the requirement of these minima repudiates the pig-head dropper's excuse. Knowledge about the Jewish minority was available on campus, yet the perpetrator in the pig-head dropping incident never tried to familiarize himself with it.

This leads me to the final important point in Medina's deliberations about epistemic responsibilities.

What is reasonable to expect of responsible agents to know about themselves, about their peers, and about their surroundings needs to be socially contextualized: the agent's epistemic obligations and our entitlement to expect and demand particular kinds of knowledge are always contextually bound.²⁸⁵

In my opinion this is an extremely important point. In situations of epistemic injustice and oppression, not everybody can equally partake in the creation and reception of knowledge. Furthermore, as we have seen throughout the thesis, implicit-bias research shows us how likely it is that a communicative and epistemic imbalance (as a result of implicit prejudice) is more prevalent than we might think. Therefore, we should not expect everybody to be adequately knowledgeable about themselves, the world and others because under conditions of oppression not everybody can equally access and produce this kind of knowledge. Medina therefore advocates for a comparative status of epistemic responsibility that takes into account to whom such knowledge is available: "My view of epistemic responsibility is one that [...] rejects any one-size-fits-all approach and one that argues that assignment and assessment of responsibilities should be done piecemeal, case by case [...]."²⁸⁶ For Medina, the level of responsibility for one's bias and ignorance should thus be attributed by taking

²⁸³ Cf., *ibid.*, 127.

²⁸⁴ *Ibid.*

²⁸⁵ *Ibid.*, 130.

²⁸⁶ *Ibid.*, 119.

into consideration the informational environment available to the individual.

4.2.3 Fricker's Concept of "Epistemic Agent-Regret"

One question still remains unanswered when we cast our thoughts back to the old-school egalitarian on the hiring committee who is influenced by implicit bias but has never before heard of this branch of psychological research. Does she really bear no responsibility at all for the influence of her implicit biases? As we have seen, Washington and Kelly relieve her from responsibility and blameworthiness since she does not meet the knowledge condition.²⁸⁷ While she contributes to a morally problematic outcome, (namely inviting only white people to be interviewed for a position) she does not know that she is implicitly biased, nor does she even know what implicit bias is. One could therefore say that although she was responsible for the outcome of the hiring committee's decision, this responsibility was not culpable.

In reference to Williams, Fricker introduces a concept, "epistemic agent-regret", that is exactly in line with such a scenario. By introducing epistemic agent-regret, Fricker comes up with a way of treating cases of no-fault epistemic responsibility for certain kinds of bad judgements.²⁸⁸ Her conception is specifically designed for cases where the agent innocently inherits bad epistemic goods from her environment but does not suspect any toxic influence from these goods.²⁸⁹ What can we say about such a case of epistemic innocence and epistemic bad luck?

First, exculpatory excuses, such as innocent inheritance of bad epistemic goods may release us from blame, but they do not change the general epistemic fault committed in instances where our judgements were influenced by these goods. However, this fault does

²⁸⁷ Cf., Washington, Kelly, "Who's Responsible for This? Moral Responsibility, Externalism, and Knowledge about Implicit Bias", 23.

²⁸⁸ Cf., Fricker, "Fault and No-Fault Responsibility for Implicit Prejudice", 35.

This is very interesting, since elaborating on implicit prejudice is the main task of this thesis. This, however, is not Fricker's emphasis in her 2007 book. And yet, since 2007 a lot has been achieved in the field of implicit-bias research. Even though Fricker does not directly apply her considerations about implicit bias to her theory of testimonial injustice in her 2016 paper, the thesis nonetheless suggests that complementing Fricker's theory of residual-identity prejudice with findings of implicit-bias research might be an undertaking with which Fricker actually would agree.

²⁸⁹ Cf., *ibid.*, 40.

not lie within the subject but rather, as Fricker puts it, “[...] flows through her.”²⁹⁰ To illustrate this, think back to the old-school egalitarian who passively and innocently inherits the racist biases of her environment but suspects no toxic influence of any kind. This is a case of epistemic innocent error, as the old school egalitarian is a blameless conduit of prejudice.²⁹¹ In other words, “[...] her judgements are epistemically bad, but it is not her fault.”²⁹² Fricker suggests that the fault, namely the motivated maladjustment to evidence that leads to the persistence of prejudice, has been committed “off-stage” by an epistemic collective of which the old-school egalitarian in our example is a member. But what can we say about the moral status of our old-school egalitarian “[...] other than that she made bad judgements through no fault of her own and so cannot, should not, be held accountable?”²⁹³

To answer this question, Fricker draws an analogy to Williams’ idea of “epistemic agent-regret”. Williams argues that when we do bad things through no fault of our own, but through epistemic or moral bad luck, the right response is to “[...] morally own these aspects of our conduct.”²⁹⁴ This notion of “owning” expresses itself through a first-personal reflexive form of regret, or, as Williams puts it, agent-regret. Agent-regret, Fricker explains, “[...] is properly considered a feeling of guilt for harm done, though clearly not of a kind entailing culpability [...]”²⁹⁵ Agent-regret is a term coined to fill an empty conceptual space in our morality system, namely for the kind of guilt that arises in instances where someone perpetrated non-culpable harm.²⁹⁶

Williams’ cases for agent-regret mostly are exemplified through situations of “outcome” luck. As a result of sheer bad luck, a lorry driver could, for instance, cause a tragic accident for which he cannot be blamed but nevertheless, he feels responsible for it. And yet epistemological concerns about responsibilities relating to implicit biases do not arise through unlucky outcomes, but rather, as Fricker puts it, unlucky inputs.²⁹⁷ To cite an example by Williams that Fricker discusses, think of the horrifying things Oedipus has non-

²⁹⁰ Ibid., 42.

²⁹¹ Cf., *ibid.*

²⁹² Ibid.

²⁹³ Ibid., 43.

²⁹⁴ Ibid., 44.

²⁹⁵ Ibid., 45.

²⁹⁶ Cf., *ibid.*

²⁹⁷ Ibid.

culpably done by killing his father and marrying his mother. “Circumstances conspired to ensure that he could not have been expected to know these things, and so this non-culpable factual ignorance entailed non-culpable moral epistemic ignorance of the significance of his deeds.”²⁹⁸ Oedipus committed these crimes without fault and when he discovered what he really had done, he blinded himself because he felt so much shame and remorse.

Similarly to Oedipus, the old-school egalitarian voluntarily conducted her actions, but could not grasp their significance. She did not know that she was biased, and neither did she know anything about implicit bias. In some way, Oedipus and the old-school egalitarian suffer from a kind of “[...] *environmental* bad luck.”²⁹⁹

In introducing the epistemic counterpart of agent-regret, Fricker thus provides a way to analyze cases in which implicit biases blamelessly influence a subject’s judgement. In these instances, the subject can be regarded as non-culpable, yet responsible. Furthermore, Fricker asserts that this non-culpable responsibility allows for epistemic obligations that the subject should fulfill by “owning” up to what she has done and consequently trying to take corrective measures.³⁰⁰ As far as the old-school egalitarian is epistemically virtuous, for instance, she will take steps to raise awareness about the possibility of implicit-bias influence on others and herself. Ultimately, the goal of the old-school egalitarian’s feeling of epistemic agent-regret should be to minimize her prejudices.

This aspect of epistemic agent-regret introduces the final project of this thesis. How can we reduce our prejudices, and more specifically our implicit prejudices, in order to prevent instances of testimonial injustice?

4.3 Remedies against Testimonial Injustice and Implicit Prejudice

Negative consequences that arise in situations of testimonial injustice are severe for the affected individual as well as for the epistemic community. Not only will the individual whose testimony receives no uptake be downgraded in her very humanity, since she is not taken seriously as a knower, but it will also harm the epistemic community as a whole. The knowledge the individual wants to convey is not integrated into the generation of meaning

²⁹⁸ Ibid., 46.

²⁹⁹ Ibid.

³⁰⁰ Cf., *ibid.*, 47.

and discourse, meaning that the community is deprived of possibly valuable information.

Moreover, as has already been pointed out, since implicit prejudices are pervasive, instances of testimonial injustice are pervasive as well. Consequently, even the most devoted egalitarian is not immune to implicit biases influencing her credibility judgements. This is highly unsettling and should not only be seen as a cause to doubt our beliefs and their formation, but should also compel us to act.

In the following, I will introduce accounts that propose remedies against testimonial injustice. Evidently, Fricker's model of the virtue of testimonial justice is the first we ought to consult when trying to come up with antidotes against testimonial injustice. I will argue that reflective self-regulation and the conscious endorsement of epistemic virtues is crucial for an amelioration of our epistemic situation. However, this should only be understood as the first move to remedy a phenomenon that is rendered more complex when we consider how implicit bias influences our judgements. As Alcoff rightly puts it: how should volitional deliberation correct for involuntary, automatic influence of prejudice?³⁰¹

I will argue that we need to broaden Fricker's model of testimonial justice by putting more emphasis on the role of habituation. Moreover, the thesis claims that a successful conduct of testimonial justice cannot be achieved merely at the individual level. Rather, egalitarian testimonial justice can only flourish in a community that is committed to implementing such a virtue.

4.3.1 Fricker's Model of the Virtue of "Testimonial Justice"

Credibility deficit in Fricker's account arises through identity prejudices that distort the hearer's perception of the speaker. For her, "[...] this raises the question whether we can identify a particular virtue that the hearer needs to have to counteract the risk of letting such prejudice distort his perceptions of speakers."³⁰² Fricker's goal is not simply to uncover the discriminatory force of testimonial injustice, but she also provides an account of epistemic virtues that aims to counteract such an injustice.

Fricker proposes that the virtuous hearer should be equipped with a certain testimonial

³⁰¹ Cf., Alcoff, "Epistemic Identities", 128.

³⁰² Fricker, *Epistemic Injustice*, 86.

sensibility. This testimonial sensibility is accomplished by a “[...] distinctly *reflexive* critical awareness.”³⁰³ In other words, the hearer should exercise an intellectual shift of gear from a spontaneous, unreflective mode into a critical reflection in order to clarify to what extent identity prejudice has influenced her credibility judgement. By conducting this intellectual shift of gear, the virtuous hearer can then neutralize the influence of prejudice on her judgements.³⁰⁴ Fricker labels this the virtue of “testimonial justice”.

Testimonial justice is not only a product of critical reflection, but it also comes about more spontaneously, namely through a process of habituation.³⁰⁵ Fricker asserts that familiarity and habituation are necessary to gradually “[...] melt away the prejudice that presented an initial obstacle to an unprejudiced credibility judgement being made.”³⁰⁶ Thus, if critical reflection functions as the initial motivation to shift gear and correct for one’s credibility judgements, habituation and familiarity are necessary modes of rehearsing and exercising the virtue of testimonial justice. I think that this is a very important point, but I also suggest that Fricker could be more precise about the role that habituation plays in the acquisition of virtue. This aspect is related to criticism that I have already voiced earlier on in the thesis: critical reflection alone cannot be sufficient to act in a testimonially just way.³⁰⁷ But before we start critically discussing Fricker’s virtue of testimonial justice, let us further investigate how it develops.

Another important aspect of testimonial justice we must understand is that it is historically contingent. Fricker illustrates this thought by drawing on her analysis of *The Talented Mr. Ripley*: Herbert Greenleaf’s inability to perceive Marge Sherwood as the important source of knowledge that she actually is must be seen in a historical context. Greenleaf lacks a critical awareness of how the gender prejudice of his time distorts his judgement. Similarly to her deliberations concerning epistemic agent-regret discussed above, Fricker construes Greenleaf’s epistemic fault as an inculpable one, at least until the “[...] requisite critical

³⁰³ Ibid., 91.

³⁰⁴ Cf., *ibid.*, 91-92.

³⁰⁵ Cf., *ibid.*, 96.

We will say more about the important role of habituation in the acquisition of virtue when we briefly discuss Zagzebski’s account of virtue epistemology.

³⁰⁶ Ibid., 96.

³⁰⁷ We will return to this criticism when we discuss Alcott’s discussion of testimonial justice.

Furthermore, I will argue that an emphasis on habituation is central when it comes to correcting for implicit bias.

consciousness of gender became available [...]” to him.³⁰⁸ Thus, Greenleaf cannot be blamed for not exercising testimonial justice for the historic circumstances of his time do not supply sufficient awareness about gender prejudice.

To escape the threat of historical contingency, Fricker sketches an ahistorical setting by drawing on Williams and his use of a fictional State of Nature scenario. Williams constructs such a scenario to characterize our most basic epistemic needs with the chief purpose of arriving at the virtue of “truthfulness”.³⁰⁹ Truthfulness, according to Williams, consists of the intellectual virtues of “accuracy” and “sincerity”.³¹⁰ Speakers thus must be accurate and sincere in order to pool knowledge and structure their epistemic relations. However, the pooling of knowledge does not only make demands on speakers, but equally concerns hearers. For Fricker, testimonial justice can satisfy this demand on hearers. “Accuracy and Sincerity sustain trust as regards *contributing* knowledge to the pool; Testimonial Justice helps sustain trust as regards *acquiring* knowledge from the pool.”³¹¹ Because hearers in the State of Nature also rely on stereotypes as heuristic shortcuts when assessing testimony, they consequently also need a disposition to avoid the influence of prejudice over their judgements of credibility.³¹²

Placing the virtue of testimonial justice in an ahistorical setting, such as the State of Nature reveals its fundamental function as an intellectual virtue for hearers. Nevertheless, Fricker does not regard testimonial justice a merely intellectual virtue, but as a hybrid that oscillates between intellectual and moral virtue.³¹³ Drawing on Linda Zagzebski’s thoughts on virtue theory and epistemology, Fricker argues that testimonial justice’s motivational component and simultaneously its immediate end is both ‘truth’ and ‘justice’.³¹⁴ Exercising the virtue of testimonial justice through neutralizing one’s prejudices will thus lead to truthful and righteous assessment of testimony.

³⁰⁸ Ibid., 100.

³⁰⁹ Cf., *ibid.*, 108.

The State of Nature is a minimal human society of minimal organization. People in the state of nature live in groups and thus share the same basic epistemic needs. Williams describes three collective epistemic needs, namely the need to possess enough truths, the need of pooling and sharing information and the need to stabilize relations of trust among individuals. Cf., *ibid.*

³¹⁰ Cf., *ibid.*, 111.

³¹¹ Ibid., 116.

³¹² Cf., *ibid.*, 115.

³¹³ Cf., *ibid.*, 120.

³¹⁴ Cf., *ibid.*, 122.

4.3.1.1 Alcoff's Criticism

Still, there are some questions and objections concerning Fricker's model of testimonial justice that are worth discussing. The primary worry I want to put forward is in line with Alcoff's criticism of Fricker and stems from the fact that identity prejudices likely operate on an unconscious, implicit level. The question arises if critical reflection really is apt to remedy these implicit kinds of prejudices. As Alcoff puts it, "[...] can a successful antidote operate entirely as a conscious practice? In other words, will volitional reflexivity be sufficient to counteract a non-volitional prejudice?"³¹⁵

Fricker acknowledges this criticism. In her 2010 paper "Replies to Alcoff, Goldberg and Hookway on *Epistemic Injustice*"³¹⁶ she defends her concept of reflective self-regulation, but she also concedes that more is necessary to effectively counteract identity prejudices, especially when they are implicit.

Referring to a psychological study by Margo Monteith, Fricker argues that it is possible to reflectively monitor one's judgements for the influence of prejudice.³¹⁷ This monitoring will, for instance, reveal affective self-directed emotions, such as guilt, "[...] associated with the awareness of having fallen below a certain subjectively approved standard."³¹⁸ To cite an illustration of this idea, Fricker refers to a white shopper, who asks a person of color for assistance in a store, even though the black person is also a customer herself. The white shopper consequently experiences a feeling of guilt about his wrongful stereotypical assumption and tries to be more sensitive in the future.³¹⁹

Though I do not want to categorically rule out the possibility of reflective monitoring as a control mechanism against implicit prejudice, I think this conception of Fricker's theory lacks an important insight. Knowledge about one's own implicit prejudices does not automatically imply that one is able to control them immediately. In my opinion, habituation must play an equally important role. Fricker, however, seems to agree. In her 2010 paper she clarifies that her conception of virtue does not necessarily operate on a purely volitional,

³¹⁵ Alcoff, "Epistemic Identities", 132.

³¹⁶ Fricker, "Replies to Alcoff, Goldberg and Hookway on *Epistemic Injustice*", in: *Episteme*, Vol. 7, Issue 2, (2010), 164-178.

³¹⁷ Cf., *ibid.*, 165.

³¹⁸ *Ibid.*

³¹⁹ Cf., *ibid.*

reflective level. Rather, “[...] motive-based virtue incorporates a reliability condition [...] and there is no reason why this reliability should not be achieved by way of sheer habit or other sub-personal mechanism.”³²⁰ Nevertheless, I am convinced that we need to be more precise when it comes to the relation between virtue, habituation and implicit prejudice. In the next section I will therefore develop a perspective that aims at integrating these three aspects.

Before we can attend to the important role of habituation in virtue theory, I shall now briefly introduce another line of criticism that concerns Fricker’s model of the virtue of testimonial justice. The challenge that many virtue accounts are confronted with is that they are merely agent-based, which means that they only emphasize the individual’s role in counteracting injustice and seem to neglect the broader, social background that equally promotes injustices. Both Alcoff and Elizabeth Anderson voice such criticism.³²¹

I adopt this line of objection and, more specifically, raise two related points. First, one could object that testimonial justice only provides a remedy for hearers and assigns a rather passive role to the speakers, who are actually the primary victims of instances of testimonial injustice.³²² Furthermore, it is uncertain if testimonial justice really is capable of counteracting structural epistemic injustices that may have “[...] locally innocent (non-prejudicial) causes, but require structural remedies.” Elizabeth Anderson raises this latter point in her paper “Epistemic Justice as a Virtue of Social Institutions”. In other words, one is more likely to become an epistemically virtuous agent, if the social environment one finds oneself in is a hospitable one that equally embraces epistemic and egalitarian virtues.

4.3.1.2 Virtues, Habituation, and Implicit Bias

To address Alcoff’s worry that volitional epistemic practices, such as the virtue of testimonial justice, might not be enough to correct for implicit prejudice, let us investigate the relationship between virtue, habit and implicit bias in more detail. By doing so, this thesis

³²⁰ Ibid., 166.

³²¹ Cf., Alcoff, “Epistemic Identities”, 135 & Cf., Elizabeth Anderson, “Epistemic Justice as a Virtue of Social Institutions”, in: *Social Epistemology*, 26:2 (2012), 163-173.

³²² Cf., Kusch, Review: “Miranda Fricker: Epistemic Injustice: Power and the Ethics of Knowing”. 173.

argues that the virtue of testimonial justice can be improved by underlining the importance of habituation in the process of virtue-acquisition.

To clarify how exactly virtue is connected to habit let us begin by consulting Zagzebski's book *Virtues of Mind*, which provides a thorough analysis of the relationship between epistemology and virtue theory. Zagzebski clarifies that a virtue is an acquired excellence that takes time to develop. Virtues thus have habitual character, but this does not mean that a virtue *is* a habit. "The features of gradual acquisition and entrenchment suggest that a virtue is a kind of habit [...]. But it would be too hasty to conclude that a virtue is *identical* with a habit [...]." ³²³ Whereas habits are only one sort of 'second nature', virtues do not operate on a purely automatic level but require intelligence in their operation. And yet both habits and virtues "[...] are acquired through a process of repetition over time [...]," ³²⁴ or, to put it differently, they need to be trained, as Aristotle maintained. ³²⁵

To further illustrate this, Zagzebski distinguishes the virtuous person from the morally strong person. Whereas the latter simply is capable of overcoming temptation in cases, "[...] in which she knows the right thing to do [...], the former has [...] acquired a habit of *feeling* as well as acting, a habit of being motivated in a certain way [...]." ³²⁶ According to Zagzebski, this means that the virtuous person has a superior form of moral knowledge because she is motivated to do the right thing without having to resist a temptation to do otherwise. ³²⁷ This thought is not new to us since we have already encountered it in Fricker's model of virtuous perception. The capacity of moral perceptual judgement is so entrenched in the virtuous subject, that she "[...] sees the world in moral color." ³²⁸ When the agent is confronted with a situation of a certain moral character, she does not have to put thought into figuring out if the situation is cruel or kind or charitable. She just *sees* it that way. ³²⁹

Furthermore, it is psychologically impossible for humans to obtain virtue without experience and interaction with the world, which means that reflection alone will not initiate

³²³ Linda Trinkaus Zagzebski, *Virtues of Mind - An Inquiry into the Nature of Virtue and the Ethical Foundation of Knowledge*. (Cambridge: Cambridge University Press, 1996), 116.

³²⁴ Ibid., 117.

³²⁵ Cf., *ibid.*

³²⁶ Ibid., 119.

³²⁷ Cf., *ibid.*

³²⁸ Fricker, *Epistemic Injustice*, 72.

³²⁹ Cf., *ibid.*

the generation of a virtue. As Zagzebski puts it, “[...] a single act of will does not a virtue make.”³³⁰ In other words, a single act of will is not enough to transform itself directly into a virtue.³³¹ This is exactly what this thesis expressed above by arguing that pure awareness of one’s residual, implicit prejudices will not be sufficient to develop the virtue of testimonial justice and consequently correct for prejudices. Rather, the virtue needs to be internalized; it has to reach a level of automatic activation that can only be achieved by a process of habituation.

To summarize, in the standard case it is in the nature of virtue that “[...] it be an entrenched quality that is the result of moral work on the part of the human agent, and that it be acquired by a process of habituation.”³³² Thus the development of a virtue relies on a process of habituation, in which the virtue is trained and automatized. However, we still need to clarify how the habitual character of virtue acquisition relates to implicit prejudice. Let us therefore briefly shed light on an account that links virtue ethics with implicit bias and habituation.

In her paper “A Virtue Ethics Response to Implicit Bias” Clea F. Rees also highlights the importance of automatization and habituation when it comes to an enduring development of virtue. She furthermore identifies a challenge that virtuous automaticity faces, namely the influence of implicit bias. Rees does so by pointing out the possibility that our virtuous automaticity is undermined “[...] by our unwitting habituation of the wrong motivations [...]”³³³ Successful acquisition of virtue therefore require not only the habituation of virtuous motivation, but the dehabituation of vicious ones too. Subsequently Rees argues that the threat posed by the automatic influence of implicit bias can be countered with its own weapon – namely the automatization of strong egalitarian goals that initiate the development of virtue.³³⁴

Rees’ proposition is similar to Fricker’s idea of reflective self-regulation. As far as Rees’ distinctively focuses on automaticity, she presents a more thorough response to how virtue

³³⁰ Zagzebski, *Virtues of Mind*, 121.

³³¹ Cf., *ibid.*

³³² *Ibid.*, 125.

³³³ Clea F. Rees, “A Virtue Ethics Response to Implicit Bias”, in: *Implicit Bias and Philosophy. Volume 2. Moral Responsibility, Structural Injustice and Ethics*, ed. Saul, Brownstein, (Oxford: Oxford University Press, 2016), 191-214, 191.

³³⁴ Cf., *ibid.*, 192.

ethics could counsel implicitly-biased people. It is important to understand that her approach does not simply suggest “willing away” implicit bias. Given the distinctly implicit structure of bias, this would be pointless. In contrast, to adopt a goal and commit to egalitarian ideas is different from simply trying to no longer be implicitly biased.³³⁵ What Rees proposes is that we have to deliberately choose egalitarian virtues and habituate these commitments through constant repetition.

To underline her proposition with empirical evidence, she cites two research programs that are concerned with the effectiveness of choosing egalitarian commitments.³³⁶ “This research supports two claims: first, consciously chosen egalitarian commitments can be automatized; second, habituated egalitarian motivations can effectively guide automatic cognition.”³³⁷ Moreover, when choosing egalitarian goals to counteract prejudice, the right kind of motivation is essential. This is a finding the first research program illustrates. A study by Devine and Plant, namely the “internal and external motivation to respond without prejudice”, demonstrates that individuals who are personally committed to not being prejudiced because they believe in egalitarian values are able to avoid implicit prejudice better than individuals whose motivation simply is not to appear prejudiced. The former is thus internally motivated (IMS), whereas the latter’s motivation is external and “[...] stems from a concern with self-presentation.”³³⁸ In short, the study found that high IMS low EMS individuals show the least bias.³³⁹ This raises the question why high IMS high EMS individuals were not equally successful in avoiding bias in their judgements. Rees cites Devine and colleagues as follows:

High IMS low EMS individuals have highly accessible, automatized egalitarian commitments which conflict with implicit stereotypes at an early enough stage of cognitive processing for the conflict-monitoring mechanism to be effective in signaling the need for increased response regulation automatically and non-consciously.³⁴⁰

In contrast, high IMS high EMS individuals can access their egalitarian commitments less readily, which consequently renders their activation less automatized. Thus, these

³³⁵ Cf., *ibid.*, 201.

³³⁶ Cf., *ibid.*, 202.

³³⁷ *Ibid.*

³³⁸ *Ibid.*, 203.

³³⁹ Cf., *ibid.*

³⁴⁰ *Ibid.*, 204.

individuals can control their implicit prejudices only through conscious cognitive processing, and not, as the former group, automatically and unconsciously.³⁴¹ The second research program Rees presents supports these findings by showing that “[...] those highly accessible, enduring egalitarian goals can inhibit the activation of stereotypes preconsciously [...]”³⁴² by an implicit motivation to control prejudice (IMCP).

These two research programs therefore are “[...] good news for the virtue ethicist.”³⁴³ To summarize, they put forward that it is possible to avoid implicit prejudice by habituating initially endorsed egalitarian motivations. The habituated virtuous motivations can then guide cognition outside conscious awareness.³⁴⁴ On this account one can agree with Fricker’s assertion that testimonial justice needs to be activated by self-reflective awareness. And yet, in order for it to be a successful antidote to implicit prejudice, testimonial justice must become a habit. Mere awareness will not be sufficient.

4.4 Ameliorated Testimonial Justice and Epistemic Friction

In conclusion, let me look beyond individual commitment of hearers and ask what speakers, who are the actual victims of instances of testimonial injustice, can do to improve their epistemic situations. Moreover, by shedding light on the fact that injustices usually operate on a structural level, I will address Anderson’s claim that testimonial justice has to become an institutionalized virtue.

To introduce these lines of arguments, it seems helpful to take a closer look at another concept of epistemic resistance, namely Medina’s concept of “epistemic friction”. When Medina talks about epistemic resistances³⁴⁵ in our cognitive lives, he clarifies that they can be both internal or external and both positive or negative. Let me first elaborate on internal,

³⁴¹ Cf., *ibid.*

One could compare this to the way Zagzebski distinguished between the truly virtuous person and the morally strong person. The morally strong person is a person who might wish to control her prejudices because of external motivations. In order not to appear prejudiced, she thus overcomes the temptation to act prejudiced because she knows that it is the right, moral thing to do. The virtuous person, however, has an intrinsic motivation not to act prejudiced.

³⁴² *Ibid.*

³⁴³ *Ibid.*, 205.

³⁴⁴ Cf., *ibid.*

³⁴⁵ Epistemic resistance for Medina means to be confronted with and consequently withstand different opposing epistemic forces. (Cf., Medina, *The Epistemology of Resistance*, 49.)

positive resistances. “For example, an opposing internal force that counteracts external epistemic influences can be positive insofar as it is critical, unmasks prejudices and biases, reacts to bodies of ignorance; and so on [...]”³⁴⁶ This comes as no surprise to us. Fricker’s model of the virtue of testimonial justice aims to do exactly the same. When someone is confronted with an internal force of epistemic resistance, such as critical monitoring for prejudice, for instance, this can lead to “beneficial epistemic friction”. The individual is encouraged to compare and contrast her beliefs with possible objections and recognize cognitive gaps in her reasoning. However, according to Medina, internal and external epistemic resistance can also be negative in the sense that it involves a certain reluctance to learn and believe. In short, this latter variation of resistance describes, “[...] the kind of stubbornness that gets in the way of knowledge,”³⁴⁷ and leads to “detrimental epistemic friction”.

Furthermore, beneficial epistemic friction depends on two guiding principles, namely the “principle of acknowledgement and engagement” and the “principle of epistemic equilibrium”:

The former dictates that all cognitive forces we encounter must be acknowledged, and insofar as it is possible, they must be in some way engaged (even if in some cases only a negative mode of engagement is possible or epistemically beneficial). The latter principle lays out the desideratum of searching for equilibrium in the interplay of cognitive forces, without some forces overpowering others, without some cognitive influences becoming unchecked and unbalanced.³⁴⁸

Beneficial epistemic friction is a concept meant to counteract epistemic vices, such as ignorance, epistemic laziness and closed-mindedness.³⁴⁹ The basic idea is that by taking divergent perspectives and viewpoints into account, one is more likely to critically assess one’s own beliefs for the influence of prejudice or other distorting factors. In other words, Medina argues that the engagement with epistemic resistances facilitates the development of virtues, whereas the absence of epistemic resistance and friction leads to epistemic vices, bad epistemic habits and epistemic disequilibria.³⁵⁰

One epistemic vice Medina highlights is “meta-blindness”, which he defines as a twofold

³⁴⁶ Ibid., 49-50.

³⁴⁷ Ibid.

³⁴⁸ Ibid., 50.

³⁴⁹ Cf., *ibid.*, 51.

³⁵⁰ Cf., *ibid.*

negative resistance to epistemic friction: on the one hand, the individual is blind to epistemic counterpoints but she is also blind to her own blindness, which means that she resists recognizing her lack of openness.³⁵¹ Following Medina, undoing meta-blindness involves a cognitive transformation process, but more crucially, it involves the “[...] restructuring of habits and affective structures.”³⁵² In a first step, the meta-blind subject, like the implicitly-biased subject thus needs to become aware of her lack of epistemic engagement, and furthermore needs to restructure her habit of perceiving the world. This can be facilitated by epistemic friction. The subject needs to actively search for alternative viewpoints and remain open to epistemic counterpoints.

I shall argue that Fricker’s account of testimonial justice could benefit from Medina’s concept of epistemic friction since his proposal of the engagement between different epistemic agents with divergent perspectives will help to set in action a process of critical awareness and self-regulation. However, as claimed above, testimonial justice remains a virtue that is principally designed for hearers, the perpetrators of testimonial injustice. Therefore, we need to address what speakers themselves can do in order to promote epistemic resistance and ensure that their testimony receives uptake.

4.4.1 “Meta-Lucidity” as a Virtue of the Oppressed

Related to Medina’s concept of epistemic friction are his deliberations about specific virtues that are more likely to be found among oppressed subjects. People living under conditions of oppression are “[...] exposed to epistemic practices and processes that erode their epistemic character.”³⁵³ Casting our thoughts back to the way Fricker describes the harms arising in situations of testimonial injustice, this becomes evident. For instance, people who are constantly denied the status of informed, qualified knowers will lose confidence in their intellectual abilities and eventually withdraw from the communicative situation. Nevertheless, some accounts in critical race theory and feminist theory argue that there are certain epistemic advantages and distinctive virtues that are characteristic of oppressed subjects.

³⁵¹ Cf., *ibid.*, 75.

³⁵² *Ibid.*, 76.

³⁵³ *Ibid.*, 42.

Drawing on certain aspects of these accounts, Medina formulates an argument that distinctively underlines the possible epistemic agency of oppressed communities and demonstrates how it is possible to resist predominant epistemologies.³⁵⁴ By doing so, he highlights three epistemic virtues, namely “humility”, “curiosity” and “open-mindedness”, which are developed when members of oppressed groups engage in positive epistemic resistance.³⁵⁵ These three virtues have a converging point which Medina describes as a special kind of “lucidity”, namely “subversive lucidity”.³⁵⁶

Epistemically humble, curious/diligent, and open-minded subjects are likely to detect and overcome blind spots and to develop new forms of lucidity that can enrich social cognition. If we add to that kind of virtuous character the experience of oppression, of not fitting, of having an alternative viewpoint, then the lucidity of the virtuous subject can have a subversive character, having the potential to question widely held assumptions and prejudices, to see things afresh and redirect our perceptual habits, to find a way out or an alternative to epistemic blind alleys, and so on.³⁵⁷

It is important to keep in mind that we cannot generalize and assume that the epistemic perspective of the oppressed is homogenous. It is not possible to attribute a specific virtue to a subject just by referring to their social location. Rather, Medina aims at showing that distinctive epistemic advantages can be found among some oppressed subjects, when they engage in positive epistemic resistance and have the necessary hermeneutical resources to decode instances of injustice. When they make demystifying social experiences, such as decoding patterns of oppression, they can develop a special kind of lucidity.³⁵⁸

Lucidity originates from the invisible position the oppressed subject occupies in her social environment. Due to this lack of visibility, positive and negative epistemic consequences arise. On the one hand, one’s cognitive abilities may suffer since one is prevented from actively participating in epistemic interactions.³⁵⁹ But on the other hand, “[...] one can also comfortably and strategically occupy one’s invisibility, exploiting the benefits of being unperceived while having access to bodies of evidence one is not assumed to know [...]”.³⁶⁰

³⁵⁴ Cf., Nancy Arden McHugh, “Epistemic Communities and Institutions”, in: *The Routledge Handbook of Epistemic Injustice*, ed. Kidd, Ian James, (New York: Routledge, 2017), 270-279, 275.

³⁵⁵ Cf., Medina, *The Epistemology of Resistance*, 42.

³⁵⁶ Cf., *ibid.*, 44.

³⁵⁷ *Ibid.*, 45.

³⁵⁸ Cf., *ibid.*

³⁵⁹ Cf., *ibid.*, 190.

³⁶⁰ *Ibid.*

Feminist standpoint theorists such as Patricia Hill Collins or Alison Wylie have analyzed the phenomenon of cognitive advantages of marginalized subjects in more detail. For instance, in her book *Black Feminist Thought*, Collins examines the standpoint of black women and characterizes their perspectives as “outsiders within”. Black women are “[...] fully cognizant of the social order in which they live but nonetheless are perceived and treated as strangers to that order.”³⁶¹ They can thus gain insight and knowledge that other, more privileged subjects could not have gained. Their lucidity allows them to decode oppressive patterns more easily since they themselves are affected by these patterns.

Based on this line of argument, Medina describes how oppressed subjects occupy blind spots in the privileged person’s field of vision. In other words, these subjects recognize epistemic gaps “[...] that only those who fall into them are aware of [...]”³⁶² Consequently, to realize one’s own invisibility is to realize that there is “[...] more to be seen than what others (some others) see.”³⁶³ For Medina, this realization constitutes a meta-attitude of being “[...] always *on the lookout for more, forever more*, which is based on the *experience that there can be more than what is seen*.”³⁶⁴ He labels this meta-attitude “meta-lucidity”. Meta-lucidity is achieved through the epistemic friction of two conflicting perspectives: the subject’s own gaze and the “[...] social gaze that does not see him.”³⁶⁵ Drawing on philosophical literature on race, one could call the position of meta-lucidity a position of “double consciousness”³⁶⁶, since the subject has the capacity to entertain two perspectives at the same time.

Collins uses the term “double consciousness” to develop an epistemology of resistance for black women. This corresponds to Medina’s conception of “meta-lucidity”. She asserts that due to double consciousness black women can create self-representations that resist the distorting and demeaning racist and sexist prejudices of black femininity.³⁶⁷ “Indeed, double consciousness brings with it the opportunity to develop the ability to shift back and forth

³⁶¹ Cf., *ibid.*, 191.

³⁶² *Ibid.*

³⁶³ *Ibid.*, 192.

³⁶⁴ *Ibid.*

(Original quote in italics as well.)

³⁶⁵ *Ibid.*

³⁶⁶ Originally, the term “double consciousness” was coined by W.E.B. Du Bois.

³⁶⁷ Cf., *ibid.*, 194.

between two ways of seeing and, hence, the ability to make comparisons [...].”³⁶⁸

For Medina, epistemically virtuous double consciousness and meta-lucidity is achieved through internal epistemic friction since successful meta-lucidity not only means that one is able to see the two perspectives, but that one has the ability to connect and critically assess them. In other words, meta-lucid subjects have the potential to decode and see through the veil of ignorance produced by predominant gazes and epistemologies. Additionally, meta-lucid subjects are not only capable of seeing through the veil of ignorance but, as Medina puts it, seeing the veil itself, which means becoming aware of epistemic limitations and obstacles.³⁶⁹

In recognizing meta-lucidity as a virtue that is distinctively characteristic for oppressed subjects, one is able to remedy a neglect that can be traced to Fricker’s account. Fricker’s virtue of testimonial justice does not consider the potential of marginalized subjects who are the actual victims of epistemic injustices. In other words, Fricker does not sufficiently address what oppressed subjects can do to improve their epistemic situation.

Another line of argument highlights the fact that resistant epistemic resources are generated through a shared experience of oppression that goes beyond individual virtuous commitments. In my last section I will therefore address the question whether epistemic injustices have to be countered by not only individual, but also by structural, and institutional virtuous commitments.

4.4.2 Structural Remedies against Testimonial Injustice

Recent accounts in social epistemology and political philosophy have underlined the importance of theorizing epistemic justice as a virtue of epistemic communities. By doing so, they have pointed out how marginalized subjects, owing to the collective experience of oppression, can develop common epistemic virtues, such as lucidity. Additionally, these accounts raise the question as to how testimonial justice could become a virtue of collectives and social institutions.³⁷⁰ In the following I will present an account by Anderson that aims

³⁶⁸ Ibid.

³⁶⁹ Cf., *ibid.*

³⁷⁰ Cf., McHugh: “Epistemic Communities and Institutions”, 274-275.

to respond to this question. But before I outline how Anderson challenges testimonial justice in regard to its status as an individual virtue I will first introduce some background assumptions she makes.

Drawing on distributive theories of justice, Anderson distinguishes between the transactional and the structural aspect of distributive justice.³⁷¹ Furthermore, she claims that such a distinction can also be applied to a theory of epistemic injustice or rather, that the distinction is important in order to come up with more substantial remedies against epistemic injustice.³⁷²

Following this analogy, the transactional aspect concerns exchanges and interactions between two people. In Fricker's theory of epistemic injustice this corresponds to testimonial injustice. As Anderson puts it: "[F]ricker's illustrations of testimonial injustice depict it as a transactional injustice."³⁷³ Testimonial injustice, Anderson clarifies, remains episodic and transactional because Fricker sets her concept of identity prejudice at its core. Although identity prejudice stems from *collective* imagination, an *individual* hearer downgrades an *individual* speaker because the hearer's judgements are influenced by these prejudices.³⁷⁴ However, in some cases, testimonial injustice becomes structural. Anderson explains this by drawing on hermeneutical injustice, which, in her view, is always structural.

Hermeneutical injustice occurs when a society lacks the interpretive resources to make sense of important features of a speaker's experience, because she or members of her social group have been prejudicially marginalized in meaning making activities.³⁷⁵

Hermeneutical injustice is structural since hearers are not able to make sense of certain social experiences. They lack the interpretative resources to decode these experiences. In hermeneutical injustice, identity prejudice, again, is the driving source to cause the injustice.³⁷⁶ In a further step, Anderson asks if credibility deficits arising in instances of

It needs to be acknowledged that Fricker addresses the possibility of institutional virtues in her 2009 paper "Can There Be Institutional Virtues". In this paper she argues that institutions can possess collective epistemic virtues which are "[...] based on the needs of the institution and those that the institutions serve." (Ibid.) In short, epistemic virtues of institutions need to be cultivated in order to promote democratic ideals. (Cf., *ibid.*)

³⁷¹Cf., Elizabeth Anderson, "Epistemic Justice as a Virtue of Social Institutions", 164.

³⁷² Cf., *ibid.*

³⁷³ *Ibid.*, 165.

³⁷⁴ Cf., *ibid.*

³⁷⁵ *Ibid.*, 166.

³⁷⁶ Cf., *ibid.*

testimonial, transactional injustice must necessarily be traced to prejudice in order for them to be unjust. In other words, Anderson asks what happens if epistemic injustice comes about not only through identity prejudice, but through other, more structural factors of ignorance that stem from a lack of hermeneutical resources?

This is an important question in order to understand how Anderson criticizes Fricker's model of the virtue of testimonial justice. She claims that Fricker's focus on individual epistemic virtues as antidotes to epistemic injustice can be challenged regarding two aspects. The first challenge demonstrates that testimonial justice may not effectively counteract testimonial injustice, since, as I have suggested above, influence of implicit bias poses a threat to an antidote that stresses self-regulation through critical reflective awareness. "Reflection, which lies at the core of testimonial justice, is cognitively taxing and impossible to keep up in environments that demand rapid responses."³⁷⁷ Although Anderson concedes that virtues can become habitual and thus can counter the influence of bias (my emphasis), individuals first have to develop the cognitive, hermeneutical resources that set the necessary initial move of critical reflection into motion. A lack of hermeneutical resources that prevents testimonial justice from being developed is one reason why Anderson believes "[...] that structural remedies need to be stressed even when the injustices at issue are transactional."³⁷⁸

This brings us to the second aspect of testimonial justice that is challenged by Anderson. She argues that structural forms of testimonial injustice may be more pervasive than Fricker acknowledges.³⁷⁹ Consequently, testimonial justice "[...] may not address certain structural epistemic injustices that may have locally innocent (non-prejudicial causes), but require structural remedies."³⁸⁰

In other words, Anderson aims for a more social account of epistemic virtues. She suggests expanding testimonial justice to the extent that it becomes not only an individual virtue, but also a structural virtue of social institutions.³⁸¹ Subsequently, structural remedies can be regarded as virtue-based remedies for collective agents, such as judicial, educational,

³⁷⁷ Ibid., 168.

³⁷⁸ Ibid.

³⁷⁹ Cf., *ibid.*, 169.

³⁸⁰ Ibid., 167.

³⁸¹ Cf., McHugh: "Epistemic Communities and Institutions", 275.

political and medical social institutions.³⁸² For Anderson, social institutions are epistemically just when its members “[...] jointly commit themselves to operating according to institutionalized principles that are designed to achieve testimonial justice, such as giving hearers enough time to make unbiased assessments.”³⁸³

To understand how groups can become structural perpetrators of testimonial justice, Anderson emphasizes that we need to look beyond a merely prejudice-based model of testimonial injustice. She considers three distinctively structural causes of group-based credibility deficits, namely differential access to markers of credibility, ethnocentrism and “shared-reality bias”.³⁸⁴

To illustrate this, let us look at the first non-prejudicial cause for epistemic injustice, differential markers of credibility. Fricker rightly claims that we rely on various markers of credibility in order to spontaneously assess whether a speaker is trustworthy or not. For instance, education could be a legitimate marker of credibility in situations “[...] where educated judgement is called for.”³⁸⁵ Therefore, in some instances it is not a testimonial injustice, if lack of education leads to a reduced credibility attribution. “Yet in societies that systematically deprive disadvantaged social groups of access to a decent education, the use of such markers in assessing credibility will tend to exclude those groups from further participation in inquiry.”³⁸⁶ This original structural injustice, namely denial of fair opportunity, generates obstacles that impede the epistemic agency of people who are less privileged. Although no testimonial injustice is committed when judging someone as less credible because she lacks a certain level of education, a structural injustice is perpetrated since the person never had the opportunity to reach the required level of education.³⁸⁷

The same mechanism applies to ethnocentrism and the shared-reality bias. Originally innocent errors can turn into vectors of hermeneutical injustice and structural testimonial injustice.³⁸⁸ In other words, these three structural phenomena can create and sustain

³⁸² Cf., Anderson, “Epistemic Justice as a Virtue of Social Institutions”, 168.

Anderson draws on Gilbert’s account of group agency in order to justify her claim that testimonial justice can become a virtue of social collectives.

³⁸³ Ibid., 168-169.

³⁸⁴ Cf., *ibid.* 169.

³⁸⁵ Ibid.

³⁸⁶ Ibid.

³⁸⁷ Cf., *ibid.*

³⁸⁸ Cf., *ibid.*, 171.

structural and group-based credibility deficits, “[...] even though when analyzed from the level of an individual epistemic agent, there was no individual epistemic failing.”³⁸⁹

For that reason, Anderson proposes that testimonial justice needs to become institutionalized. “A structural remedy for epistemic injustice is a virtue of large-scale systems of inquiry. Just as individuals are accountable for how each acts independently, we are accountable for how we act collectively.”³⁹⁰ In short, the virtue of testimonial justice will only be successful if individuals *and* institutions adopt it to promote justice and equality. To create sustainable epistemic justice, we need to reconstruct our epistemic institutions to the extent that they prevent epistemic injustice.

To summarize, Fricker’s model of testimonial justice can be a successful antidote in cases where the hearer gains awareness of her distorting prejudices towards a speaker and aims at correcting for them. However, I identified two problems related to testimonial justice. First, I tried to demonstrate that reflective monitoring, as a control mechanism against prejudice is rendered significantly harder when it comes to implicit prejudice. Therefore, I suggested to stress the role of habituation in virtue theory and argued for the importance of automatization of egalitarian values in order to counter the influence of implicit prejudice. Second, I aimed at looking beyond what hearers and individuals can do to prevent testimonial injustice. By drawing on Medina’s deliberations on meta-lucidity as a virtue of the oppressed, I wanted to provide an alternative to Fricker’s account of testimonial justice since it neglects the potential agency of those who are the victims of testimonial injustice. Ultimately, I believe that testimonial justice and lucidity should not be understood as divergent, but rather, the two virtues should be acknowledged as convergent. Ameliorated testimonial justice is therefore an epistemic virtue that should strive to include the perspective of those whose epistemic situation it tries to improve.

Finally, since oppression is a structural phenomenon, we should aim at formulating an account of epistemic virtues that outgrows individual commitments and emphasizes the importance of structural remedies.

³⁸⁹ Cf., McHugh: “Epistemic Communities and Institutions”, 275.

³⁹⁰ Anderson, “Epistemic Justice as a Virtue of Social Institutions”, 171.

5. Conclusion

In this thesis I have sought to illustrate how Fricker theorizes testimonial injustice as an injustice that is intrinsically epistemic. Fricker describes how the reception of testimony of people from marginalized groups is arbitrarily corrupted on the basis of prejudice. By doing so, she shows how epistemic injustice operates systematically through a specific kind of prejudice, that is, identity prejudice. Moreover, Fricker outlines how a wrongfully attributed status of credibility disadvantages the individual affected and harms the epistemic community as a whole.

My objective was to provide a thorough analysis of Fricker's account. Furthermore, I detected some shortcomings which relate to her considerations about implicit prejudice. I argued that Fricker neglects the role residual implicit prejudice plays in cases of testimonial injustice. Although she introduces the concept of residual internalizations to deal with this issue, she hardly makes use of it in the course of her argument. Moreover, Fricker claims that implicit prejudice is pervasive and thus likely the main reason for testimonial injustice. Still, her account is primarily based on actively endorsed prejudices. This constitutes an inconsistency I have aimed to resolve.

Fricker's focus on actively endorsed prejudice also resonates with the way she proposes to counter testimonial justice, that is, through virtuous self-critical monitoring. In general, I supported Fricker's proposition to draw on virtue ethics in order to minimize cases of testimonial injustice. And yet, I have also pointed out that Fricker's account lacks an important insight: we need to come up with an antidote that takes the pervasiveness of implicit prejudice into consideration. Mere awareness about one's implicit prejudices does not automatically imply that one is able to control them.

That is why this thesis put special emphasis on implicit-bias research. I argued that findings in implicit-bias research offer a way to expand Fricker's notion of residual internalizations, as well as her entire account of testimonial justice. Integrating implicit-bias research into a theory of epistemic injustice can therefore be seen as a fruitful strategy for broadening our understanding as to how implicit prejudices lead to cases of testimonial injustice.

In order to outline this improvement of Fricker's theory, I first indicated the parallels and intersections between implicit-bias research and Fricker's account. I showed that Fricker's

conception of prejudice exhibits many parallels with accounts of implicit biases. For instance, by drawing on Gendler's concept of aliefs, the thesis tried to show that Fricker's description of identity prejudices intersects with aliefs in terms of their associative, image-like structure and their mode of activation. My goal was to point out that mere knowledge of the existence of a certain prejudice does not rule out being unconsciously influenced by it.

Moreover, Fricker underlines the fact that identity prejudices are resistant to counter-evidence.³⁹¹ Social psychology confirms this. Often prejudice-confirming behavior is recognized, whereas prejudice-disconfirming behavior is ignored. These two mechanisms make it even more likely that implicit prejudices can flourish.

Anderson also reflects upon these two aspects of implicit prejudice and she coins the terms "attention bias" and "attribution bias". She uses them in order to analyze media reactions to "[...] stranded people's behavior in the aftermath of Hurricane Katrina."³⁹²

News media described stranded blacks as "looting" grocery stores for necessities [...], abstracting from the desperate circumstances brought on by the storm, and the fact that the flood would have otherwise destroyed these groceries. The "looting" frame fit their actions into the narratives of inner-city riots, invoking the stigma of inherent black criminality. By contrast, stranded whites hauling groceries from stores were generously inferred to have merely "found" them by innocent luck.³⁹³

This example illustrates how pervasive implicit biases are. They influence our perceptions, our judgements and our spontaneous assessments of situations. The findings provided by the "IAT", the Implicit Association Test, are further proof for the ubiquity of implicit bias. This test measures implicit attitudes in individuals and demonstrates how we are affected by implicit biases in our judgements. That is why this thesis argued that in a lot of cases, testimonial injustice likely is the product of implicitly endorsed identity prejudice.

I have claimed that implicit-bias research thus is the best scientific explanation available for the role of implicit prejudice. At an earlier point in the thesis I stated that this argument can be framed as a methodological naturalist argument. In order to deepen our understanding of how implicit prejudices function within Fricker's account, one should look at the best scientific theories available to explain implicit attitudes. These are provided by implicit-bias

³⁹¹ Cf., Fricker, *Epistemic Injustice*, 35.

³⁹² Cf., Medina, *The Epistemology of Resistance*, 165.

³⁹³ Elizabeth Anderson, *The Imperative of Integration*, (Princeton, NJ: Princeton University Press, 2011), 165.

research.

In the last section of this thesis, I addressed normative questions that result from relating implicit-bias research to a theory of testimonial justice. First, I introduced an account by Saul that analyzes to what extent implicit bias gives rise to a specific kind of skepticism, that is, bias-related doubt. Saul argues that implicit biases affect us in unsettling ways. For instance, they are able to influence us even before we consciously try to make sense of the world because they “[...] affect our very perceptions of the world [...] in worrying ways.”³⁹⁴ Bias-related doubt therefore suggests that we are likely to make errors when evaluating testimony. I reasoned that it is crucial for a theory of testimonial injustice to address phenomena such as bias-related skepticism in order to reflect upon the possibility that some of our credibility judgements are not reliably formed.

In a further step, I elaborated on epistemic responsibilities and obligations. I argued that agents are responsible for their actions, even when the beliefs, setting their behavior in motion are informed by implicit bias. I presented two exculpating conditions for blameworthiness, that is, knowledge and control, and showed why in most cases implicit biases do not meet these exculpating conditions. Scientific knowledge about the probability of being influenced by implicit bias exists and is steadily increasing. When we are familiar with this knowledge, we can therefore be held accountable for judgements that are affected by implicit bias. And yet, by referring to Medina, I also claimed that knowledge about implicit bias is not equally accessible to everyone. Thus, rather than holding every person accountable for their implicit biases to the same degree, we need to come up with a comparative system of responsibility that takes into consideration what kind of knowledge is available to whom. Finally, I briefly discussed an account by Fricker in which she provides a way of thinking about epistemic responsibilities for non-culpable implicit-bias influence, that is, epistemic agent-regret.

In the last section I tried to introduce a remedy against testimonial injustice that takes the possibility of implicit-bias influence into consideration. I pointed out that testimonial justice needs to be expanded in order to successfully block implicit prejudice as well.

First, I focused on the important role that habit plays in the acquisition of virtues. Although Fricker also mentions the relationship between habit and virtue, she does not see

³⁹⁴ Saul, “Skepticism and Implicit Bias”, 245.

it as a way of preventing implicit prejudice. I have emphasized this relationship and have argued that automatically activated implicit prejudice can only be successfully defeated when the virtues countering it is equally automatically activated.

In a last step, I criticized that Fricker's model of testimonial justice downplays the agency of the victims of testimonial injustice. That is why I drew on Medina's concept of "metaculidity" as a virtue that likely occurs among oppressed subjects.

Finally, I referred to Anderson and claimed that we should not only strive for just individual, but primarily for structural solutions when counteracting epistemic injustice. Anderson argues that testimonial justice ought to become a justice of social institutions in order to promote equality and diversity.³⁹⁵ "Just as it would be better and more effective to redesign economic institutions so as to prevent mass poverty in the first place, it would be better to reconfigure epistemic institutions so as to prevent epistemic injustice from arising."³⁹⁶ We live, work, learn, speak and listen to each other in a society, in which certain groups are deprived of educational and other opportunities. That is why the way in which we assess testimony should acknowledge and consider these different social standings. As Anderson puts it, "[...] an original structural injustice – denial of fair opportunities for education – generates additional structural inequalities in opportunities for exercising full epistemic agency, which is an injustice to the speakers."³⁹⁷ Before we dismiss someone's testimony as uninformed or unqualified because of their group affiliation, foreign accent or the lack of an academic degree, we should therefore ask ourselves which opportunities were given to them, and which opportunities I would grant them, in taking their testimony seriously. An inclusive society is one that promotes epistemic equality and fairness. With this in mind, we should not only be concerned with explicit prejudices. Rather, we should ask in which ways we might be affected by implicit attitudes ourselves. As I repeatedly pointed out, implicit prejudice is pervasive. To consider oneself immune against it would mean to perpetrate an epistemic vice such as closed-mindedness.

In conclusion, I want to underline that it typically takes a position of epistemic power and privilege to be able to elaborate on phenomena such as epistemic injustice and implicit bias.

³⁹⁵ Cf., Anderson, "Epistemic Justice as a Virtue of Social Institutions", 171.

³⁹⁶ Ibid.

³⁹⁷ Ibid., 169.

For instance, am I really in a position to address, without perpetuating a certain epistemic hierarchy, the kind of epistemic oppression experienced by black or Muslim-women? When formulating a theoretical account of epistemic injustice and discrimination one should be cautious not to fall into the trap of “speaking for others”³⁹⁸, as Alcoff puts it. Doing so presents a kind of epistemic smothering of individuals and groups who might not have the resources or opportunities to speak up about their epistemic situations themselves. Still, the concepts of “epistemic allies” and “active bystander-ship”³⁹⁹ have gained increasing attention in philosophy. I conclude that accounts on injustice in general, and epistemic injustice in particular, should ideally be produced in consultation with members of the affected groups. The question we need to ask ourselves is the following: will our presentation of the experience of an epistemically oppressed group enhance the groups’ empowerment and epistemic standing?

³⁹⁸ Cf., Linda Alcoff, “The Problem of Speaking for Others”, in: *Cultural Critique*. No. 20. (1992), 5-32.

³⁹⁹ The term of the “active bystander” was coined by Rachel MacKinnon as a reaction to increasing criticism regarding a certain kind of “ally culture”. Cf., Rachel MacKinnon, “Allies Behaving Badly. Gaslighting as Epistemic Injustice”, in: *The Routledge Handbook of Epistemic Injustice*, ed. Kidd, Ian James, (New York: Routledge, 2017), 167-174.

LITERATURE

- Alcoff, Linda. (2010). "Epistemic Identities". *Episteme*. Vol. 7. Issue 2. 128-137.
- Alcoff, Linda. (1992). "The Problem of Speaking for Others. *Cultural Critique*. No. 20. 5-32.
- Anderson, Elizabeth. (2012). "Epistemic Justice as a Virtue of Social Institutions", in: *Social Epistemology*. 26:2. 163-173.
- Anderson, Elizabeth. (2011). *The Imperative of Integration*. Princeton, NJ: Princeton University Press.
- Aronson, Eliott, Wilson, Timothy, Akert, Robin M. (2013). *Social Psychology*, Hoboken, NJ: Pearson Education Inc.
- Frankish, Keith. (2016). "Playing Double. Implicit Bias, Dual Levels and Self-Control", in: *Implicit Bias and Philosophy. Volume 1. Metaphysics and Epistemology*. Saul, Brownstein (ed.). Oxford: Oxford University Press. 23-46.
- Foucault, Michel. (1982). "How is Power Exercised?", trans. Sawyer, Leslie from Afterword in: *Michel Foucault: Beyond Structuralism and Hermeneutics*. Dreyfus, H.L. and Rabinow, P: Hemel. (ed.). Hempstead: Harvester Press.
- Fricker, Miranda. (2007). *Epistemic Injustice. Power & the Ethics of Knowing*. New York: Oxford University Press.
- Fricker, Miranda. (2017). "Evolving Concepts of Epistemic Injustice". in: *The Routledge Handbook of Epistemic Injustice*, Kidd, Ian James. (ed.). New York: Routledge. 53-61.
- Fricker, Miranda. (2016). "Fault and No-Fault Responsibility for Implicit Prejudice – A Space for Epistemic Agent Regret", in: *The Epistemic Life of Groups. Essays in the Epistemology of Collectives*. Michael S. Brady, Miranda Fricker. (ed.). Oxford: Oxford University Press. 33-51.
- Fricker, Miranda. (2010). "Replies to Alcoff, Goldberg and Hookway on *Epistemic Injustice*", in: *Episteme*. Vol. 7. Issue 2. 164-178.
- Gendler, Tamar Szabó. (2011). "On the Epistemic Costs of Implicit Bias", in: *Philosophical Studies N°156*. 33-63.
- Gendler, Tamar Szabó. (2010). *Intuition, Imagination & Philosophical Methodology*, Oxford: Oxford University Press.

- Goguen, Stacey. (2016). "Stereotype Threat, Epistemic Injustice and Rationality", in: *Implicit Bias and Philosophy. Volume 1. Metaphysics and Epistemology*. Saul, Brownstein. (ed.). Oxford: Oxford University Press. 216-237.
- Hamilton, David, Sherman, Jeffrey. (1994). "Stereotypes", in, *Handbook of Social Cognition*. (2nded. Vol.2). R.S. Wyer Jr. & T.K. Srull (eds.). Hillsdale, NJ: Erlbaum.
- Kidd, Ian James, Medina, José and Pohlhaus, Gaile Jr. (2017). "Introduction to the Routledge Handbook of Epistemic Injustice", in: *The Routledge Handbook of Epistemic Injustice*. Kidd, Ian James (ed.). New York: Routledge. 1-11.
- Kusch, Martin, Review. (2009). "Miranda Fricker: Epistemic Injustice: Power and the Ethics of Knowing", in: *Mind, Volume 118, Issue 469*. 170–174.
- Lippmann, Walter. (1922). *Public Opinion*. New York: Harcourt, Brace.
- McHugh Arden, Nancy. (2017). "Epistemic Communities and Institutions", in: *The Routledge Handbook of Epistemic Injustice*. Kidd, Ian James. (ed.) New York: Routledge. 270-279.
- McKinnon, Rachel. (2017). "Allies Behaving Badly. Gaslighting as Epistemic Injustice", in: *The Routledge Handbook of Epistemic Injustice*, Kidd, Ian James (ed.). New York: Routledge. 167-174.
- Medina, José. (2013). *The Epistemology of Resistance. Gender and Racial Oppression, Epistemic Injustice and Resistant Imaginations*, New York: Oxford University Press
- Minghella, Anthony. (2000). *The Talented Mr. Ripley* – Based on Patricia Highsmith's Novel. London: Methuen.
- Nussbaum, Martha. (1990). "The Discernment of Perception: An Aristotelian Conception of Private and Public Rationality.", in: *Love's Knowledge: Essays on Philosophy and Literature*. Oxford: Oxford University Press. 54-105.
- Puddifoot, Katherine. (2017). "Epistemic Discrimination", in: *The Routledge Handbook of the Ethics of Discrimination*. Kaspar Lippert-Rasmussen (ed.). New York: Routledge. 54-67.
- Rees, Clea F. (2016). "A Virtue Ethics Response to Implicit Bias", in: *Implicit Bias and Philosophy. Volume 2. Moral Responsibility, Structural Injustice and Ethics*. Saul, Brownstein. (ed.) Oxford: Oxford University Press. 191-214.
- Saul, Jennifer, Brownstein, Michael. (2016). *Implicit Bias and Philosophy. Metaphysics and Epistemology. Volume 1*. Oxford: Oxford University Press.

- Saul, Jennifer. (2017). "Implicit Bias, Stereotype Threat and Epistemic Injustice", in: *The Routledge Handbook of Epistemic Injustice*. Kidd, Ian James, (ed.). London, New York: Routledge. 234-243.
- Saul, Jennifer. (2012). "Skepticism and Implicit Bias", in: *Disputatio, Vol. V, No. 37*. 243-263.
- Schwitzgebel, Eric. (2010). "Acting Contrary to Our Professed Beliefs, or the Gulf between Occurrent Judgement and Dispositional Belief", in: *Pacific Philosophical Quarterly* 91(4). 531-553.
- Shklar, Judith. (1990). *The Faces of Injustice*. New Haven, London: Yale University Press
- Stangor, Charles, Ford, Thomas. (1992). "Accuracy and Expectancy-Confirming Processing in Orientations and the Development of Stereotypes and Prejudice, in: *European Review of Social Psychology. Vol. 3:1*.57-89
- Vavova, Katia. (2018). "Irrelevant Influences", in: *Philosophy and Phenomenological Research, Vol. XCVI, No. 1*. 134-152.
- Washington, Natalia, Kelly, Daniel. (2016). "Who's Responsible for This? Moral Responsibility, Externalism, and Knowledge about Implicit Bias", in: *Implicit Bias and Philosophy. Volume 2. Moral Responsibility, Structural Injustice and Ethics*, Saul, Brownstein. (ed.). Oxford: Oxford University Press. 11-37.
- Williams, Bernard. (1993). *Shame and Necessity*. Berkley: University of California Press.
- Zagzebski, Linda. (1996). *Virtues of Mind - An Inquiry into the Nature of Virtue and the Ethical Foundation of Knowledge*. Cambridge: Cambridge University Press.
- <https://plato.stanford.edu/entries/implicit-bias/>