UDC 101 | ISSN 1845-8475

EUROPEAN JOURNAL OF ANALYTIC PHILOSOPHY

Vol. 21, No. 1, 2025

BOOK SYMPOSIUM ON HOW STEREOTYPES DECEIVE US

STEREOTYPES DECEIVE US, BUT NOT IN THE WAY WE COMMONLY THINK: INTRODUCTION TO THE BOOK SYMPOSIUM, **Marina Trakas** | ON THE DISCOMFORT WITH THE ACCURACY CRITERION: REASSESSING PUDDIFOOT'S CRITICISM OF THE DUAL FACTOR VIEW, **Federico José Arena** | ACCURATE STEREOTYPES AND TESTIMONIAL INJUSTICE, **Leonie Smith** | SOME CRITICAL THOUGHTS ON "HOW STEREOTYPES DECEIVE US", **Federico Picinali** | SOME STARTLING CONSEQUENCES OF HOW STEREOTYPES DECEIVE US, **Jennifer Saul** | REPLIES TO CONTRIBUTORS, **Katherine Puddifoot**

BOOK REVIEW

Lisa Bortolotti THE EPISTEMIC INNOCENCE OF IRRATIONAL BELIEFS, Oxford University Press, 2020, **Dino Jelčić**

EUROPEAN JOURNAL OF
ANALYTIC PHILOSOPHY

UDC 101
ISSN (Print) 1845-8475
ISSN (Online) 1849-0514
https://doi.org/10.31820/ejap

Publisher, Editorial office
University of Rijeka, Faculty of Humanities and Social Sciences, Department of Philosophy
Address: Sveučilišna avenija 4, 51000 Rijeka, Croatia (EU)
Phone: +385 51 669 794
E-mail: eujap@ffri.uniri.hr
Web address: https://www.eujap.uniri.hr

The publication of the journal is financially supported by the Ministry of Science, Education and Youth of the Republic of Croatia under a programme agreement with the University of Rijeka.

TABLE OF CONTENTS

BOOK SYMPOSIUM

Stereotypes Deceive Us, but Not in the Way We Commonly Think: Introduction to the Book Symposium
Marina Trakas……1-18

On the Discomfort with the Accuracy Criterion: Reassessing Puddifoot's Criticism of the Dual Factor View
Federico José Arena……..9-24

Accurate Stereotypes and Testimonial Injustice
Leonie Smith…….25-38

Some Critical Thoughts on "How Stereotypes Deceive Us"
Federico Picinali……..39-55

Some Startling Consequences of How Stereotypes Deceive Us
Jennifer Saul…..57-65

Replies to Contributors
Katherine Puddifoot…..67-90

BOOK REVIEW

Lisa Bortolotti THE EPISTEMIC INNOCENCE OF IRRATIONAL BELIEFS, Oxford University Press, 2020
Dino Jelčić……..91-93

ABSTRACTS (SAŽECI).……..(AB)

AUTHOR GUIDELINES AND MALPRACTICE STATEMENT…….(AG)

STEREOTYPES DECEIVE US, BUT NOT IN THE WAY WE COMMONLY THINK: INTRODUCTION TO THE BOOK SYMPOSIUM

Marina Trakas[1]

[1] Instituto de Investigaciones Filosóficas (IIF), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Editor's note – Received: 04/02/2025 Accepted: 19/02/2025

ABSTRACT

This is an introduction to the book symposium on Katherine Puddifoot's *How Stereotypes Deceive Us*.

Keywords: epistemic benefits; epistemic costs; stereotypes.

Correspondence: marina.trakas@conicet.gov.ar

Katherine Puddifoot's *How Stereotypes Deceive Us* (2021) is a fascinating reflection on the theoretical and philosophical underpinnings of stereotypes. Drawing on research in psychology, the main objective is to illuminate the various ways in which stereotyping makes people susceptible to errors in perception and judgment, with a focus on the epistemic dimension of stereotypes. It demystifies the notion that stereotypes are misleading simply because they are always inaccurate, as well as the idea that only inaccurate stereotypes are epistemically harmful. It also proposes innovative theoretical frameworks to understand the (epistemic) complexities surrounding acts of stereotyping and to evaluate such acts.

Following a brief introduction in Chapter 1, which outlines the main ideas and provides an overview of the book, Chapter 2 focuses on defending a non-normative account of stereotypes against normative accounts. Normative perspectives regard stereotypes as inherently "bad", assuming they are always inaccurate and distorting. In contrast, non-normative accounts recognize that stereotypes can be either accurate or inaccurate, sometimes distorting judgments and other times improving accuracy. Puddifoot critiques several arguments against the normative view before ultimately defending a pragmatic argument for the non-normative account: acknowledging that stereotypes are not necessarily inaccurate allows for a more nuanced study of their functioning and facilitates their integration into broader research on cognitive structures that are not inherently deviant or flawed. However, this approach is criticized by two commentators: whereas Federico Arena stresses the importance of thoroughly discussing the concept of stereotype, Leonie Smith questions whether stereotypes can really be accurate, given that its accuracy is always related to a given population.

Additionally, in this chapter, Puddifoot defines stereotypes as social attitudes that associate certain traits more strongly with members of some social groups than with others. Yet, this definition is also deemed unsatisfactory by Federico Picinali and Jennifer Saul, the other two commentators in this special issue, and Arena, who emphasizes the importance of considering the normative force of stereotypes beyond their descriptive aspect. For Puddifoot, her definition has the advantage of remaining neutral regarding the psychological underpinnings of stereotypes, emphasizing that they are neither concepts nor necessarily beliefs, as they can also exist as implicit attitudes.

Chapter 3 introduces one of Puddifoot's key contributions: a multifactorial view of stereotyping, which asserts that multiple factors influence whether applying a stereotype increases or decreases the likelihood of making an

accurate judgment about an individual or case. Puddifoot challenges both the single-factor view, which holds that stereotypes merely reflect some aspect of reality, and the dual-factor view, which considers both the accuracy of the stereotype and its application only when evidence about a case is ambiguous.

According to Puddifoot, accuracy goes beyond simply reflecting some aspect of social reality—it requires alignment with base-rate information and the statistical distribution of traits across different populations. A stereotype may capture certain aspects of reality yet fail to prompt the person using it to respond in a way that aligns with statistical reality. For this reason, the dual-factor view is preferred over the single-factor view.

However, even the dual-factor view is, according to Puddifoot, insufficient. Beyond assessing whether a stereotype reflects social reality, is statistically accurate, or is applied only in cases of ambiguous evidence, Puddifoot argues that additional crucial factors must be considered. One such factor is whether the stereotype is applied in situations where it is entirely irrelevant (e.g., if it is triggered by an individual's wounded ego). Another is whether the stereotype introduces distortions that affect case-specific information, whether that information is ambiguous or unambiguous. These considerations lead her to formulate the multifactorial view of stereotypes, which incorporates these various elements into the epistemic evaluation of stereotypes.

Puddifoot's arguments face criticism from two commentators, though. Picinali contends that her rejection of the single-factor and dual-factor views suffers from conceptual weaknesses and shortcomings, and that her notion of "relevance" is imprecise. Meanwhile, Arena contends that the dual-factor view of stereotypes can already account for the additional factors Puddifoot identifies, making the multifactorial view unnecessary.

Chapter 4 challenges the idea that having thoughts that reflect reality is always the most beneficial from an epistemic perspective. Puddifoot argues that stereotypes that reflect reality can incur epistemic costs by increasing the chance of misperceptions and misjudgments. Conversely, stereotypes that fail to reflect reality can bring epistemic benefits, even when they do not align with base-rate or background statistical information, sometimes outweighing the epistemic costs of their inaccuracy. Regarding the first point, Puddifoot expands on ideas presented in the previous chapter, providing detailed real-world examples of how stereotypes that reflect some aspects of social reality can lead to misperceptions about individuals and misexplanations of their behavior. These stereotypes can influence judgments about other aspects of social

reality, fail to highlight associations that would reflect social reality, or encourage associations that do not. They can also lead to overlooking important differences and similarities between members of different groups, discrediting testimony, and other distortions.

On the second point, Puddifoot argues that there can be epistemic benefits to holding social attitudes and beliefs that do not reflect reality, especially when these attitudes portray the world as more egalitarian than it actually is. This is framed through Lisa Bortolotti's notion of *epistemic innocence*, which suggests that epistemically faulty cognitive states can be valuable for achieving epistemic goals, particularly when no alternative cognition would provide the same benefits without the associated costs. While some egalitarian social attitudes may carry epistemic costs due to their failure to reflect reality, Puddifoot contends that the epistemic benefits often outweigh these costs. However, the supposed epistemic benefits of egalitarian social attitudes have been criticized by both Picinali and Arena. Ultimately, the main claim that Puddifoot wants to make is that, in some cases, neglecting statistical information—typically considered an irrational act—and focusing on accessing and processing case-specific information can actually be the best epistemic strategy for achieving epistemic goods.

In this lengthy chapter, Puddifoot also evaluates the implications of her analysis for ethical and epistemological accounts of stereotyping, including Blum's analysis of the moral objection to stereotypes based on their falsity, Fricker's account of epistemic injustice grounded solely in prejudice, and moral encroachment views that invoke moral considerations to explain the epistemic wrongness of stereotyping. Puddifoot's criticisms have nonetheless generated controversy among commentators. Smith argues that, since stereotypes can be accurate in relation to a specific population, prejudice may lie behind accurate stereotypes that are epistemically faulty, making Fricker's notion of epistemic injustice a useful explanation for the epistemic harms in these cases. On the other hand, Picinali contends that Puddifoot's dismissal of the moral encroachment approach is too quick, defending the idea that this approach can also account for cases where stereotyping is justified.

Building on the previous discussion in the last part of the previous chapter on moral encroachment views, Chapter 5 further explores the relationship between ethical and epistemic goals, critiquing conceptions that oversimplify the issue by presenting stereotyping as an epistemic-ethical dilemma. According to this simplified view, failing to apply an accurate stereotype leads to epistemic errors, while applying it is considered unethical. Puddifoot argues that the dilemma is far more complex, as ethical and epistemic demands do not always conflict. The real dilemma

lies in discerning whether, in specific contexts, epistemic and ethical goals—either separately or together—are more likely to be achieved by stereotyping or by avoiding it. To explore this, she focuses on a case study: stereotyping in healthcare.

First, she rejects a potential solution based on her previous argument: the idea that there is no conflict between moral and epistemic goals because avoiding stereotyping is always the best epistemic strategy. Although in many situations ethical and epistemic principles align—such that avoiding stereotyping is the best approach from both perspectives—there are times when these principles point in different directions. Puddifoot argues that stereotypes are not inherently misleading. In some cases, the social group to which a person belongs may be highly relevant for making an accurate judgment, such as in the context of medical diagnoses. In these cases, the epistemic benefits of stereotyping can clash with the ethical costs. But in other instances, stereotyping may also align with ethical goals: treating people fairly sometimes requires making correct judgments, and in certain situations, this may involve stereotyping. In conclusion, while epistemic and ethical goals can converge—making stereotyping either the most or the least ethical course of action in some cases—they do not always align. Puddifoot suggests that, given the potential epistemic benefits of stereotyping, the ideal strategy to avoid its negative consequences is not to eliminate stereotypes, but to take control of them and their influence on our mental processes.

Like the previous chapter, Chapter 6 has not sparked controversy among commentators. It offers a conceptual analysis based on a specific case study: individuals with mental health conditions, although its insights can be extended to other social identities and group memberships. Puddifoot's main aim is to examine the other side of the coin—the dilemma faced by people who are stereotyped—through the lens of the multifactorial approach to stereotypes. In some cases, disclosing mental health issues increases the likelihood of being stereotyped and, consequently, misperceived. Conversely, choosing not to disclose such information can also have negative consequences, as it may hinder others from properly understanding a person's dispositions, mental states, and behaviors, leading to misperceptions. These misperceptions can carry significant costs, including pragmatic disadvantages and psychological harm, such as being seen as having a flawed character. Given this, individuals face a difficult dilemma: whether to disclose or conceal information about their mental health. According to Puddifoot, an adequate solution requires primarily social changes, but also actions from individuals other than those with certain social identities. This would create conditions in which people can disclose information without fear of misperception. Therefore, the

dilemma cannot be fully resolved by those who are, or are at risk of being, stereotyped.

In Chapter 7, Pudifoot argues that Western analytical accounts of epistemic rationality and justification fail to capture all the epistemic flaws associated with stereotypes. According to these accounts, one can be considered rational and justified in believing certain stereotypes, even when such beliefs entail significant epistemic costs. This limitation makes these accounts inadequate for fully evaluating the epistemic status of stereotypical beliefs. The author identifies three main types of epistemological accounts and provides real-world examples of each (which I will not detail here). Upstream accounts focus on the causal history of a belief, considering factors such as the quality of evidence and the reliability of the belief-formation process. Downstream accounts emphasize the consequences and practical outcomes of holding a belief, such as in the epistemic innocence approach. Finally, static accounts concentrate on the internal features of the belief itself, such as its coherence with other beliefs.

According to Puddifoot, upstream accounts effectively capture epistemic flaws related to inaccurate stereotypes—those that do not align with the evidence or result from unreliable belief-forming processes. However, they fall short when it comes to stereotypes that are well-supported by evidence and formed through reliable processes but still lead to distorted views of individuals. In such cases, these accounts may evaluate the belief positively despite its problematic epistemic consequences. Downstream accounts are better suited to addressing these epistemic flaws because they focus on the broader impacts of the belief. However, as Puddifoot argues, relying solely on consequences, without considering the evidential basis, undermines the prescriptive force of concepts like justification and rationality. The causal history of a belief is still relevant to its epistemic evaluation. Static accounts also fail to capture all the epistemic flaws of stereotypes because they overlook the role of consequences. For example, within a coherentist framework, a false stereotype might be fully coherent with the belief system of a highly prejudiced individual, thereby being considered justified and rational within that context.

Since none of these accounts, taken individually, can fully address the epistemic faults of stereotypical beliefs, the author proposes a pluralistic approach. This approach, called “evaluative dispositionalism” and developed in Chapter 8, the final chapter of the book, integrates insights from all three perspectives: upstream, downstream, and static, to offer a more comprehensive evaluation of the epistemic status of stereotypes. It is also one of Puddifoot’s key contributions.

According to evaluative dispositionalism, evaluating an act of stereotyping requires focusing on the full set of epistemic dispositions associated with believing the target stereotype. This includes both dispositions that are manifest in the act of believing and those that are possessed as a result of holding the belief. The latter includes dispositions such as the tendency to adopt related beliefs, to misremember or misinterpret evidence, and so on. Importantly, dispositions are not limited to tendencies to act; they also encompass tendencies to think, speak, and consider what one would do in hypothetical scenarios.

Puddifoot distinguishes her position from similar views in the literature and illustrates the explanatory advantages of her framework through specific cases. These advantages include:

> (a) the ability to capture epistemic faults associated with stereotypes supported by available evidence, as well as cases where poorly supported stereotypes yield positive consequences;
> (b) the unification and simplification of the epistemic evaluation of acts of stereotyping;
> (c) intuitive appeal in post-theoretical contexts;
> (d) the capacity to explain how two individuals can hold the same stereotype yet differ in the epistemic standing they deserve due to differences in the dispositions associated with their belief;
> (e) a significant social role in articulating the epistemic wrongs that can be committed on a societal level;
> (f) the ability to distinguish morally objectionable stereotypes from those that are not.

Puddifoot's evaluative dispositionalism is also one of the points that have sparked controversies. Picinali criticizes this approach to assess stereotype beliefs as being poorly formulated to do the task for which it was formulated. On the other hand, Saul argues that Puddifoot's position has more radical consequences than she initially imagined: it is impossible to know all the dispositions associated with an act of stereotyping or to single out specific cognitions (such as particular stereotypes or individual acts of believing in a stereotype) for praise or blame. Saul therefore proposes that evaluative dispositionalism leads to the abandonment of the idea that we can successfully evaluate whether individual acts of stereotyping exhibit epistemic faults or merits and thus deserve praise or blame.

All the contributions in this symposium offer a rich debate on the already substantial material in Puddifoot's book, either by highlighting its

controversial points and potential problems or by building on her ideas. They are followed by a reply from Puddifoot herself, which clarifies many of the points raised by the commentators and deepens her ideas. I now invite readers to enjoy this book symposium, which is both engaging and informative in its own right, and which may inspire further exploration of Puddifoot's ideas in her book.

REFERENCES

Puddifoot, Katherine. 2021. *How Stereotypes Deceive Us*. Oxford: Oxford University Press.

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 1 | 9-24
https://doi.org/10.31820/ejap.21.1.1

ON THE DISCOMFORT WITH THE ACCURACY CRITERION: REASSESSING PUDDIFOOT'S CRITICISM AGAINST THE DUAL FACTOR VIEW

Federico José Arena[1,2]

[1] Universitat de Girona, Spain
[2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Discussion – Received: 01/07/2023 Accepted: 01/11/2023

This paper is part of a book symposium on Katherine Puddifoot's *How Stereotypes Deceive Us* guest edited by Marina Trakas (IFF, CONICET)

ABSTRACT

Given the pervasive presence of stereotypes' negative effects, there is a widely shared view according to which almost all stereotypes are harmful. However, some scholars have claimed that we should distinguish between stereotypes: those with and those without statistical support. In her book, Katherine Puddifoot claims that the statistical criterion falls short of what we need to develop a full theory of the epistemology of stereotyping, so she advocates for a Multifactorial view. While I share Puddifoot's discomfort with the accuracy criterion, in these comments I will put forward some critical considerations. Firstly, I will introduce a methodological concern regarding the discussion about the normative versus non-normative conception of stereotypes. Secondly, I will introduce some doubts about the extent to which the additional factors pointed out by Puddifoot's Multifactor view, are actually a challenge to the accuracy criterion. Thirdly, I will also critically comment on the argument that not having stereotypes or having egalitarian stereotypes that do not reflect some aspect of social reality improves our chances of attaining certain epistemic ends regarding the perceptions of individuals. Finally, I will briefly introduce a further factor of deception: the normativity of stereotypes, which is not considered within Puddifoot's proposal.

Keywords: conceptual analysis; accuracy criterion; generalizations; normative stereotypes.

Correspondence: fjarena@conicet.gov.ar

1. Introduction

Focus on stereotypes and their negative effects has increased in the legal domain in recent years. Although the battle against the discriminatory effects of stereotypes has a long history in some legal systems such as Sweden and the United States—where the term "stereotype" has appeared in legal debates and decisions since the 1960s [1] and the 1970s, [2] respectively—in other legal systems the term has only emerged very recently.[3] In the legal context stereotypes may produce their negative effects in several ways. First, they may hide behind norms that at a first sight might seem not only harmless but also beneficial for the stereotyped group. For instance, the norm that gives some financial aid to women who are the only caregiver of an elder member of her family. At first glance, the norm appears to introduce a benefit for women. However, it "hides the pervasive and mutually reinforcing stereotype that women are responsible for performing (unpaid) family care, and men are responsible for providing their families with financial support" (Franklin 2010, 139). Second, stereotypes may influence the assessment of evidence. For instance, they often play a significant role in how judges evaluate the credibility of victim testimony in cases of gender-based violence. Here I have in mind stereotypes such as "a victim of rape must try to resist and escape in all possible ways, but she is to be suspected of fabricating allegations of rape if she has the presence of mind to collect evidence after the assault" (CEDAW/C/82/D/148/2019) or "women accuse their partners of rape in order to obtain an economic advantage, such as keeping the family home" (Asensio et al. 2010, 87).

Given the pervasive negative effects of stereotypes, a widely shared view holds that almost all stereotypes are harmful and must be addressed using available legal instruments.[4] However, some scholars argue that we should draw on empirical sciences and use statistical advancements to distinguish between stereotypes with and without statistical support.[5] Although legal scholars seem to agree that the accuracy or statistical criterion improves our understanding of how stereotypes work, there is growing discomfort

[1] See, for instance, MYRDAL Y KLEIN (1956), a key text within the Swedish debate.
[2] See, for instance, Moritz v. Comm'r, 469 F.2d 466, 467 (10th Cir. 1972) and Weinberger v. Wiesenfeld, 420 U.S. 636, 641 n.7 (1975).
[3] In the American legal context that took place mostly after the ICHR decision on the "Campo algodonero" case. (Caso González y otras vs. México, 16/11/2009). In the European context the milestone is the ECHR decision on the Kiyutin case (ECHR, Kiyutin v. Rusia, no. 2700/10, 11/03/2011).
[4] See Franklin (2010) and Timmer (2011) advocating in favour of an anti-stereotype approach regarding stereotypes behind norms, and Di Corleto (2015) introducing the requirements of feminist epistemology regarding the assessment of evidence.
[5] See Jussim et al. (2009) for debates within psychology, and Schauer (2003) and Appiah (2000) for the legal and political domains.

regarding some of its applications (Arena 2022; Ghidoni and Morondo Taramundi 2022). The discomfort stems mostly from the awareness that there are wrongs brought about by stereotypes that are not captured by that criterion, such as stereotype threat, disregarding individuality, and reinforcing discriminatory practices.

In her book, Katherine Puddifoot (2021) provides additional arguments supporting the discomfort with the accuracy criterion. What sets Puddifoot's critique apart from previous analysis is her focus on the epistemic dimension of stereotypes and stereotyping (i.e., the application of a stereotype to an individual):[6] "the various ways that stereotyping makes people susceptible to making errors in their perception and judgment" (2021, 5). Puddifoot is interested in a specific kind of judgements, to wit, those about the traits of an individual: for instance, whether an individual Afro-American is a drug dealer or whether a particular woman is an expert in STEM disciplines, etc.

Puddifoot identifies two previous approaches to the epistemology of stereotyping: the Single Factor view and the Dual Factor view. Both theories address "whether the application of the stereotype increases or decreases the chance of an accurate judgement being made" (2021, 31). According to the Single Factor view, stereotyping produces an epistemic benefit when the applied stereotype reflects some aspect of social reality. In contrast, the Dual Factor view holds that two conditions must be met for an epistemic benefit to arise: the stereotype must accurately reflect reality by aligning with precise statistical information, and the act of stereotyping must be sensitive to unambiguous individual information. Puddifoot claims that both the Single and the Dual factor views fall short of what we need in order to develop a full theory of the epistemology of stereotyping. I will focus here on her criticisms against the Dual factor view.

Puddifoot presents her objections and her perspective in several steps. She begins by offering a pragmatic argument in favour of what she terms a non-normative conception of stereotypes, which holds that our definition of a stereotype should not include the condition that stereotypes are inaccurate or distorting. She then advocates for a Multifactorial view regarding the

[6] Even though Puddifoot defines the term "stereotyping" at the beginning of the book, an ambiguity in its usage remains (see, also, Saul in this issue of EuJAP). According to her definition, stereotyping consists in "the application of a social attitude that associates members of some social group more strongly than others with certain traits to an individual or individuals who are perceived as a member of the relevant social group, leading that individual or those individuals to be associated with the trait" (2021, 13). However, there are cases where the term refers to the expression of a social attitude (i.e., a stereotype) rather than its application to an individual. It seems to me that such instances include the Puddifoot's example of Trump (see 2021, 4, 20) and the discussion about the changes introduced by the Fair Sentencing Act of 2010 (see 2021, 47-48).

criteria for the epistemic evaluation of stereotyping. Subsequently, she claims, on the one hand, that even stereotypes with statistical support can cause epistemic harm and, on the other hand, that having false stereotypes (that do not reflect some aspect of social reality) may also be epistemically beneficial—that is, they may increase the likelihood of making an accurate individual judgement (see also Saul's discussion in this issue of EuJAP). While I share Puddifoot's discomfort with the accuracy criterion, in the remainder of the paper, I will present some critical considerations regarding certain steps in her arguments.

Firstly, I will introduce a methodological concern regarding the discussion of the normative versus non-normative conception of stereotypes. My main aim here is to clarify what appears to be a (fragment of a) conception of philosophical method underlying the way Puddifoot evaluates arguments put forward by other proponents of the non-normative view. Secondly, I will raise doubts about the extent to which the additional factors identified by Puddifoot's Multifactor view genuinely challenge the Dual Factor view. My central argument is that, while Puddifoot highlights important critical issues within the epistemology of stereotyping, the Dual Factor view provides the tools to address these concerns. Thirdly, I will critically examine the argument that not having stereotypes, or having egalitarian stereotypes that do not reflect some aspect of social reality, improves our chances of achieving certain epistemic goals regarding the perception of individual traits. I will argue that having egalitarian stereotypes would not eliminate the pitfalls of stereotyping; rather, what is needed is a form of epistemic sensitivity to new information. Finally, I will briefly introduce a hidden factor of deception: the normativity of stereotypes, i.e., the fact that some stereotypes are used to impose specific roles on members of a social group. This characteristic of stereotypes has been recognised by psychologists (e.g., APA 1991; Burgess & Borgida 1999) and political theorists (e.g., Appiah 2000). However, in my view, it has not received sufficient attention in the debate on the epistemology of stereotyping.

2. The philosophical status of the terminological move

There is a story often shared among Genoa University's PhD candidates about two analytic philosophers who were very good friends: one Argentinian and the other Italian. Both shared a love for Italian wine and food but disagreed on one issue—Italian bread. According to the Argentinian, Italian bread was simply bad. One day, the Italian philosopher, in a mildly nationalist gesture, offered the Argentinian a piece of bread without revealing its origin. The Argentinian tasted it and, satisfied, declared: "It's very good". Triumphantly, the Italian exclaimed:

"It's Italian!" To which the Argentinian wittily replied: "Then it's not bread".

Beyond the question of whether the story may be considered humorous, it serves to raise an important point: does a terminological move qualify as a philosophical move? This question is particularly relevant in the context of Puddifoot's approach to the debate between non-normative and normative conceptions of stereotypes. According to non-normative conceptions, stereotypes may be accurate, whereas normative conceptions hold that stereotypes are always inaccurate and distorting. While Puddifoot favours non-normative conceptions, she argues that the debate is ultimately a matter of framing or terminology:

> If the arguments of the current book were framed in a way that is consistent with the normative approach, they would still be important and interesting (…). I could be viewed as identifying the conditions under which the application of the social attitude amounts to stereotyping. (2021, 16)

From the way Puddifoot engages in the discussion between normative and non-normative conceptions of stereotypes, it appears that she treats the terminological move as a philosophical move. For example, she considers the strategy of pointing out all the accurate stereotypes to be a bad argument against the normative view:

> The problem with this argument is that it begs the question against the defender of the normative view. They could just reply that the beliefs that [the non-normative view] identifies, which are accurate social beliefs, are not stereotypes because stereotypes are, by definition, inaccurate. (2021, 17)

But is this terminological move truly a philosophical move—that is, one that contributes to our understanding of a concept? The answer to this question is undoubtedly tied to one's preferred approach to conceptual analysis. While this is not the place to fully address that issue, it seems to me that a terminological move does not constitute a method for advancing a conceptual proposal. The success of a conceptual proposal can be evaluated based on counterexamples, logical coherence, and ordinary language intuitions, among other criteria, but not on its ability to preclude the possibility of a terminological move.[7]

[7] Furthermore, as Puddifoot points out, one way to constrain a terminological move is by demonstrating that the counterpart has a specific goal and that making the terminological move would render the goal

For example, a first type of argument might consist in pointing out intuitive examples of the concept excluded by the proposed analysis. Jussim et al. (2009), proponents of the non-normative view, have highlighted certain social attitudes found to be accurate—such as sex distributions in various occupations or beliefs about demographic differences between African Americans and other Americans. Even though these examples fall outside of the scope of a normative concept of stereotypes, they are intuitive examples of stereotypes that have been even identified as such by advocates of the normative view.[8]

A second type of argument might involve considerations about how well a conception aligns with ordinary language usage. For example, the fact that people often use the term "stereotype" as an accusation is frequently cited as an ordinary language-based argument in defence of the normative view (Puddifoot 2021, 21). However, the accusation implied by the term "stereotype" is not always, nor necessarily, based on the inaccuracy of the attitude. At times, the accusation may instead arise from other failures, such as disregarding individual information, being overly generalised, or violating the duty to treat others as individuals. For instance, the statement "Women do not ride buses" is accurate in nearly all cities in Argentina and widely recognised as such. Nevertheless, the attitude expressed by this statement would still be labelled a stereotype. Thus, adopting the non-normative view does not rule out the possibility of the term retaining its negative expressive content

Even though I find these arguments convincing, the point here is that it is this type of argument, rather than a terminological move, that determines the outcome of a conceptual debate. Otherwise, if a terminological move were to count as a philosophical move, then pointing out the existence of stereotypes such as "African Americans are musical" or "Women are caregivers" would have to be regarded as a bad argument against the claim that only attitudes associating negative traits qualify as stereotypes. As the supposedly humorous story about Italian bread illustrates, in the context of conceptual analysis, a terminological move represents philosophical surrender. While it is always an available option, it amounts to raising the white flag.

unattainable. However, it is sufficient to have a different goal to justify making the terminological move regardless. This seems to apply to the pragmatic considerations Puddifoot advances as a stronger strategy against the normative conception. Even the reasons provided by Ashmore and Del Boca in favour of a non-normative view, as cited by Puddifoot, leave the terminological choice underdetermined. That is, one could accept those reasons and still proceed with the terminological move.

[8] Afro-Americans as "occupationaly unstable", see Allport (1954, 197).

3. The scope of the Dual Factor View

One key point in Puddifoot's book is the criticism of two principled approaches to stereotyping that she labels the Single and the Dual factor views. Principled approaches purport to identify in which cases stereotyping may have epistemic quality, in the sense of increasing our chance of making an accurate judgement about the traits of an individual. In that context, Puddifoot distinguishes between a stereotype that reflects some aspect of social reality and a stereotype that fits accurate statistical information. An example of the distinction may be found in the difference between the stereotype "women are not good at math" and the stereotype "Chilean young women get lower results than Chilean young men in math exams for University admission" (see del Río et al., 2016). Given the accuracy of the second stereotype, the first one reflects some aspect of social reality even if itself is not accurate. On this basis she distinguishes between the Single and the Dual factor view. The Single factor view claims that the only factor that determines the epistemic quality of an act of stereotyping "is the extent to which the stereotype reflects social reality" (32). The Dual factor view claims that the epistemic quality of an act of stereotyping is determined by two factors: "the accuracy of the stereotype and the nature of the available evidence about the individual (whether or not it is high quality and unambiguous)" (32). Puddifoot does not purport to deny the importance of the factors identified by these two approaches. On the contrary, she challenges both approaches by showing that there are further factors that determine the epistemic quality of an act of stereotyping. These factors are: (i) Stereotype Accuracy; (ii) Stereotype Relevance; and (iii) Response to Case-Specific Information.

The first factor, as Puddifoot notes, provides a challenge only to the Single factor view. There are stereotypes that, even if they reflect some aspect of social reality, given that they do not fit statistical information, may lead a person to form a distorted judgement of an individual. The stereotype about the relationship between women and math in Chile is a good example of that risk. I agree with this point; my concern here will be with the two remaining factors as they purport to represent a challenge to the Dual factor view. The point is not to deny the importance of these further factors, but to comment on their value as a criticism against the Dual Factor view.

The second key factor when assessing the epistemic quality of stereotyping is *relevance*: "individuals are not appropriately sensitive to contexts in which any statistical information that might be encoded in a stereotype is relevant to the judgement" (46). Puddifoot introduces a psychological explanation of this lack of sensitivity linked to the way in which stereotyping is triggered: wounded ego and the desire to justify the current

social system. Given these psychological causes of stereotyping, it is often the case that bearers of stereotypes would apply them even when irrelevant. And, when the stereotype is irrelevant, its application does not increase the chance of an accurate judgement being made.

The first difficulty in assessing this challenge to the Dual Factor view is that some examples (the black man who committed a minor traffic violation and a black person treated as threatening (46-47)) are based on inaccurate stereotypes. So, those cases show indeed that relevance is important, but do not challenge the Dual Factor view. The second, and key, difficulty, is that there is only a generic definition of "irrelevance" as "influenc[ing] judgements about other aspects of social reality" but from the examples there seem to be several possible interpretations of the meaning of relevance.

First, relevance may be interpreted as "the individual is within the scope of the stereotype". This seems to be a possible interpretation of relevance in the case of the black man who committed a minor traffic violation and received the application of the stereotype associating Black people with crime and the case of the woman who entered a university to work in STEM disciplines and received the application of the stereotype associating men more strongly than women with scientific expertise. In both cases, it cannot be claimed that the stereotypes are irrelevant in the sense of influencing judgements about other aspects of social reality because in both cases the judgements are about the aspect to which the stereotype refers: crime and scientific expertise. The irrelevance there seems to stem from the fact that both individuals are not within the scope of the more accurate formulation of the stereotype—e.g., "Black people *under certain contextual conditions* are prone to certain kinds of crimes" or "Women *with specific characteristics* have less scientific expertise than men".

Second, relevance may also be interpreted as "non-spurious statistical correlation". It seems to be a good interpretation of relevance in the case of stereotypes regarding Black people or career women. In the first case, stereotypes of Black people while slavery was legal in the United States, such as that they are happy, childlike, and affectionate, are irrelevant in order to associate one of those traits with an individual Black person because they express a spurious correlation. A similar interpretation of relevance may follow from the examples of stereotypes about career women as lacking creativity given their monolithic lifestyle.

Finally, relevance may even be interpreted as "argumentative force". It seems to be a good interpretation of irrelevance in the case of the debate

on the changes to the US criminal justice system regarding the punishment attached to the possession of crack and powder cocaine. Within this debate, the stereotype strongly associating Black people with crime is irrelevant because it has no argumentative force regarding the question "whether specific sentences are appropriate for particular crimes" (48).[9]

The important point here is that the Dual factor view would agree with the claim that an act of stereotyping has epistemic quality (produces epistemic benefits) when it is applied to an individual that belongs to the group associated by the attitude with certain trait and when that association expresses a non-spurious relationship, i.e., when it has statistical basis. The reason for that agreement is that both requirements follow from the condition of accuracy. The stereotype according to which Chilean women got lower results than men in math exams for the admission to universities has statistical basis and will produce epistemic benefits only if it is applied to a woman that fits the description: a Chilean woman that has taken the math exam.

The last factor is *response to case-specific information*. The point here is "whether or not the application of the stereotype leads diagnostic case-specific information to become inaccessible or distorted" (48). Puddifoot introduces several findings in psychology showing that stereotypes may distort case-specific information in different ways that leads to misperception of the individual traits:

> They may lead (1) ambiguous evidence to be misinterpreted as fitting with a stereotype; (ii) information to be remembered in a distorted manner; (iii) details about individuals to be missed due to assumptions about the similarities within and dissimilarities between groups; (iv) false explanations to be developed; (v) testimonial injustice and testimonial smothering to occur; (vi) inaccurate associations to be made or accurate associations not to be made. (2021, 73)

I agree with Puddifoot to the extent that these factors illustrate the complexity of the task of gathering individual information and assessing the traits of an individual. However, given the emphasis that the Dual factor view puts on case-specific information, it seems to me that many of the cases of stereotyping pointed out by the examples would be considered by the Dual factor view as lacking epistemic quality, given that the ambiguity of individual information was not adequately ruled out. Take for instance the example of the Black female barrister who suffers the

[9] Note, however, that in this last case we are not dealing with irrelevance in the context of stereotyping.

application of the stereotype associating barristers with being male and White. This stereotype, while being accurate (according to the example it fits statistical information), may produce epistemic costs given that people will fail to notice that a Black female may not share the characteristic attributed by the stereotype to the group and therefore will fail to notice that she is a barrister.

It may also produce epistemic costs given that people will tend to take as non-ambiguous ("she is a defendant") information ("a Black woman entering the court building or the courtroom") that is actually ambiguous ("she is either a defendant or a barrister"). However, from my point of view, this seems to be a case in which individual information was not taken into consideration in the right way. Remember that one of the conditions that, according to the Dual factor view, must be satisfied for the act of stereotyping to bring epistemic benefits is to take into consideration non-ambiguous individual information. This means that individual information must be carefully considered to rule out ambiguity before applying the stereotype. In this example, the problem was that the ambiguity of the information ("a Black woman entering the court building or the courtroom") was not ruled out before applying the stereotype. If this is correct, then the Dual Factor View provides the resources to determine that, in this case, the act of stereotyping did not bring about epistemic benefits.

4. On the epistemic quality of an egalitarian attitude

According to Puddifoot, stereotypes that reflect aspects of social reality may nevertheless produce epistemic costs due to the different ways, listed in the previous section, in which stereotypes may distort case-specific information. Given these effects, Puddifoot claims that "there can be epistemic benefits associated with lacking these beliefs or other social attitudes, and even from having false social beliefs or other social attitudes" (73). One kind of false social beliefs that may bring epistemic benefits is "egalitarian social attitudes", that is, social attitudes that represent the world as being egalitarian. To illustrate this point, Puddifoot introduces the example of Tim, who grew up in a society where men and women were equally well represented in science but now lives in one where they are not. Given the conditions in which he was raised, Tim holds an egalitarian attitude, believing that men and women are equally represented in science. While this belief was accurate in his home country, it does not reflect the reality of the society he currently lives in (in this case, the UK). According to Puddifoot, this attitude, which falsely represents the world as more egalitarian than it really is, may yield epistemic benefits—

most notably, the benefit of "avoiding the distortion of case-specific information" (76). For instance,

> [Tim] does not treat ambiguous behaviour of women scientists as demonstrating that they lack expertise, instead judging the behaviour as it should be, as ambiguous. He does not remember evidence suggesting that any woman scientist lacks expertise more strongly than evidence suggesting that a woman scientist has expertise (…). (2021, 76)

That is why "the possession of [egalitarian] attitudes can protect against various epistemic errors, e.g., misinterpreting ambiguous evidence, having distorted memories, and so on" (2021, 75).

Let's consider a further example to evaluate the argument. As mentioned above, in Chile the generalization that women get lower results than men in the math exams for admission to universities is an accurate generalization, i.e., it fits statistical information. Therefore, according to the Dual Factor View, applying this generalization to a specific female Chilean student who recently took the exam—associating her with the trait of having obtained a lower score than men—should increase our chances of making an accurate judgment about her individual traits. However, Puddifoot argues that this is not always the case, as even accurate stereotypes can produce effects that reduce our chances of making accurate judgments. For instance, suppose that after lunch with colleagues, the stereotyped student correctly splits the bill among all participants but then miscalculates the tip for the waiter. Due to the stereotype, we are more likely to remember the second event and not the first, leading to inaccurate judgments about her mathematical abilities. Now imagine a math professor who grew up in Argentina, where university admission does not require mandatory math exams. He holds egalitarian views about the distribution of math ability between men and women and later moves to Chile. According to Puddifoot, his egalitarian outlook would improve his chances of remembering all relevant information and forming accurate judgments about individuals, as he would not be influenced by the Chilean inegalitarian stereotype. And this would hold true even if, as in the example, the stereotype is statistically accurate.

However, is it necessarily the case that egalitarian attitudes bring epistemic benefits? Selectiveness is an inherent consequence of viewing circumstances through the lens of a generalization encoded in a stereotype. The use of a generalization creates a blind spot—namely, the facts excluded by the generalization. As Frederick Schauer puts it: "In focusing on a limited number of properties, a generalization simultaneously

suppresses others, including those marking real differences among the particulars treated as similar by the selected properties" (Schauer 1991, 21–22). If selectiveness results from applying a generalization, then adopting a different or opposing generalization merely introduces another blind spot. Holding an *inaccurate* egalitarian attitude about the distribution of mathematical abilities may lead the math professor to overlook information that aligns with the *accurate* inegalitarian stereotype. For instance, he may be condescending toward a student with weak math skills or, as Puddifoot suggests, may forget the tip miscalculation incident. The point here is that the distortion of case-specific information is a consequence of basing the assessment of individual traits on a generalization, regardless of whether it has egalitarian or inegalitarian content.

Moreover, an individual with an inegalitarian attitude—that reflects the reality of the society where she was raised but inaccurate in her current society—may actually avoid distortions caused by an *accurate* egalitarian attitude that misrepresents case-specific *inegalitarian* information. Consider a Chilean-born and raised math professor who moves to Argentina. Her inegalitarian attitude accurately reflects her upbringing in Chile but is inaccurate in Argentina. However, if she encounters a female student with weak math abilities, an egalitarian attitude might distort her perception of the student, decreasing the likelihood of making a correct judgment. A similar issue arises, following Puddifoot's example, with a UK-born physicist who travels to Tim's home country and meets a female colleague with poor expertise.

It is certainly true that an egalitarian attitude can lead to better interpersonal interactions, mitigating the negative effects of discrimination. However, this does not necessarily mean that having egalitarian attitudes improves our chances of making accurate judgments about an individual's traits. A person with an egalitarian bias may make the same errors as someone whose attitude reflects social reality, even if that reality is inegalitarian. The risks of ignoring relevant inegalitarian information, misinterpreting inegalitarian facts as unambiguously egalitarian, or forming distorted egalitarian memories of inegalitarian events remain present.

5. A hidden factor: Stereotype normativity

In the case of Price Waterhouse v. Hopkins, the American Psychological Association's amicus curiae noted the importance of distinguishing between descriptive and normative stereotypes about women. The authors of the amicus curiae claimed that: "descriptive stereotypes characterize

women in a way that undermines their competences and effectiveness; normative stereotypes label women whose behaviour is inappropriately masculine as deviant" (APA 1991, 16). After a decade of further research, this distinction has been generalized in order to distinguish between a descriptive and a prescriptive component of stereotypes:

> [T]he descriptive component of gender stereotypes consists of beliefs about the characteristics that women do possess, whereas the prescriptive component consists of beliefs about the characteristics that women should possess. (Burgess and Borgida 1999, 665-666)

When used descriptively, a stereotype aims to provide information about the world—it seeks to describe a state of affairs, specifically the traits of a group. Its direction of fit is stereotype-to-social-group. Therefore, descriptive stereotypes can be evaluated based on their accuracy—whether they accurately reflect aspects of social reality.

On the contrary, when used normatively, stereotypes seek to impose certain roles on members of a social group. In this case, the stereotype follows a social-group-to-stereotype direction of fit. This means that, for those who uphold it, a mismatch between the world and the stereotype is a reason to change the world rather than the stereotype itself. Since normative stereotypes do not claim to describe reality, they are not subject to empirical evaluation. Thus, in principle, it does not make sense to ask whether normative stereotypes are accurate or reflect social reality. This implies that an approach to stereotypes based solely on accuracy will overlook their normative dimension.

However, normative stereotypes can also have epistemic consequences, as norms about the behavior of certain social groups may distort perceptions of individual group members. The most dangerous effect is the inversion of the purpose of the epistemic endeavour, in the sense that the bearer of a normative stereotype may end up abandoning epistemic goals, limiting himself to disapproving of the behaviour of those who do not conform to the normative stereotype and constructing the facts in such a way as to make it possible to inflict some type of punishment. For example, someone who upholds the *decent woman* stereotype may, when faced with a victim who does not fit this ideal—i.e., a woman who does not behave according to its expectations—deny her the status of victim and dismiss the possibility of a crime, not based on evidence but as an expression of moral disapproval. In such cases, by transforming an epistemic inquiry into a normative judgment, the stereotype reduces the likelihood of an accurate

assessment. A full theory of the epistemology of stereotyping should then take into consideration this further factor of distortion.

6. Summing up

In my opinion Puddifoot's book is an essential reading for any legal scholar seeking a serious approach to stereotypes. These brief comments have focused only on the first chapters, but in the subsequent sections, Puddifoot develops further insights and arguments on the epistemology of stereotyping that also merit detailed discussion. The structure of my arguments here reflects my agreement with most of Puddifoot's conclusions. I share her non-normative conception of stereotypes, even though I do not find her objections to alternative arguments for that same conception convincing. Like Puddifoot, I consider statistical accuracy a necessary—though not sufficient—criterion for assessing both the epistemic import of stereotypes and the epistemic quality of stereotyping.

Reading Puddifoot's book, I have learned a great deal about how stereotypes can hinder epistemic inquiry due to various psychological factors that may lead their bearers to make incorrect judgments. My only disagreement concerns whether the existence of these factors constitutes an objection to the Dual Factor View. Relevance and the quality of individual or case-specific information are necessary criteria for assessing the epistemic import of stereotypes, but they are merely implications of the conditions set forth by the Dual Factor View.

The only fundamental disagreement concerns the benefits of egalitarian attitudes. As I have argued, some pitfalls of inegalitarian stereotypes stem from the fact that stereotypes function as generalizations. Since egalitarian attitudes are also generalizations, they may incur the same epistemic costs. Avoiding these costs requires a form of epistemic sensitivity—being open-minded and willing to revise one's beliefs when reality proves them inaccurate.

Finally, I have introduced what appears to be an additional factor affecting stereotyping and contributing to epistemic costs: the normativity of stereotypes. This presents a challenge to the Dual Factor View, at least insofar as the accuracy criterion becomes ineffective in assessing the epistemic risks posed by stereotypes that do not aim to describe reality but rather to regulate the behavior of the target group.

Acknowledgments

Work on this paper has been supported by grant CEX2021-001169-M, funded by MICIU/AEI/10.13039/501100011033.

REFERENCES

Allport, Gordon Willian. 1954. *The Nature of Prejudice*. Reading: Addison-Wesley.

American Psychological Association. 1991. *In the Supreme Court of the United States: Price Waterhouse v. Ann B. Hopkins. Amicus Curiae Brief for the American Psychological Association. American Psychologist* 46: 1061–1070.

Appiah, Kwame Anthony. 2000. "Stereotypes and the Shaping of Identity." *California Law Review* 88 (1): 41-53. https://doi.org/10.2307/3481272

Arena, Federico José. 2022. *Los estereotipos detrás de las normas*. Córdoba: Toledo.

Asensio, Raquel, Julieta Di Corleto, Valeria Picco, Leah Tandeter, and Magdalena Zold. 2010. *Discriminación de género en las decisiones judiciales: justicia penal y violencia de género*. Buenos Aires: Defensoría General de la Nación.

Burgess, Diana, and Eugene Borgida. 1999: "Who Women Are, Who Women Should Be. Descriptive and Prescriptive Gender Stereotyping in Sex Discrimination." *Psychology, Public Policy, and Law* 5 (3): 665-692. http://doi.org/10.1037/1076-8971.5.3.665

Del Río, María Francisca, Katherine Strasser, and Maria Inés Susperreguy. 2016. "¿Son las habilidades matemáticas un asunto de género? Los estereotipos de género acerca de las matemáticas en niños y niñas de kínder, sus familias y educadoras." *Calidad en la educación* 45: 20-53. http://doi.org/10.4067/S0718-45652016000200002

Di Corleto, Julieta. 2015. " La valoración de la prueba en casos de violencia de género." In *Garantías Constitucionales en el Proceso Penal*, edited by Florencia Plazas and Luciano Hazan, 589-606. Buenos Aires: Editores del Puerto.

Franklin, Cary. 2010. "The Anti-Stereotyping Principle in Constitutional Sex Discrimination Law." *New York University Law Review* 85: 83-173. https://ssrn.com/abstract=1589754

Ghidoni, Elena, and Dolores Morondo Taramundi. 2022. "El papel de los estereotipos en las formas de la desigualdad compleja: algunos apuntes desde la teoría feminista del derecho

antidiscriminatorio." *Discusiones* 28 (1), 37-70. https://doi.org/10.52292/j.dsc.2022.2325

Jussim, Lee, Thomas R. Cain, Jarret T. Crawford, Kent Harber, and Florette Cohen. 2009. "The Unbearable Accuracy of Stereotypes." In Handbook of *Prejudice, Stereotyping and Discrimination*, edited by Tood D. Nelson, 199-227. New York: Psychology Press - Taylor and Francis.

Myrdal, Alva, and Viola Klein. 1956. *Women's Two Roles. Home and Work*. London: Routledge and Kegan Paul.

Puddifoot, Katherine. 2021. *How Stereotypes Deceive Us*. Oxford: Oxford University Press.

Schauer, Frederick. 1991. *Playing by the Rules: A Philosophical Examination of Rule-based Decision-making in Law and in Life*. Oxford: Clarendon Press.

Schauer, Frederick. 2003. *Profiles, Probabilities and Stereotypes*. Cambridge, Mass.: Harvard University Press.

Timmer, Alexandra. 2011. "Toward an Anti-Stereotyping Approach for the European Court of Human Rights." *Human Rights Law Review* 11 (4): 707-738. https://doi.org/10.1093/hrlr/ngr036

References to legal documents

Committee on the Elimination of Discrimination against Women, General Recommendation, CEDAW/C/82/D/148/201

European Court of Human Rights, Kiyutin v. Rusia, no. 2700/10, 11/03/2011.

Interamerican Court of Human Rights, Caso González y otras ("Campo algodonero") vs. México, Sentencia de 16 de noviembre de 2009. Serie C No. 205.

Moritz v. Comm'r, 469 F.2d 466, 467 (10th Cir. 1972)

Weinberger v. Wiesenfeld, 420 U.S. 636, 641 n.7 (19

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 1 | 25-38
https://doi.org/10.31820/ejap.21.1.2

ACCURATE STEREOTYPES AND TESTIMONIAL INJUSTICE

Leonie Smith[1]

[1] Lancaster University, United Kingdom

Discussion – Received: 01/04/2024 Accepted: 01/06/2024

This paper is part of a book symposium on Katherine Puddifoot's *How Stereotypes Deceive Us* guest edited by Marina Trakas (IFF, CONICET)

ABSTRACT

In *How Stereotypes Deceive Us*, Katherine Puddifoot provides a convincing non-normative account of what stereotypes are, and of the conditions under which we appropriately rely on them in achieving our epistemic and ethical goals. In this paper, I focus on Puddifoot's discussion of what she takes to be the non-prejudicial use of accurate stereotypes and their role in causing or perpetuating harm. Such use can cause harm but does not, on the face of it, appear to be wrongful in the way that ordinary cases of prejudicially motivated use of stereotypes are. This raises a challenge for identifying when our use of such stereotypes might be unjust or wrongful (and why). In response, I first suggest that prejudice might be located within the context in which one uses a stereotype, rather than within the content of the stereotype itself. In this way, we can indeed distinguish prejudicial (and therefore wrongful) use of accurate stereotypes from non-prejudicial (innocent) use of accurate stereotypes. And second, I suggest that we also ought to question whether the stereotypes being invoked in all cases really are accurate, given the context and scope of application.

Keywords: testimonial injustice; stereotypes; context; epistemic injustice.

Correspondence: leonie.smith@lancaster.ac.uk

1. Introduction

In *How Stereotypes Deceive Us*, Katherine Puddifoot has produced a convincing non-normative account of what stereotypes are, and of the complex relationship between epistemic and ethical goals, and the conditions under which we appropriately rely on stereotypes in achieving these ends. I found myself agreeing with much of the book. What follows are therefore not substantial worries for Puddifoot's account, but rather just a few points to consider with regard to how Puddifoot's analysis and cases fit with other cases in the literature on epistemic ethics, specifically, with cases of testimonial injustice. I hope this will serve to build on the invaluable work that Puddifoot has contributed to our understanding, and perhaps offer some alternative ways to think about cases of testimonial injustice more generally.

2. Testimonial injustice, reliability and faulty stereotypes

A person who is marginalised in a society can fail to be treated fairly as an epistemic agent, through the unjust epistemic behaviours of those who she interacts with. One way in which this can happen is, as Puddifoot identifies, through testimonial injustice (Fricker 2007). Someone who experiences testimonial injustice is "wronged specifically in her capacity as a knower" (Fricker 2007, 20) when she attempts to take part in a testimonial exchange, or when she offers up her own knowledge. In central cases, on Fricker's account, the hearer fails to take up, believe, or otherwise appropriately epistemically process the testimony of the speaker, when she otherwise ought to, because of her (implicit or otherwise) acceptance of an identity-prejudicial stereotype which "embodies a prejudice that works against the speaker" (Fricker 2007, 17). The speaker's testimony is treated as less credible than it ought to be by the hearer. The hearer may also distort the speaker's meaning, attach unintended meaning to the speaker's words, or even attach meaning to the fact of that person speaking at all, which is not justified by the actual content or context of the testimony provided.

Fricker provides a case from literary fiction (the screenplay of "The Talented Mr Ripley"), which she takes to be a paradigm example of systematic, prejudicial identity-based testimonial injustice. A woman (Marge Sherwood) has her credible—and importantly valuable—evidence downplayed as "feminine intuition", by a male acquaintance (Herbert Greenleaf), who otherwise holds her in warm personal regard (Fricker 2007, 14). Sherwood experiences a kind of subconscious downgrading of her testimony, and her attempts to provide it to Greenleaf, due to prejudices held towards people with her social identity as a woman within that society,

at that point in time. This can be observed in the following interaction, suggests Fricker:

> GREENLEAF, speaking to Tom Ridley: "This theory, the letter he left for you, the police think that's a clear indication he was planning on doing something (…) to himself."
> SHERWOOD: "I just don't believe that!"
> GREENLEAF: "You don't want to, dear. I'd like to talk to Tom alone—perhaps this afternoon? Would you mind? Marge, what a man may say to his sweetheart and what he'll admit to another fellow-"
> (Minghella, quoted in Fricker 2007, 87).

Despite his desperate need for information relating to his missing son, Dickie, Greenleaf "fails to see Marge as the source of knowledge about Dickie that she manifestly is" (Fricker 2007, 88). Greenleaf's uptake of negatively-valenced prejudicial stereotypes about the rational capabilities of women cause him to downgrade Sherwood's testimony to that of a hysterical woman, too delicate to understand her fiancé's infidelities and mind, whose testimony can be dismissed as that of someone who simply wants to believe the best of her partner. This is summed up in the line: "Marge, there's female intuition, and then there are facts" (Minghella, quoted in Fricker 2007, 14). In the novel and screenplay, however, we discover that Sherwood is correct and that Ripley has murdered Marge's fiancé.

There are a number of key features in this account of paradigm cases of testimonial injustice:[1]

(i) The speaker is a member of a group about which there are widely held stereotypes, associations between that group and particular attributes (Fricker 2007, 30), in her society.
(ii) These stereotypes are negatively valenced: they ascribe traits to the group which are derogatory within the specific context.
(iii) These stereotypes "embody" prejudice: they are resistant "*to counter-evidence owing to some affective investment on the part of the subject*" (Fricker 2007, 35, author's italics).
(iv) The hearer is influenced, knowingly or otherwise, by these stereotypes.[2]

[1] Note that Fricker does allow for various exceptions and non-central cases, including "incidental" testimonial injustice which can be ethically serious and practically damaging but which "does not render the subject vulnerable to any other kinds of injustice (legal, economic, political)" (2007, 27).
[2] Whether this counts as "acceptance" or "take up" of the stereotype is contested.

(v) This leads to the hearer's epistemically culpable mistreatment of the speaker's testimony in a specific interaction.[3]

All of this suggests that there is something *wrong* with the stereotypes in play in cases of testimonial injustice. The reason that these stereotypes ought not to influence the hearer's assessment of the speaker is that they are faulty or inaccurate; they are prejudicial and "resistant to evidence". The use of these *faulty* stereotypes therefore makes the hearer epistemically and ethically culpable for their failure to treat the speaker appropriately in a testimonial interaction.[4]

3. The problem of *accurate* stereotypes

3.1 An important insight

However, Puddifoot draws our attention to very important cases which appear to mirror the epistemic and ethical harms of testimonial injustice, despite the hearer acting in ways which do not, on the face of it, appear to be epistemically or ethically suspect in the way that paradigm cases of testimonial injustice suggest. These are cases in which *accurate* stereotypes are in play but someone is still treated unjustly in terms of their testimony.

For example, it may be the case that in wider society, women are less likely to have scientific expertise than men are. This may be due to any number of reasons, including a lack of access to opportunities, or the discriminatory treatment of women in educational settings within that society. Nevertheless, the stereotype that "women are less likely to have scientific expertise than men", is *accurate* for that society. When a woman who works in STEM then attempts to provide her expert insight, her credibility might be unfairly downgraded because of the holding and application of this stereotype by her audience. Viewed as a woman, and therefore as someone who is less likely to have scientific expertise than her male colleagues, she might be ignored and overlooked in discussions with her colleagues, or in providing information to a wider public.

[3] Again, Fricker does allow that there might be a form of cumulative testimonial injustice, whose effects are felt only in the impact of multiple interactions over time, not in any one testimonial interaction (Fricker 2007, 20-21), but these are not the central cases (or, relevant to this paper, equivalent to the cases Puddifoot examines).

[4] Note that Fricker does not argue that all stereotypes are misleading (Fricker 2007, 15, 30). But in these original paradigm cases of testimonial injustice, the key is that the stereotype must be *prejudicial* and derogative, implying that it is an unfair or misrepresentative stereotype.

In this case, the stereotype that "women are less likely to have scientific expertise than men" is one which (a) accurately "reflects reality", and (b) appears to have been gained in an epistemically appropriate way, given that it is *responsive to evidence* about the number of women vs men who have scientific expertise: for the audience, the stereotype is "a fitting response to facts found in their social environment" (Puddifoot 2021, 68). Puddifoot therefore suggests, on my reading of her account, that the use of the stereotype does not embody prejudice of the kind found in Frickerian testimonial injustice (or perhaps that the use of the stereotype is not prejudicially motivated).

Puddifoot's key claim then, is that the application of a stereotype which *accurately* reflects "some aspect of social reality" "can (…) lead to the undue dismissal of the testimony of members of some social groups" (Puddifoot 2021, 190). The non-prejudicial use of accurate stereotypes mean that these types do not fit well under Fricker's account of paradigm cases of testimonial injustice, in which our prejudicial attitudes are at the root of the wrongdoing, and in which we can be held epistemically culpable and blameworthy for the outcomes because of this. The speaker experiences the epistemic harm of not being treated fairly as a knower (with the risk of subsequent harms to her career, and perhaps even to how she herself views her own epistemic capabilities, with further knock-on effects to her attempts to contribute to science). But despite mirroring the epistemic harms found in Frickerian cases (and perhaps the sense that something has still gone "wrong" in how the individual has been epistemically treated), prejudice does not seem to be at the root of the problem.

Puddifoot's identification of cases such as this is a hugely important contribution to understanding the landscape of epistemic harm. It seems vital, for anyone concerned with epistemic justice or the avoidance of epistemic harm in structural oppression, that we recognise that in many ordinary cases of testimonial interaction the mechanism by which our epistemic judgements can cause harm is not immediately obvious through the presence of false (and either prejudicial or prejudicially motivated) identity-based beliefs.

3.2 A worry for ethical assessment

One upshot of identifying these types of cases, however, is that it becomes harder to recognise whether the use of stereotypes as heuristics will lead, or have led, to epistemic harm in our testimonial interactions. This creates a challenge for evaluating the ethical behaviour of hearers who rely on (accurate) stereotypes in situations which result in epistemic harm for the

speaker. If it is indeed the case that accurate and fairly formed stereotypes can cause us to epistemically harm speakers on some occasions, while not doing so on others, it becomes difficult to argue that an *ethical* harm has been perpetrated by the hearer, despite there being someone who appears to be on the receiving end of identity-based, systematic mistreatment of her testimony.

Puddifoot proposes that we rely on evaluative dispositionalism to determine whether or not a person ought to adopt and use a heuristic or not. This is a process in which we need to understand and evaluate the complicated and wide-ranging dispositions of the actors involved taken as a whole, in order to determine whether a person has acted, or will act, rightly or wrongly in relying on a particular stereotype in any given context or situation.

There may be worries around how demanding this process would be. However, a more pressing worry for many might be that, even when applied well, this process will lead to counter-intuitive results. Specifically: that the weighing up of the balance of all possible upstream and downstream dispositions of the speaker—the dispositions displayed in forming beliefs, and the dispositions possessed as a result of holding those beliefs—might in some instances suggest that the hearer *did* act in a reasonable or fair way in adopting and relying on a stereotype, even though the speaker experienced epistemic harm in her testimony not being taken up when it was in fact credible.

And if this is the case, then this would leave no real recourse for identifying an ethical, epistemically culpable, wrongdoing or wrongdoer, despite the presence of an epistemically unfair outcome for the speaker. This might appear to let those who downgrade credibility in this way "off the hook" for what, intuitively, still seems in some way to be bad behaviour, in and of itself. Or at the very least, to suggest that there is no ethical injustice for someone who experiences testimonial credibility downgrading under these circumstances.

We could, of course, bite the bullet on this, and accept that there was no ethical wrong. Alternatively, it does in principle remain possible that an exhaustive and detailed analysis of the dispositions involved in every case we can think of might in fact turn up that there are *not* cases in which there is *no* residual ethical wrong in the adoption and application of an accurate stereotype in scenarios which lead to testimonial harm.

However, biting the bullet requires giving up an important ethical stance many might want to take, and completing a detailed analysis of every

possible case that initially seems problematic is something that arguably falls beyond the bounds of a reasonable expectation for our ethical judgements (and interpretation of right action for ourselves and others in practice).

I wonder, therefore, whether the worry around ethical culpability for epistemic harms in some of the cases Puddifoot identifies might justify revisiting whether or not these types of cases are distinct from "ordinary" testimonial injustice—in which we can identify a wrongdoer and wrong action—after all.

A first thought in this direction might question whether or not it is right to say that Puddifoot-style cases do not, in fact, involve the operation of prejudice, construed in some suitable manner. If this is the case, then this alone would be enough to categorise them as "ordinary" testimonial injustices involving clear wrong action (albeit not necessarily ones that Fricker's original 2007 account accommodates).

A second thought relates to the question of what makes a stereotype accurate or otherwise in the first place. Is it in fact the case that a stereotype reflecting "some aspect of social reality" and being mis-applied in a context, is in itself enough to demonstrate that testimonial injustice comes about as a result of an *accurate*, yet irrelevant, stereotype? It is the fact that these stereotypes appear to be accurate that creates the challenge of the missing ethically wrong action. If the stereotypes turn out to not reasonably be considered accurate after all then, again, this problem goes away, and these apparently different cases may not actually differ from what we might think of as central cases of Frickerian testimonial injustice, where the wrongdoing is connected to the use of *faulty* stereotypes, ones which are resistant to counter-evidence that the hearer is epistemically responsible for not recognising.

I expand on these thoughts in turn in sections 4 and 5 below.

4. Prejudice: "Resistance to counter evidence" in stereotype content vs context

> *When credibility deficits of this type occur, there will at least sometimes be an absence of prejudice, so the effect is not a case of testimonial injustice, as defined by Fricker, because this occurs due to the operation of prejudice*. (Puddifoot 2021, 190)

Testimonial injustice relies on the use of stereotypes which embody prejudice. As such, the first thing we might question is whether, in cases such as the woman in STEM who is not treated with due credibility by her colleagues or the public, there really is no prejudicial content either in the stereotype, or its use.

Applying a stereotype inappropriately without appropriate contextual boundaries, as a result of some "ethically bad affective investment" on the part of the hearer (Fricker 2007, 35), involves prejudicial treatment. For Fricker, it seems that the specific stereotype must *itself* embody a negative association with members of the group it targets, which is resistant to counter-evidence (ibid., 35). As such, for any case of testimonial harm in which the stereotype is accurate, given the available evidence, and in which the individual hearer would update their stereotype should the underlying facts change, it seems prejudice has not played a part in the hearer's actions, where their downgrading of a speaker's credibility is due to their belief in an accurate stereotype.

However, it is not clear to me that testimonial injustice *ought* to only involve instances in which the resistance to counter evidence is within the stereotype itself, rather than *within the context in which one uses the stereotype*. Prejudice can enter into our actions at alternative points, related to the use of the stereotype within a given context, rather than the content of the stereotype itself.

Puddifoot highlights that stereotypes are (regularly) applied in contexts to which they are not relevant. As she notes: "A stereotype can reflect some aspect of social reality but influence judgements about other aspects of social reality" (Puddifoot 2021, 62). In this situation, a person acquires a stereotype which is responsive to evidence in their wider environment. However, they go on to act in ways which generate testimonial harm through applying that stereotype in irrelevant contexts. In the case of women in STEM being downgraded in testimonial interactions with their colleagues:

> Women who enter universities and workplaces to work in STEM subjects are stereotyped as lacking expertise. But the stereotype is not relevant in these cases, in which women have (…) gained scientific expertise. (Puddifoot 2021, 63)

In this case we might say then that someone has made a prejudgement without proper regard to the evidence about a more relevant group: that of "women who work in STEM".

The question we should ask is whether this prejudgement itself was due to prejudice. If prejudice is resistant to counter-evidence, due to "affective investments", then in the case of women in STEM, it may well be the case that while the initial stereotype is accurate, the failure to consider further relevant readily-available contextual evidence (that the specific woman is a STEM professional and must therefore have scientific expertise), comes about due to the same prejudicial views about women that make it the case that in the society one lives in, fewer women than men have received a STEM education or range of career opportunities. The very fact that the stereotype is true is indicative of prejudice in the background conditions of society. The willingness to ignore further evidence (including evidence that the stereotype may not apply in this case) in the presence of that stereotype might very likely be due to the same prejudice towards that group of persons.

Consider the following two scenarios by way of illustration of the difference between prejudicial and non-prejudicial use of accurate stereotypes:

CROWDSOURCING

> Jonno the journalist needs to quickly get a few fairly innocuous anonymous reactions, regarding Manchester City's victory at the Etihad Stadium earlier that day, to include in a quick summary in a local news report going out live in fifteen minute's time. There is a good chance that many City fans are still in the city centre, celebrating the win. Jonno has limited time so, relying on the stereotype that "men are more likely than women to attend football matches", they approach a group of men sat outside the pub, rather than a group of women who are sat outside a different pub over the road, to try and find a few supporters to speak to. Had a woman approached Jonno proactively, they would have happily asked her for a quote. But given time, Jonno's strategy is to just maximise their chance of getting a few good soundbites in hand in time for the live transmission.

In this case:

(i) The stereotype "men are more likely than women to attend football matches" was accurate.
(ii) The stereotype "men are more likely than women to attend football matches" was relevant to the scenario (trying to

maximise the chance of a good quote in a timely fashion in Manchester centre).

(iii) No epistemic injustice occurred[5] because (a) nobody was wronged in their capacity as a knower; (b) there was no epistemic cost to Jonno or their audience in Jonno's focus on men; (c) there was no prejudicial aspect to Jonno's behaviour (they would have responded quickly to counter evidence from any woman who spoke to them).

POST-MATCH PUB CHAT

> Elsie, a City fan wearing a City shirt, is sat in the pub after attending that same City match earlier that day. The rest of the crowd around her are men. Everyone is joining in with a large discussion, dissecting the match, the strengths of players on both sides, and the actions of the referee, in great detail. Elsie tries to explain why she thought the ref made a bad call early in the game, but when she speaks, she is ignored. A few minutes later, a man in the pub makes the same point and everyone joins in with their views in response to him.

In this second case:

(i) The stereotype "men are more likely than women to attend football matches" was accurate.

(ii) However, the stereotype "men are more likely than women to attend football matches" was irrelevant in the given context of listening to Elsie. Additional content—that the specific woman present, Elsie, was known to have been at the football match—over-ruled any reasonable exclusion of Elsie from the discussion about what happened in the game on the grounds of rules about the general population.

(iii) Epistemic harm occurred because: (a) Elsie was wronged in her capacity as a knower; and (b) there was an epistemic cost to the group who failed to hear Elsie's insights.

Was there also (c), a prejudicial aspect to the group's behaviour, which would make this testimonial injustice?

[5] Let's stipulate that the quotes Jonno was looking to include were quite banal—"the atmosphere was electric", "good to see City finally get a win in"—and that there is no good reason to suppose that women City fans would have provided any significantly different comments.

Arguably, yes. The prejudice may have entered in through background prejudicial beliefs and acts which lead to fewer women watching football matches in the first place, and the other members of the group having poor responsiveness to evidence due to holding those prejudicial views about women. All of this together was likely to have led to the downgrading of Elsie's testimony, in the presence of the accurate stereotype about the general population.

There is no easy way to prove that this is the case. And there may be cases where it is not true. But we might say that where an accurate negatively valenced stereotype is true in virtue of conditions of prejudice within the broader environment and the acts this leads to, it is very likely that one cannot have escaped the prejudicial beliefs that led to that stereotype being true. As such, one is not ethically non-culpable for epistemic harm in a testimonial interaction with someone whose credibility is assessed based on that stereotype. This would not entirely correspond with Fricker's use of the concept but it does, I think, represent a reasonable adjustment to it which takes into account the important sub-genre of cases Puddifoot has identified, and which accords with the spirit of what testimonial injustice is intended to capture: prejudicial treatment through use of stereotypes.

5. The role of accurate stereotypes in ordinary testimonial injustice

That said, I think that there may be something more fundamental about the role of context when considering any special considerations we ought to make for cases of testimonial epistemic harm, and the use of accurate stereotypes. And this is the question around whether or not the stereotypes being invoked really *are* accurate, given the context of application.

To examine this, first consider the stereotypes at play in paradigm cases of testimonial injustice. What stereotypes was Greenleaf drawing on, in unjustly failing to give Sherwood her epistemic due, in Fricker's own example? Candidates, with their accuracy, include:

a. That women rely on intuition more than men (could be accurate, given social factors).
b. That women are more intuitive than men (as above, could be accurate).
c. That women are not capable, or are less capable than men are, of rational thought (may or may not be accurate for a given interpretation of rational—if, for example, rationality is understood in a way that requires a particular kind of thought

that women are less likely to be trained in, then women may count as less rational on that account).

d. That women are less likely to deploy a rational approach to problem solving than men are (may or may not be true in a given context, and for a given interpretation of rational, as with suggestion c above).

e. That women are irrational and unable to think rationally (inaccurate as a blanket statement about women; may be accurate under conditions of particular oppression; same caveats as previously around what rational is taken to be).

In the Sherwood case there would also have been, given the time and location of the setting, many other beliefs and biases lurking in the background, which were not stereotypes about a given group (women), but which might come about due to stereotypes about a given group: that intuition plays no part in rational thought; that a person who is emotionally invested cannot be rational; that rationality of a certain kind is actually needed to understand the character of someone who has gone missing and to therefore be of use in finding out what has happened to him, etc. These are all aspects of the social reality of the time period and place.

My point here is not to say that in this paradigm case from literary fiction, or in parallel ones from real life dismissal of women's testimony, testimonial injustice is *in fact* based on accurate stereotypes. Rather, that there are a number of readings where the stereotype being held by Greenleaf could be one that *reflects some aspect of social reality*, given the society and time in which the story is set and the characters are speaking. Nevertheless, somewhere along the line, an *inaccurate* stereotype came into play, for that specific context, in a way which made Greenleaf irresponsive to evidence in a given testimonial interaction, and it is this which arguably drives the poor epistemic behaviour and leads to this being a case of testimonial injustice. And this may not be dissimilar to what happens in the type of cases Puddifoot introduces.

Stereotypes are only accurate with regard to a given population. They are only reliably accurate insofar as they are applied to that same population. They are non-prejudicially applied within a given context only when they are responsive to evidence in egalitarian ways. In the case of the woman in STEM, those around her who don't take her scientific expertise seriously, and who downgrade her input, are not merely relying on accurate stereotypes. They might be understood as inferring *beyond* what the empirical evidence supports, based on any number of further background beliefs and assumptions, and therefore relying on an (implicit or tacit) inaccurate stereotype.

One way to understand this is to attempt to make visible the hidden sub-clauses which are unstated but required for a stereotype to be genuinely accurate. The stereotype:

> "women lack scientific expertise",

is, for example, only accurate within a given context if we add sub-clauses:

> "women (taken as a whole in the base rate population of this particular society) lack scientific expertise (because fewer women than men receive scientific training)".[6]

While this is accurate, it can now clearly be seen that this stereotype is actually irrelevant in assessing the credibility of a person who *does* have scientific training. And so, the person using this stereotype is guilty of one of two things. Either they have failed to attach good reasons to their judgements (because the stereotype is irrelevant), in which case they are epistemically culpable for this in itself and likely to be driven by other motives (of prejudice or simple culpable irrationality) in making their assessment. Or, they have formed a judgement which is *not* based in the claimed stereotype at all, but in a more precisely expanded and *inaccurate* stereotype, such as that:

> "women (who work in science) lack scientific expertise".

If this is the case, then these instances more closely fit the standard model of testimonial injustice. And the problem of a lack of an ethically culpable wrong or wrongdoer in Puddifoot-style cases dissipates.

6. Conclusion

Katherine Puddifoot has drawn attention to the important fact that stereotypes which are accurate can be seen to feature in important cases of testimonial harm. This is an invaluable addition to the scope of what we need to consider when it comes to Fricker's concept of epistemic injustice and to thinking about epistemic harm more broadly.
Further, it might be difficult, given that the stereotypes in these cases are not inaccurate and do not initially appear to involve prejudicial attitudes and act, to recognise epistemic harms in these cases as instances of

[6] As an example, the facts are more complicated than this in terms of rates of women entering into STEM in different societies. Note that the statement is also true of men, when rated against the base rate population, simply because few people are in STEM occupations, relative to the population base of a given society.

injustice, for which hearers and audiences are culpable. This might be thought to leave a worrying responsibility gap in attributing responsibility which dispositional evaluation is not fully equipped to deal with.

In this short response I have, however, offered two ways in which these newly categorised Puddifoot-style cases might still be thought to be cases of testimonial injustice after all: (i) they may involve prejudice, although this may not be located where Fricker initially proposed in her account; or (ii) they may turn out to involve inaccurate stereotypes. If I am right in either case then it is possible that we can identify that in Puddifoot cases, there are ways in which persons are culpable for their use of stereotypes. As such, we can also conclude that an epistemic injustice, not merely unfortunate epistemic harm, has been enacted after all.

REFERENCES

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing.* Oxford: Oxford University Press.

Puddifoot, Katherine. 2021. *How Stereotypes Deceive Us*. Oxford: Oxford University Press.

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 1 | 39-55
https://doi.org/10.31820/ejap.21.1.3

SOME CRITICAL THOUGHTS ON "HOW STEREOTYPES DECEIVE US"

Federico Picinali[1]

[1] London School of Economics and Political Science, United Kingdom

Discussion – Received: 01/07/2023 Accepted: 01/11/2023

This paper is part of a book symposium on Katherine Puddifoot's *How Stereotypes Deceive Us* guest edited by Marina Trakas (IFF, CONICET)

ABSTRACT

The goal of this paper is to identify and discuss the weaker aspects of some of the arguments in Kathy Puddifoot's fascinating and thought-provoking book. Section 1 deals with Puddifoot's treatment of the "single factor view" and the "dual factor view" of stereotyping. Section 2 deals with Puddifoot's treatment of egalitarian attitudes. Section 3 deals with Puddifoot's treatment of the moral encroachment approach to stereotyping. Finally, section 4 deals with Puddifoot's theory of evaluative dispositionalism. The sections can be read independently.

Keywords: stereotyping; egalitarian attitudes; moral encroachment; pragmatic encroachment; evaluative dispositionalism.

Correspondence: f.picinali@lse.ac.uk

Introduction

The goal of this paper is to identify and discuss the weaker aspects of some of the arguments in Kathy Puddifoot's book, such that Puddifoot and other scholars interested in stereotypes may improve, at least marginally, on these arguments. Whether or not my criticisms are persuasive, it would be unfair to forgo the praise that Puddifoot deserves for her important work.

The book is replete with thought-provoking ideas, hypotheticals and arguments, making it a very valuable read for anyone interested in epistemology, and social epistemology in particular. However, there are two contributions that, in my view, stand out for their significance. The first, offered in Chapter 6, is the compelling analysis of the dilemma represented by an individual's choice whether to disclose their mental health condition and/or their social identity in certain contexts. Disclosure increases the risk of being stereotyped and of triggering a series of epistemic pitfalls of stereotyping, detailed by Puddifoot in earlier chapters, that will likely damage the individual facing the dilemma; non-disclosure may well mean that the interlocutor will misunderstand the needs of this individual, as well as their behaviour and attitudes. Puddifoot highlights a genuine problem that warrants further scrutiny. The brief discussion of strategies to tackle the problem (see, in particular, Puddifoot 2021, 128-132) is a valuable starting point for future inquiry.

The second contribution that stands out consists in the claim, defended in Chapters 7 and 8, that the rationality of holding a stereotyping belief *also* depends on the dispositions[1] that are possessed due to holding such a belief, or "downstream" dispositions. As pointed out below, I am not entirely convinced by the way in which Puddifoot defends this claim. The claim itself, though, is intriguing, plausible, and potentially disruptive of the mainstream views on the justification of beliefs.

I now turn to the criticisms, which are grouped in four distinct sections. Section 1 deals with Puddifoot's treatment of the "single factor view" and the "dual factor view" of stereotyping. Section 2 deals with Puddifoot's treatment of egalitarian attitudes. Section 3 deals with Puddifoot's treatment of the moral encroachment approach to stereotyping. Finally, section 4 deals with Puddifoot's theory of evaluative dispositionalism. The sections can be read independently.

[1] Dispositions are defined by Puddifoot as "what a person does, says, thinks, and would do and think in various circumstances" (2021, 164).

1. The single and the dual factor views of stereotyping

A first set of problems pertains to Puddifoot's critical discussion of what she calls the "single factor view" and the "dual factor view" of stereotyping. Both views concern the conditions under which stereotyping increases the chances that the agent makes an accurate judgement about a member of the group to which the stereotype refers.

According to the single factor view

> [T]here is only one feature of any act of stereotyping that determines whether the application of the stereotype (…) increases (…) the chance of an accurate judgement being made: whether or not the stereotype that is applied reflects some aspects of reality. (2021, 32)

On the same page, Puddifoot states that a stereotype

> [R]eflects some aspect of social reality as long as there is a regularity found within society and the stereotype leads a person to respond in a way that reflects the regularity. (Ibid.)

To show that this view is fallacious, Puddifoot offers the example of the stereotype associating Black people more strongly than non-Black people with drug use.[2] According to Puddifoot, this stereotype reflects an aspect of US social reality, this being the high arrest rate for suspected drug use[3] amongst Black people (ibid. 45, 61). And yet, Puddifoot says, the stereotype may also lead "to judgements that fail to fit accurate statistical information about actual rates of *drug use*, which are similar across Black and non-Black populations" (ibid., 45, emphasis in the original). This would allegedly show that the single factor view is wrong:

> An act of stereotyping might involve the application of a stereotype that reflects some aspect of social reality, which would mean that on the single factor view it should increase the chance of an accurate judgement being made. However, because the stereotype does not dispose the person who engages in the stereotyping to respond in a way that reflects the

[2] At pages 45 and 61 Puddifoot phrases the stereotype as referring to "White", rather than "non-Black" people. But in her discussion of the example, she switches between the two attributes. Moreover, the stereotype is sometimes phrased by Puddifoot as referring to "drug crime", sometimes as referring to "drug use". I have attempted to bring consistency to the example. Also consider the following footnote.

[3] NB: in England and Wales the law does not punish the "use" of drugs *per se*. Rather, the Misuse of Drugs Act 1971 targets conducts such as the importation, exportation, production, possession, and supply of controlled drugs.

> statistical reality, the stereotyping could reduce rather than enhance the chance of an accurate judgement being made (2021, 45).

To clarify, the stereotype in question may lead someone to conclude that a Black acquaintance is more likely than a non-Black acquaintance to use drugs. This conclusion—Puddifoot suggests—reflects the high arrest rate for suspected drug use amongst Black people; but it does not reflect the "statistical reality" concerning drug use.

My objection to Puddifoot's analysis is that the example she relies on does not show that the single factor view is wrong. Contrary to what is argued by Puddifoot, the stereotype about comparative drug use does not reflect the aspect of social reality represented by arrest rates, such that the single factor view would not insist that reliance on it increases the chances of making an accurate judgement concerning drug use. In other words, this view would likely lead to the same conclusion reached by Puddifoot about the accuracy of a judgement that relies on the stereotype.

To begin with, it is not clear in what sense the stereotype in question can be said to reflect the high arrest rate concerning Black people, given that the former is comparative (as all stereotypes are, according to Puddifoot's definition, to which I'll soon return), whereas the latter does not provide comparative information. The (spurious) proposition that Black people are more likely than non-Black people to use drugs is consistent with any arrest rate concerning Black people alone—insofar as this is higher than the rate amongst non-Black people—and may, therefore, "lead a person to respond in a way that reflects" a very low arrest rate amongst Black people. I take it, then, that with "high arrest rate" Puddifoot actually means that the arrest rate is *higher* for Black people than for non-Black people. Even so, the stereotype in question cannot, in fact, be said to reflect this comparative arrest rate.[4]

The stereotype may well lead someone to conclude that a Black acquaintance is more likely than a non-Black acquaintance to use drugs. Pace Puddifoot, though, this judgement about the comparative likelihood of drug use would not *reflect* the higher arrest rate for suspected drug use amongst Black people. The same holds for the stereotype on which the judgement is based. The judgement (and the stereotype) could meaningfully be said to reflect this comparative arrest rate only under the assumption that the comparative arrest rate tracks the comparative rate of

[4] By "comparative rate" I mean a construct that compares the rate for one group with the corresponding rate for the other group.

drug use, such that an inference could be reliably drawn between the two. To see why, consider the case in which the comparative arrest rate does not track the comparative rate of drug use, perhaps, because, due to racism in law enforcement, the arrest rate amongst Black people for suspected drug use is substantially higher than the rate of drug use amongst Black people. In such a case, no generalisation or individualised judgement about the comparative likelihood of drug use of Black and non-Black persons could reliably be formed based on the comparative arrest rate. After all, this rate is disconnected from the state of affairs about drug use, being determined, instead, by forces such as hatred, dislike, and suspicion towards Black people. If, however, the comparative arrest rate does not provide us with information on which to base reliably a generalisation or an individualised judgement about the comparative likelihood of drug use, it is mystifying to claim that the generalisation or judgement *reflect* the rate. They could reflect the comparative arrest rate only under the assumption that arrests are evidence of drug use. But this assumption is, *ex hypothesi*, false.

In the circumstances of Puddifoot's example, in fact, the assumption of a correspondence between the comparative arrest rate and the comparative rate of drug use cannot apply. In societies such as the US and England and Wales, where structural racism is a significant issue and where police forces are affected by institutional racism,[5] Black people are overrepresented among those arrested, both in general and with regard to drug-related offences considered separately. What is more important, a Black individual is significantly more likely than a White individual to be arrested, [6] notwithstanding that "BAME groups are less likely to commit crime".[7] Given the disconnect between the comparative arrest rate and the comparative rate of drug use, it is not clear how any judgement (or

[5] As far as the Metropolitan Police Service is concerned, see Baroness L. Casey, *Final Report. An Independent Review into the Standards of Behaviour and Internal Culture of the Metropolitan Police Service* (2023), available at https://www.met.police.uk/SysSiteAssets/media/downloads/met/about-us/baroness-casey-review/update-march-2023/baroness-casey-review-march-2023a.pdf (accessed 9 August 2023).

[6] For data about England and Wales see Ministry of Justice, *Black, Asian and Minority Ethnic Disproportionality in the Criminal Justice System in England and Wales* (2016), in particular, at 12, 22, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/639261/bame-disproportionality-in-the-cjs.pdf (accessed 9 June 2023) and D. Lammy MP, *The Lammy Review: Final Report. An Independent Review Into the Treatment of, and Outcomes for, Black, Asian and Minority Ethnic Individuals in the Criminal Justice System* (2017), available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/643001/lammy-review-final-report.pdf (accessed 9 June 2023). For data about the US see Bureau of Justice Statistics, 'Race and Ethnicity of Violent Crime Offenders and Arrestees, 2018' available at https://bjs.ojp.gov/content/pub/pdf/revcoa18.pdf (accessed 9 June 2023) and R. Camplain and others, 'Racial/Ethnic Differences in Drug- and Alcohol-Related Arrest Outcomes in a Southwest County From 2009 to 2018' (2020) 110 American Journal of Public Health 85.

[7] Ministry of Justice, n 6, at 12.

stereotype) about the comparative likelihood of drug use could be said to *reflect* the comparative arrest rate; and it is not clear why a defender of the single factor view would maintain that said judgement (or stereotype) reflects said rate.

If the single factor view is indeed wrong, then, the example of a stereotype concerning the comparative rate of drug use amongst the Black and the non-Black populations does not contribute to show this. This is not an example in which the single factor view would insist that the stereotype reflects a relevant aspect of social reality (other than racism, that is). Notice, instead, that if the stereotype actually reflected the comparative arrest rate—due to the fact that the comparative arrest rate actually tracks the comparative rate of drug use—it would be perfectly sensible to argue that reliance on such a stereotype increases the chances of judging accurately whether a Black acquaintance is more or less likely to use drugs than a non-Black acquaintance. In this case the stereotype would indeed reflect the "statistical reality" concerning drug use.

Consider now the definition of the "accuracy" of a stereotype, introduced by Puddifoot in her critical discussion of another account of stereotyping, the dual factor view. Puddifoot states that

> A stereotype can be deemed to be accurate if it leads a person to respond in a way that is fitting with accurate statistical information about the distribution of traits across groups. (2021, 40)[8]

According to this definition, the accuracy of a stereotype does not reside in the accuracy of the generalisation that the stereotype represents: it consists, instead, in the accuracy of the judgement that is produced relying on the stereotype. This definition, though, seems to turn the dual factor view into tautology, since this view is construed by Puddifoot as claiming that

> [J]udgements produced as a result of acts of stereotyping are more likely to be accurate than alternative judgements (…) if and only if (a) the stereotype that is applied is accurate (…). (2021, 38)[9]

[8] Surprisingly, on page 56, the author replaces the verb "to lead" with the less demanding verb "to dispose".
[9] That is, if the stereotype leads to accurate judgements.

Moreover, Puddifoot's definition of the accuracy of a stereotype seemingly undermines her criticism of the dual factor view. According to Puddifoot, such view is flawed because it does not acknowledge a series of factors that may compromise the accuracy of the judgement resulting from stereotyping: the irrelevance of the stereotype to the particular decision problem, the misinterpretation, ignorance or selective recollection of the evidence, the discrediting of the testimony of the stereotyped person etc. If, however, a stereotype's accuracy is defined in terms of the accuracy of the resulting judgement, the factor of the stereotype's accuracy, which is central to the dual factor view, would already encompass all the epistemic factors that, according to Puddifoot, such a view mistakenly ignores (see 45-55). Indeed, these are all factors that may intervene between the endorsement of the stereotype and the ensuing judgement in each case; thus, factors that may compromise the accuracy of the judgement and, hence, according to Puddifoot's definition, also of the stereotype. If a stereotype's accuracy is defined in terms of the accuracy of the resulting judgement, then, requiring the former means requiring that the epistemic factors highlighted by Puddifoot do not materialise—or that, if they do materialise, they do not affect the judgement's accuracy.

Perhaps I am mistaken: Puddifoot never meant to define the accuracy of a stereotype in terms of the accuracy of the resulting judgement (there are indications to this effect in Chapter 2, Section 5). If so, an alternative interpretation must be given of the phrase "to respond in a way that is fitting with accurate statistical information", which is central to Puddifoot's definition on page 40. If this phrase does not mean "to judge accurately", what does it mean? A first possibility is that it means "to rely on accurate statistics in making one's judgement"—where the judgement may, however, turn out to be inaccurate. For a stereotype to lead the agent to rely on accurate statistics, one would expect that the stereotype itself must reflect such statistics. If so, though, why not defining the accuracy of a stereotype simply in terms of the accuracy of the generalisation it represents, rather than by reference to what it leads the agent to do? A second possibility is that the phrase means "to rely on accurate statistics in making one's judgement *and* to give to such statistics the appropriate weight". Again, the ensuing judgement may well be inaccurate, but for the stereotype to be accurate it must provide the agent with an accurate generalisation, and it must lead the agent to use this generalisation correctly. This means, for example, that the generalisation should not bring the agent to overlook case-specific evidence indicating that the case does not fit the generalisation, nor should it enable the recollection of only the evidence indicating that the case fits. It is unclear whether Puddifoot intended to define the accuracy of a stereotype according to this second hypothesis. It is worth noting, though, that this hypothesis presents the

problem discussed in the previous paragraph. If the accuracy of a stereotype is defined in terms of the appropriate weight given by the agent, in the particular decision problem, to the accurate generalisation reflected in the stereotype, then the dual factor view already includes some of the epistemic factors that, according to Puddifoot, are not part of it.

To conclude on Puddifoot's critical discussion of the single and dual factor views of stereotyping, Chapter 3 surprisingly lacks a definition of "relevance", notwithstanding that this notion is central to the discussion. Puddifoot argues that a flaw of these views is their failure to recognise that a factor that may undermine the accuracy of a judgement based on stereotypes is the possibility that the stereotype be triggered even when irrelevant to the judgement. To illustrate this point, she gives the following example of an irrelevant stereotype:

> A police officer approaches the car of a Black male, which has been pulled over for a minor traffic violation, e.g. one of his headlights is not working. The police officer asks the man to step out of the vehicle but he responds slowly and cautiously to the command. The police officer is offended at what he takes to be a threat to his authority. This triggers a stereotype associating the innocent man with crime; the police officer evaluates the man as a criminal and treats him with hostility; and this leads to an escalation of tension and hostility between the two individuals. The stereotype associating Black people with crime is triggered although the Black man has not committed a crime, only a minor traffic violation. (2021, 46)

Perhaps the omission of a definition of "relevance" is only remarkable for someone who, like me, works in the field of evidence law, where this notion is a cornerstone. But Puddifoot's example cannot be elucidated by falling back on the general understanding of this term as referring to the quality of being "related or useful to what is happening or being talked about".[10] In light of this understanding, I fail to see how the stereotype in the example, while admittedly spurious, is irrelevant to the decision-making of the officer. An officer's job includes preventing and detecting crime. True, the officer may have stopped the car for the sole purpose of fining the driver for a malfunctioning headlight. However, if, after stopping the car, the officer forms a suspicion that the driver has committed, is committing or will commit a crime, it is part of the officer's job to act on that suspicion. The stereotype being about the relationship between a social group and criminality, it is surely relevant to forming the

[10] See https://dictionary.cambridge.org/dictionary/english/relevance (accessed 9 June 2023).

suspicion; hence, to the decision-making of the officer in the case at hand. Again, the stereotype is spurious and may, therefore, lead the officer astray. But a charge of inaccuracy is not the same as a charge of irrelevance.

2. Egalitarian attitudes and epistemic pitfalls

On page 3, Puddifoot defines the concepts of "stereotype" and of "stereotyping" in comparative terms. Both involve "a social attitude that associates members of some social group *more strongly than others* with certain traits" (emphasis added). According to Puddifoot, what is required is not that the attitude reflects a stronger association than the actual one. After all, she makes clear that, according to her definition, stereotypes and stereotyping need not be inaccurate or lead to inaccurate judgements. The definition requires, instead, that the attitude reflects a stronger association for group members than for others. Therefore, an "egalitarian" attitude (that is, an attitude associating group and non-group members with equal strength with a certain trait) cannot be a stereotype. I am not convinced by the inclusion of this comparative definitional element: if, for example, vegans have a low carbon footprint and non-vegans have a higher footprint, isn't there a sense in which vegans are stereotyped if people associate both vegans and non-vegans with the carbon footprint that characterises the latter? Be that as it may, I won't dwell on the definition here.

In Chapter 4, Puddifoot discusses, among other things, the epistemic benefits of avoiding stereotyping. Consider an attitude that associates scientific expertise more strongly with men than women. It is a stereotype according to the above definition. Moreover, it is a stereotype that, to use Puddifoot's words, reflects an aspect of social reality in the United Kingdom: there are significantly more men than women in the sciences. This notwithstanding, Puddifoot argues that reliance on this stereotype involves significant epistemic pitfalls, introduced in Chapter 3. Conversely, an egalitarian attitude according to which men and women are equally likely to have scientific expertise does not track reality. And yet, Puddifoot argues that this epistemic cost is the lesser evil, compared with the epistemic pitfalls that would be avoided by endorsing such an attitude (ibid., 79). The bottom line is that, at least in some cases, avoiding stereotyping is preferable from an epistemic point of view, even if the stereotype tracks reality in some salient respect. In other words, in some cases endorsing an egalitarian attitude is preferable from an epistemic point of view, even if the attitude is inaccurate.

Puddifoot's conclusion may well be correct. Contrary to what she suggests, though, the egalitarian nature of the alternative attitude cannot alone guarantee the less costly epistemic outcome. The author writes that

> The possession of the attitude that women and men are equally likely to have scientific expertise will guard against various tendencies that accompany the stereotype associating scientific expertise more strongly with men than women. (2021, 75)

The reader will remember from Chapter 3 that these tendencies, or pitfalls, include misinterpreting, ignoring, or remembering selectively the evidence, so that the recognised evidence fits and confirms the stereotype, as well as giving a credibility deficit to the testimony of those who are stereotyped. Now, if I endorsed an attitude according to which both men and women are *extremely unlikely* to have scientific expertise, I would probably fall prey to (at least some of) these pitfalls when assessing the testimony of a woman (or of a man) scientist—e.g., the base rate that I endorse may lead me to give this testimony a credibility deficit. This is notwithstanding that my attitude is egalitarian (hence, not a stereotype). The egalitarian nature of the attitude may mean that the distribution of epistemic errors is also egalitarian: other things being equal, errors will equally affect the men and the women whose expertise I judge. Pace Puddifoot, though, it does not necessarily mean that the above epistemic "tendencies" are avoided, or that they are less pronounced than in the case of stereotyping.

3. The moral encroachment approach to stereotyping

Chapter 5 analyses the relationship between epistemic and ethical demands in the case of stereotyping. Puddifoot argues that this relationship is a complex one: depending on the context, epistemic and ethical demands may align or may clash. Here is a brief illustration of her account. Reliance on a stereotype that reflects an accurate generalisation (say, associating Black people more strongly than non-Black people with a certain medical condition) may lead to an accurate judgement (say, an accurate medical diagnosis of a Black patient). But it may also trigger further stereotypes (say, that associating Black patients more strongly than non-Black patients with uncooperativeness), which may foster a poor interaction between the individuals involved (perhaps with detrimental medical consequences for the Black patient). This is a case in which epistemic and ethical demands appear to clash: only the former justify stereotyping. In earlier chapters, though, Puddifoot has argued that stereotypes (including those that reflect accurate generalisations) may lead to significant epistemic pitfalls, such as the misinterpretation, ignorance or selective recollection of evidence, and

the unwarranted discounting of the testimony of the stereotyped person. Hence, reliance on a stereotype may lead to a series of inaccuracies (say, unwarranted distrust in the patient's description of their symptoms). In this case, epistemic and ethical demands may align: neither justifies stereotyping.

Puddifoot contends that the moral encroachment approach to stereotyping cannot capture the complexity of the relationship between epistemic and ethical demands. Her conclusion, though, seems too quick. To see why, let's start with Puddifoot's brief illustration of the approach. According to its defenders

> [E]thical and epistemic demands do not conflict because moral considerations determine whether it is epistemically permissible to engage in stereotyping (…). Where there are high moral stakes in a situation in which a judgement is made, high evidentiary standards need to be met in order for a judgement to be justified or rational or to constitute knowledge. In cases where people might engage in stereotyping, there will often be high moral stakes, and these stakes will raise the evidentiary standards. Those who engage in stereotyping will not meet the high evidentiary standards. Therefore, stereotyping will not be justified or rational. (2021, 114)

Puddifoot highlights two main problems with this approach. First, it is "overly simplistic" in its failure to acknowledge that epistemic and moral considerations may, indeed, clash. Second, it mistakenly assumes that stereotyping is never justified in high-stakes situations: "sometimes high-stakes situations demand stereotyping, because stereotyping can be an efficient way to achieve both ethical and epistemic goals, like correct diagnoses and treatment decisions" (ibid, 115). I am not particularly interested in engaging with the first criticism. The question as to which are the possible arrangements between the epistemic and the ethical "vectors" strikes me as more formal than substantial. What matters most in assessing an approach to stereotyping is whether the approach offers appropriate guidance to those who may stereotype. In this regard, I agree with Puddifoot that it is mistaken to foreclose the possibility of stereotyping in high-stakes situations. However, I believe that a moral encroachment approach to stereotyping can account for cases in which stereotyping is (epistemically and ethically) justified, notwithstanding the high stakes involved. In order to defend this thesis, though, I will first defend the equivalent thesis formulated with reference to the distinct theory of pragmatic encroachment.

The essential tenet of pragmatic encroachment is that whether someone knows *that p* (say, that a patient has a given condition) depends on the importance that getting the matter right has for this individual, thus on their practical interests involved. It is possible to operationalise the pragmatic encroachment approach by modelling a decision problem using decision theory. This will allow us to account for the stakes, as well as to identify the evidentiary standard (or probability threshold) that must be satisfied for someone to be justified in acting as if the proposition at issue were true, given the stakes. Under pragmatic encroachment, being justified in acting this way means having an "outright" belief in the proposition—a belief necessary for knowledge. A full illustration of this model is beyond the scope of this article.[11] It suffices to point out the following. The model requires identifying the possible outcomes of the decision problem (say, the problem whether to diagnose a condition). These are: the correct outcomes consisting in acting as if the proposition were true when it is, indeed, true (say, diagnosing a condition when the patient has such a condition) and in acting as if the proposition were false when it is, indeed, false (say, not diagnosing a condition when the patient does not have such a condition); and the mistaken outcomes consisting in acting as if the proposition were true when it is, indeed, false (say, diagnosing a condition when the patient does not have such a condition); and in acting as if the proposition were false, when it is, indeed, true (say, not diagnosing a condition when the patient has such a condition).

Once the possible outcomes are identified, a value should be assigned to each outcome, reflecting the preferences, or practical interests, of the decision-maker. The model then allows to identify a probability threshold the satisfaction of which justifies acting as if the proposition at issue were true, given such values (i.e., the stakes). For the purposes of the pragmatic encroachment approach, this probability threshold corresponds to the attitude of outright belief: the decision-maker outright believes the proposition at issue if and only if, their degree of belief in that proposition corresponds to a probability equal to, or greater than, such threshold. Therefore, outright belief in the proposition at issue implies that the agent is justified in acting as if that proposition were true.

Notably, the model does not foreclose the possibility that the decision maker has an outright belief in a proposition notwithstanding that their degree of belief in that proposition is fairly low *and* the decision problem hinging on the truth of that proposition involves high stakes. It is possible

[11] In essence, however, a decision problem hinging on the probability of a particular state of affairs (e.g., whether a patient has a given condition) can be modelled similarly to the decision problem of criminal adjudication, which hinges on the probability of the defendant's guilt. For a decision-theoretic approach to the latter, see Picinali (2022, chs. 3 and 4)

that acting as if the proposition were false when it is, indeed, true, is an extremely costly outcome, whereas acting as if the proposition were true when it is false, presents moderate or little cost. For example, not diagnosing a condition when the patient has such a condition may quickly lead to the patient's death, whereas diagnosing the condition when the patient does not have it may result in subjecting the patient to a treatment with some beneficial consequences and few side effects. Depending on the values of the correct outcomes, in this situation the model may well indicate that the evidentiary threshold for outright belief in the proposition at issue is low: the agent may outright believe and, hence, know that the patient has the condition (and may be justified in acting accordingly, that is, in diagnosing the condition and giving treatment) even in the presence of weak evidence that the patient, indeed, has it. This is a high-stakes situation (one of the mistaken outcomes, the false negative, is very costly) in which stereotyping may well be warranted in accordance with the essential tenet of pragmatic encroachment, notwithstanding that the stereotype linking members of a group to which the patient belongs with the condition (more strongly than non-members) may not provide robust epistemic support.

The essential tenet of moral encroachment is that whether someone knows *that p* depends on the moral stakes of the decision problem. The practical interests of the decision maker, with which pragmatic encroachment is concerned, may not track the moral stakes (e.g., the agent may care very little about whether the patient will suffer harm in case of a false negative, being chiefly concerned with the costs of treatment for the hospital's coffers if the condition is indeed diagnosed, and with the repercussion these may have on a desired salary increase). Therefore, pragmatic encroachment and moral encroachment may lead to different conclusions as to whether the agent knows *that p*.

In standard decision theory the value function reflects the decision maker's preferences, which is why pragmatic encroachment lends itself to modelling through decision theory. However, one can construct the value function as tracking moral value, that is, as reflecting the preferences of the morally conscientious agent.[12] Once this condition is added, the argument can be rerun with reference to moral encroachment. In essence, the point is that there will be situations in which the false negative has such a high moral cost that the threshold for outright belief (and, hence, for knowledge) will be relatively low; sufficiently low to be satisfied by stereotyping. In the

[12] One may object that decision theory cannot capture essential aspects of a deontological theory, such that a deontologist cannot model moral problems with decision theory. On this issue see the work of Lazar (2017).

medical example in which a false negative (missed diagnosis) leads to a quick death and a false positive (false diagnosis) leads to some health benefits and mild side effects, the doctor may well be morally warranted to follow a stereotype linking members of a group to which the patient belongs with the medical condition at issue (more strongly than non-members).

4. Evaluative dispositionalism

On page 164, Puddifoot introduces her theory of "evaluative dispositionalism" (ED) as a theory to assess the justification for holding "stereotyping beliefs", that is, "beliefs that encode generalizations about social groups, associating all group members more strongly than non-group members with some feature" (2021, 145). This theory is introduced after arguing that existing accounts of epistemic appraisal (that is, upstream, downstream,[13] and static accounts) cannot capture some of the epistemic faults associated with holding stereotyping beliefs. According to ED

> [A] complete epistemic evaluation of an act of believing should focus on both (a) the dispositions that are displayed in believing, and (b) the dispositions that are possessed due to believing. (2021, 164).

Effectively, ED is a combination of the existing accounts.

A preliminary issue with Puddifoot's analysis is that it fluctuates between presenting ED as a theory for the appraisal of beliefs tout court and as a theory for the appraisal of stereotyping beliefs only. On pages 158 and 188, Puddifoot states that she is only concerned with the more modest task of offering a theory that targets stereotyping beliefs. And yet, the formulation of ED that I have just reproduced is couched in general terms. What is more puzzling is that, in illustrating the theory, Puddifoot relies on examples of beliefs that are not stereotyping beliefs (e.g., the belief that Manchester City is going to win the Premier League). Perhaps this fluctuation betrays greater (and perfectly sensible) ambitions than those declared. Be that as it may, ED is not an entirely convincing theory even when circumscribed to stereotyping beliefs; or, at least, Puddifoot does not show that it is.

On pages, 181-2 Puddifoot offers a hypothetical case in which two individuals harbour the same stereotyping belief and yet, according to her, this belief deserves different epistemic evaluation in the two cases (it is

[13] To be sure, downstream accounts are, by and large, a creation of Puddifoot (see 2021, 141-143).

justified in one case only, perhaps) because only one individual has dispositions to make epistemic mistakes due to holding the stereotyping belief. Here is the hypothetical case.

> Consider two people. Nora is a female scientist who has 30 years of experience. She is a feminist and as a result pays close attention to the representation of women in the sciences. She notices over time that a gender gap in the sciences never goes away: there are consistently more men than women in sciences, and therefore consistently more men than women with scientific expertise. Nora therefore harbours a stereotype associating men more strongly than women with scientific expertise. She harbours and endorses the social attitude *men are more likely than women to have scientific expertise*. However, the stereotype does not distort Nora's judgements of individual women scientists and their levels of expertise. The stereotype does not make her assume that women scientists are more similar to each other than they really are, or that they are less similar to men scientific experts than they really are. She does not misremember the features of women scientists due to the operation of the stereotype. And so on. Instead, Nora judges women scientists on the basis of the skills, expertise, and potential that they display in their work, with the stereotype only operating to allow her to understand the challenges that they are likely to have faced as a minority in the profession.
>
> Ned is also a scientist with 30 years of experience. He is not a feminist. He has also registered that women are underrepresented in the sciences and therefore harbours a stereotype associating men more strongly than women with scientific expertise. He endorses the social attitude *men are more likely than women to have scientific expertise* and thereby harbours the same stereotype as Nora. However, the stereotype that he harbours permeates his thought, influencing all of the judgements that he makes about individual women scientists, and about the relative merits of men and women scientists. He makes errors such as misremembering the attributes of his women colleagues, misinterpreting ambiguous behaviours as indicating a lack of expertise, assuming women colleagues are more similar than they really are, and so on.
>
> These two characters harbour the same stereotype: that men are more likely than women to have scientific expertise. Their stereotypes are formed on the basis of the same evidence:

> evidence about the underrepresentation of women in the sciences. But on the evaluative dispositionalist account the characters and their act of believing deserve different epistemic evaluations because the characters differ in the dispositions that they have due to possessing the stereotyping belief: Ned has dispositions to respond poorly to the evidence while Nora does not. This seems to be precisely the right result.

Both Nora and Ned hold the stereotyping belief that men are more likely than women to have scientific expertise. As it happens, this belief is accurate in Nora's and Ned's society (remember that, under Puddifoot's definition, stereotypes need not be inaccurate). Since they both hold this belief, one would expect that both will fall prey to a series of epistemic pitfalls which, as argued by Puddifoot in earlier chapters, are produced by harbouring stereotypes. And yet, this is not the case: only Ned incurs the pitfalls. Puddifoot suggests that this is because Nora is a feminist, whereas Ned is not. Consequently, while holding the stereotyping belief, Nora will not, e.g., misinterpret evidence concerning the scientific expertise of a woman or give a credibility deficit to the testimony of a female scientist. In this scenario—Puddifoot concludes—ED indicates that Nora is justified in holding the stereotyping belief, whereas Ned is not.

What is perplexing about this conclusion is that, as accepted by Puddifoot, the epistemic pitfalls which Ned incurs are not, or not just, "due to" his holding the stereotyping belief. If they were just due to Ned's holding the belief, one would reasonably expect Nora to incur such pitfalls as well. Perhaps the fact that Ned holds the stereotyping belief is indeed irrelevant to the occurrence of the pitfalls, these being entirely brought about by pre-existing dispositions such as the rejection of feminism. Perhaps holding the belief is not irrelevant after all: it acts as an enabler or as an enhancer of pre-existing dispositions that bring the agent to incur the epistemic pitfalls. Now, if holding the stereotyping belief is irrelevant, then there is no reason to accept that the pitfalls should be factored into the assessment of the justification of holding such belief. In fact, ED says that they shouldn't, since they are not "possessed due to" holding the belief. If, instead, holding the stereotyping belief enables or enhances pre-existing conditions, then there is an explanatory story to be told and evaluated, the complexity of which is not accounted for in the current formulation of ED.

I suggest that ED would be improved by clarifying how significant should be the role of the belief in bringing about the pitfalls for the pitfalls to be factored into the assessment of the justification of holding the belief. In the current formulation of ED, any causal (or enhancing) role of the belief seems sufficient. However, this may be too strict a position to take. After

all, it is doubtful whether it would be epistemically unjustified to hold a stereotyping belief that is accurate (as that in the example) and contributes only minimally to bringing about the epistemic pitfalls. It is also possible that the question of the role played by the stereotyping belief in the aetiology of the pitfalls is intractable: there is simply no way of ascertaining the nature and extent of this role in any given case. If so, ED is in even greater trouble.

5. Conclusion

This article was an attempt to highlight some flaws in Kathy Puddifoot's book. First, I criticised her treatment of the "single factor view" and of the "dual factor view" of stereotyping, raising questions about the examples and the conceptual apparatus that Puddifoot relies on. Second, and contra Puddifoot, I argued that endorsing an egalitarian attitude may not avoid the epistemic pitfalls associated with stereotyping. Third, I argued against Puddifoot's claim that a moral encroachment approach cannot justify resorting to stereotypes in high-stakes situations. Fourth, and finally, I argued in favour of enhancing Puddifoot's "evaluative dispositionalism" with a clarification of the causal role of the stereotyping belief vis-à-vis downstream dispositions.

These flaws notwithstanding, the book is a fascinating and engaging read, highly recommended to epistemologists and legal scholars alike.

Acknowledgments

I am grateful to Kathy Puddifoot and Federico José Arena for their comments on an earlier draft.

REFERENCES

Lazar, Seth. 2017. "Deontological Decision Theory and Agent-Centered Options." *Ethics* 127 (3): 579–609. https://doi.org/10.1086/690069.

Picinali, Federico. 2022. *Justice In-between: A Study of Intermediate Criminal Verdicts*. Oxford and New York: Oxford University Press. https://doi.org/10.1093/oso/9780198864592.001.0001.

Puddifoot, Katherine. 2021. *How Stereotypes Deceive Us*. Oxford: Oxford University Press.

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 1 | 57-65
https://doi.org/10.31820/ejap.21.1.4

SOME STARTLING CONSEQUENCES OF CONSEQUENCES OF HOW STEREOTYPES DECEIVE US

Jennifer Saul[1]

[1] University of Waterloo, Canada

Discussion – Received: 01/11/2023 Accepted: 01/06/2024

This paper is part of a book symposium on Katherine Puddifoot's *How Stereotypes Deceive Us* guest edited by Marina Trakas (IFF, CONICET)

ABSTRACT

This paper argues that Puddifoot's arguments in *How Stereotypes Deceive Us* have more radical consequences than those argued for in the book. It does this by pointing out two problems for evaluating stereotypes via Evaluative Dispositionalism, Puddifoot's view. The first problem concerns the very large number of dispositions associated with any stereotype, and the second the difficulty of evaluating a stereotype in isolation from other elements of a person's psychology. The paper suggests that, when we take seriously Puddifoot's arguments, we'll end up concluding that it's not possible to assess the epistemic worth of any individual belief or stereotype. We could still discuss the epistemic merits of how it was formed, and discuss the epistemic consequences that this belief or stereotype has for a particular person in a particular situation. But overall epistemic evaluations of stereotypes, or even acts of stereotyping, would be something that we should try to avoid.

Keywords: stereotypes; dispositions; beliefs; epistemic evaluation.

Correspondence: jennifer.saul@uwaterloo.ca

Introduction

How Stereotypes Deceive Us (2021) is a truly wonderful book. It argues for some rather startling conclusions, but it does so incredibly well and incredibly carefully. The startling claims come to seem not just inevitable but even commonsensical. This is a remarkable feat, and an indication of the book's importance. It is the sort of book that really has the potential to change the way that people see some very important issues. What I will do here is to suggest that the book might have even more radical consequences than Katherine Puddifoot has herself argued for. First, I will rehearse some key arguments from the book, then I will suggest the possibility that they lead to even more startling conclusions than those which Puddifoot has already drawn out.

1. Key arguments from how stereotypes deceive us

I'll start by laying out some of the startling, or at least surprising, claims made.

1. Social attitudes that do not reflect reality may sometimes be epistemically good.
2. Acts of truthful stereotyping may sometimes be epistemically bad.

These claims are, we will see, importantly interrelated: according to Puddifoot, it is because 2 is true that 1 is true.

As is obvious from these claims, Puddifoot rejects the idea of any summary judgement of stereotyping as always bad. She adopts a very broad and non-normative definition of "stereotype" and "stereotyping". On her definition, stereotypes are "social attitudes that associate members of some social group more strongly than others with a certain trait or traits" (2021, 23). I will pause to note here that this is an extremely broad definition of stereotype, one that is in certain ways at odds with ordinary usage. (This is one of very few places where an opponent might find something to object to in Puddifoot's arguments.) To see this, consider the group of people who teach at Durham University. This is clearly a social group. I associate this social group more strongly than certain others (e.g. people who teach at Southampton University) with teaching in the north of England. It seems strange to say that I *stereotype* people who teach at Durham University as teaching in the north of England. On Puddifoot's view, however, this is precisely what I am doing. I don't think that she needs to be troubled by this, though, as her chief goal is not one of perfectly capturing ordinary

usage. But it seems perhaps worth noting that there is something here to think about.

How, then, does Puddifoot argue for her surprising conclusions? A crucial step here is Puddifoot's view that there are many aspects we must consider when deciding whether an attitude is epistemically good or bad. (She calls her view Evaluative Dispositionalism.) Amongst these, crucially, are:

> Does the stereotype dispose the person who applies it to respond in a way that is fitting with accurate statistical information (…). Does the application of the stereotype lead information about the specific case to be distorted or ignored? For instance does the application of the stereotype lead to: distorted remembering, the misinterpretation of ambiguous evidence, false assumptions about similarities/dissimilarities among groups and group members, aspects of the social identity of the person who is stereotyped being missed, testimonial silencing, testimonial injustice? (Puddifoot 2021, 56).

The next important step is realising that acts of truthful stereotyping can be epistemically bad. Take, for example, the belief that men are more likely than women to be scientists. This is true, and on Puddifoot's view it is a stereotype. It has some obviously epistemic strengths: along with being true, it will allow one to make a reasonably good guess in the absence of other information about whether any randomly chosen scientist is a woman, for example. But it can also lead to errors, due to the power of the largely automatic associations that are likely to come with it. For example, someone with this belief who is evaluating a job candidate would likely be prone to:

> (i) memory distortions that would make them selectively remember features of the candidate; (ii) viewing ambiguous behaviours of the candidate as evidence of lack of expertise; (iii) failing to notice differences between the candidate and other, previously encountered female scientists; (iv) failing to notice similarities between the candidate and male scientists; (v) the tendency to assume that any behaviours that are stereotypical of non-experts (…) are indicative of the dispositions of the candidate rather than the situation that she is placed in; and, finally, (vi) the tendency to make associations with a candidate that are inaccurate. (Ibid. 79-80)

This is, in brief, the case for accepting that some acts of truthful stereotyping can be epistemically bad—and it's a convincing case.

The fact that truthful stereotyping can be epistemically bad is crucial to the argument that it may sometimes be epistemically good to have social attitudes that don't reflect reality. A key case for Puddifoot is that of a person—call him Albert—who lacks the stereotype that men are more likely than women to have scientific expertise, and instead has the false belief that women and men are equally likely to have scientific expertise. This false belief, Puddifoot argues, brings significant epistemic benefits: Albert will respond to a woman candidate for a position in science in such a way as to avoid the many epistemic faults just outlined above. The case Puddifoot makes here is both startling and undeniable: it is clear that Albert's false belief is doing a tremendous amount of good epistemically. Puddifoot firmly establishes, then, that when making epistemic judgements we must look to far more than the truth or falsity of the beliefs or stereotypes at issue.

This leads to a further suggestion of Puddifoot's—that Albert's false belief may be *epistemically innocent*: "a cognition is epistemically innocent if there is no alternative cognition that would confer the same benefits without the costs" (2021, 78).[1] She suggests that this may well be the case for a false belief like that described above. After all, it is extremely difficult to avoid automatic stereotyping if one has beliefs like *women are less likely than men to have scientific expertise*. Albert's false belief confers quite considerable epistemic advantages which may well outweigh its costs. Puddifoot admits that it may in fact be difficult to determine whether there is an alternative cognition available (ibid., 79), and also that there are methods which can be used in order to avoid such automatic stereotyping-meaning that there is a real possibility of alternative cognitions with the same benefits and fewer costs. (We will return to this latter point shortly.)

Puddifoot's preferred method of assessing beliefs/stereotypes is Evaluative Dispositionalism, which requires examining both the dispositions that give rise to the cognition in question and the dispositions flowing from it. This would allow us to say that the person in the example above has an epistemically flawed belief, both in terms of its falsehood and in terms of the epistemic faults that gave rise to it, while at the same time appreciating the epistemic benefits that it brings. We can also identify epistemic flaws that may flow from it—such as the disposition to not appreciate arguments for working to increase the representation of women in science (we see these in Roger, a later example Puddifoot discusses). There is a wonderful

[1] For discussion of the notion of epistemic innocence, see Bortolotti (2020).

complexity in all of these discussions, one which makes very clear the folly of simple-minded judgments of stereotypes or beliefs as simply good or bad.

Nonetheless, Puddifoot insists that her approach can give clear guidance in how to evaluate acts of believing stereotypes:

> [T]he approach provides clear prescriptions about how to approach acts of believing stereotypes: check the dispositions associated with so believing. It allows us to distinguish satisfactorily between two different acts of stereotyping, explaining, for example, how two people can believe the same stereotype, under the same circumstances, will be deserving of different levels of praise and criticism. (2021, 12)

2. Reflections on *How Stereotypes Deceive Us*

As I reflected further on this rich and interesting picture, I began to wonder whether Puddifoot should take this complexity yet further—by abandoning the very idea of epistemically assessing individual beliefs or stereotypes themselves at all (except perhaps in terms of truth or falsehood). Or, at the very least, abandoning the idea that this is a simple matter—and that Evaluative Dispositionalism can guide us successfully in deciding which individual cognitions deserve praise or blame.

My suggestion here is that it is no easy matter to "check the dispositions associated with so believing" (ibid., 12). We can never rest content that we have looked at all the relevant dispositions in order to correctly apportion praise and blame. We can certainly praise or blame people for the dispositions that lead up to their beliefs or stereotypes. And we can praise or blame them for particular acts which flow from those beliefs or stereotypes. But we cannot possibly know enough about all the dispositions that they have in order to properly apportion praise and blame for those. Moreover, I will suggest that it will often be impossible to single out particular cognitions for praise or blame. I have two broad reasons for suggesting this: the very many dispositions that there are, and the difficulty of singling out a particular cognition as responsible for any dispositions.

2.1 Very many dispositions

The concern which I call *Very Many Dispositions* is that any act of stereotyping or believing something will give rise to a huge range of dispositions, arguably an infinite number of them. At any rate, the number

is so great that we can never reach any firm evaluation. Consider, for example, our evaluation of Albert's false belief as having more epistemic benefits than costs. This seems right, given the dispositions that we considered. But now let's add some more. Suppose that if Albert were to be asked to approve funds for a study of the under-representation of women in science. This would have huge costs not just in terms of the epistemic injustice done to the person suggesting this, but also in terms of the potential loss of the knowledge which could be gained by such a study. However, things might not go that way. Rather than being asked about implementing such a study, Albert might be asked to make a huge number of hiring decisions, thereby increasing the benefits that come from his lack of tendency to stereotype women as less expert at science. Or perhaps Albert might leave science for a career as a bartender, and have no notable further dispositions arising from his beliefs about gender and science. Any of these things might happen, and Albert would have dispositions with respect to each of them. And I have only begun to scratch the surface. Any evaluation of Albert's belief in terms of the *very* many dispositions it gives rise to begins to look pretty unfeasible.

2.2 Which cognition?

It is also important to note that no cognition gives rise to dispositions completely on its own. Any act of believing or act of stereotyping is carried out by a person with various other beliefs, dispositions, habits, skills, preferences, and the like. This point has already been touched on, by noting that one thing a person might do if they realize that there are more men than women in science is take action to prevent themself from acting automatically on the basis of this stereotype. Now consider the case of Betty, who believes that men are more likely than women to have scientific expertise. This belief gives rise to automatic associations which cause her to make all the bad epistemic moves that Albert avoided, leading Betty to underrate the competence of women scientists that she encounters, and making her more likely to hire men than women as junior scientists. It may seem obvious at first that this is a case in which Betty's stereotyping belief is at fault—despite its truth, it has a large range of negative epistemic consequences and should receive a poor epistemic evaluation on an Evaluative Dispositionalist account. So far so good for that account.

But now imagine something further. Betty has been to an equality and diversity training session about implicit bias. She knows that there is a real risk of her true belief giving rise to automatic stereotypes which caused her to underestimate the worth of women scientists. The trainers were extremely skilled, and explained a variety of techniques that Betty could use to try to keep her true belief about gender and science from giving rise

to biased future cognitions. Betty was not interested in doing this and decided to shred the helpful handout that she was given. Let's also imagine that Betty's colleague, Caleb, reacted very differently to the training. He held the same belief as Betty about the relative likelihood of men and women having expertise in science, but he was horrified to learn of the consequences this might bring. He decided to work very hard to keep this belief from bringing with it future biased cognitions. He kept the handout, did further research, and tried out all the techniques that he could find. Through these efforts, he succeeded in blocking that belief from bringing about biased cognitions. After all of this is in place, it starts to seem quite strange to place the blame for the biased cognitions on the belief about the frequency of men and women in science. It seems as though the real blame should focus instead on Betty's belief that it isn't worth doing anything about future biased cognitions, on her decision not to take any action to combat future biased cognitions, and so on.

Puddifoot herself would readily admit that these two different cases deserve different judgments. Indeed, one advantage that she cites of Evaluative Dispositionalism is the "ability to explain how two people with same stereotype, believed under same circumstances, may deserve different levels of praise and blame" (2021, 12). However, crucially: for Puddifoot, these different levels of praise and blame attach *to the act of believing the stereotype*. My question here is whether that's really what the praise and blame should attach to. My suggestion is that in real cases, once we spell out enough details to understand the differences between the believers involved, it becomes difficult to single out any one cognition as the proper target for praise or blame. It seems more likely to me that, to put it in a Quinean way, our cognitions face the tribunal "not individually, but only as a corporate body" (Quine 1951, 38).

Let's consider another example to motivate this thought. Imagine that Dorinda is a woman scientist. She is well aware of the under-representation of women in science, and indeed quite devoted to fighting it. As a result, she is aware that men receive unfair advantages relative to women, and has adopted the policy of assuming that where a man and a woman in science look equally well-qualified on paper, the woman is actually more qualified—since she has managed these achievements despite the barriers of being a woman in science. Now consider Edith, who is just like Dorinda except for one thing: Edith is also very aware of the under-representation of Black men in science. When Edith encounters a man and woman in science who look equally well-qualified on paper, she also takes time to think about the man's race before concluding that the woman will have faced more barriers. Dorinda and Edith will behave quite differently when they encounter the CV of a Black man in science. Dorinda will falsely

assume that he is unlikely to have encountered many barriers, since he is a man. Edith will assume that he is very likely to have encountered barriers, since he is a Black man.

Once more, Puddifoot can certainly capture the difference between these cases. She can say that the stereotyping belief held by both Dorinda and Edith leads to different results in the two cases, and assess it as epistemically problematic in Dorinda's case but not Edith's. But once more, it seems to me puzzling to single out the belief in this way. It seems like we capture the situation better if we look at the whole picture of Dorinda's and Edith's beliefs. When thinking about praise and blame, surely we should focus on their attention or lack of attention to race, rather than just on their shared belief about gender.

My thought here is that what dispositions people have arising from any belief depends on many other facts about them—including, crucially, what other things they believe. Given this, I wonder why an Evaluative Dispositionalist should want to single out a particular belief for evaluation.

3. Conclusion

The *Very Many Dispositions* worry makes me wonder how an Evaluative Dispositionalist could ever confidently make a judgment about whether someone's dispositions are on balance good or bad. The *Which Cognition?* worry makes me wonder whether an Evaluative Dispositionalist should ever single out a particular cognition for evaluation anyway. Together, these lead me to the thought that perhaps Puddifoot's arguments could lead us to an even more surprising place—one where we can't assess the epistemic worth of any individual belief or stereotype. We could still discuss the epistemic merits of how it was formed, and discuss the epistemic consequences that this belief or stereotype has for a particular person in a particular situation. But overall epistemic evaluations of stereotypes, or even acts of stereotyping, would be something that we should try to avoid.

Acknowledgments

I'm very grateful to Marina Trakas for organising this symposium and special issue, and to Leonie Smith for helpful comments on this paper.

REFERENCES

Bortolotti, Lisa. 2020. *The Epistemic Innocence of Irrational Beliefs*. New York: Oxford University Press.

Puddifoot, Katherine. 2021. *How Stereotypes Deceive Us*. Oxford: Oxford University Press.

Quine, Willard Van Orman. 1951. "Two Dogmas of Empiricism Symposium." *The Philosophical Review* 60: 20–43

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 1 | 67-90
https://doi.org/10.31820/ejap.21.1.5

REPLIES TO CONTRIBUTORS

Katherine Puddifoot[1]

[1] Durham University, United Kingdom

Discussion – Received: 01/11/2024 Accepted: 25/01/2025

This paper is part of a book symposium on Katherine Puddifoot's *How Stereotypes Deceive Us* guest edited by Marina Trakas (IFF, CONICET)

ABSTRACT

This paper provides responses to the 4 commentaries by Federico José Arena, Leonie Smith, Federico Picinali, and Jennifer Saul under the main headings: "Definition of stereotypes"; "Single/dual factor view", "Epistemic benefits of egalitarian beliefs", "Beyond stereotyping beliefs", "Which disposition?", "More radical implications of evaluative dispositionalism", "Stereotypes, reality and testimonial injustice", "Normative stereotypes", and finally "Moral encroachment".

Keywords: epistemic benefits; dispositions; stereotypes; single/dual factor view.

Correspondence: katherine.h.puddifoot@durham.ac.uk

Introduction

How Stereotypes Deceives Us (HSDU) primarily aims to provide a characterisation of the ways that stereotypes lead to misperceptions and misunderstandings of people and events. It argues that stereotypes can have this negative impact in various ways, including when a stereotype reflects an aspect of social reality. I defend a multifactorial approach to stereotyping, according to which multiple factors determine whether any act of stereotyping increases or decrease the likelihood of a misunderstanding or misperception. I develop a view that I call *evaluative dispositionalism*, which says that given the multiple ways that stereotypes can deceive us, we ought to consider, when evaluating any act of believing a stereotype, both the dispositions displayed when acquiring the belief and the dispositions that a person acquired by believing. I articulate implications of the multifactorial view and evaluative dispositionalism for when stereotyping is ethically wrongful, the ways that medical decision-making should be conducted, and for how we should approach any decision about whether to disclose information relating to a stigmatized social identity.

HSDU attempts to integrate a wide-ranging literature, drawing from sources not based on their disciplinary background but instead their quality and relevance to the topics under discussion. Given the breadth of the topic, I could not hope to do justice to, or integrate insights from, all relevant literature, but my hope is that the book might stimulate further discussion—within my home discipline philosophy, but also perhaps more broadly—about the nature and potential epistemic pitfalls associated with stereotyping. This symposium is therefore pleasing because the contributions represent careful and considered responses by experts from inside philosophy and beyond. Below I outline responses by theme rather than by author because in some cases the arguments presented by contributors dovetailed. I aim to show how my ideas have developed in response to symposium contributions.

1. Definition of stereotypes

Let us begin by focusing on the definition of stereotypes. In HSDU I provide a defence of a non-normative conception of stereotypes and stereotyping, that does not define stereotypes as false or stereotyping as misleading (for other defences of a non-normative approach, see Beeghly 2015; Fricker 2007; Jussim et al. 2012; Kahneman 2011). My reason for adopting this approach is pragmatic: to define stereotyping in a way that emphasises the continuity between the mental states involved

with stereotyping and other cognitive states, rather than assuming a discontinuity, with stereotypes separated from other biases like heuristics on the basis that they are always false and misleading (I am influenced here by Ashmore and Del Boca 1981 and Beeghly 2015). In addition to this, I argue that stereotypes are necessarily comparative: they involve a comparison between social groups, suggesting that members of one social group are more likely than some others to possess a particular trait or traits. Here is the definition of stereotypes:

> Social attitudes that associate members of some groups more strongly than others with a certain trait or traits. (HSDU, 13)

Saul raises a potential counterexample to this definition. Suppose someone associates people who teach in Durham, a city in the North of England, more strongly than others with teaching in the North of England. Saul suggests that this association could count as a stereotype on my definition, but that this doesn't seem right. I agree that an association between people living in North of England and teaching in Durham does not seem at face value to be a stereotype. So, how should I respond?

There are at least three general options available to me. The first is *bullet biting*. Although it might seem strange to classify this as an example of stereotyping, people's intuitions about what counts as a stereotype are inconsistent. As I discuss in the book, some people seem to intuitively endorse a normative account of stereotyping, according to which stereotypes are always false and misleading (cf. Blum 2004). Others seem to be open to saying that stereotyping can be useful, and necessary, because stereotypes can be accurate heuristics (cf. Beeghly 2015; Fricker 2007; Jussim et al. 2012; Kahneman 2011). Any specific definition of stereotypes is not going to satisfy everyone's intuitive judgements, and so will require some adjustments to classificatory practices. It might be that one appropriate adjustment to classificatory practices is to accept that attitudes like the one associating living in the North of England with teaching at Durham can be stereotypes.

A second option would be to maintain the current definition, but stipulate that definitions or definition-like propositions cannot be stereotypes, for example:

> Social attitudes that associate members of some groups more strongly than others with a certain trait or traits *but are not definitions or definition-like*.

Why might one make this move? The thought is as follows. It might seem that one is not stereotyping when more strongly associating people who teach in Durham with working in the North of England because working in Durham, i.e. that characteristic that makes them a member of the target group, by definition involves having the attribute that is ascribed by the stereotype, i.e. working in the North of England. Or, more precisely, it might be said that the characteristic that makes the person a member of the target group almost by definition involves having the attribute ascribed because there may be some exceptional cases, such as people who work in Durham but wholly online, who have never been to Durham, and so forth. If one were to say that people who teach in Durham work in the North of England one would be saying something that is almost true by definition. By stipulating that propositions that are definitions or definition-like are not stereotypes, one could thereby avoid accepting that the target proposition is a stereotype. (Similar cases include: French people are more likely than others to be born in France, or Irish passport holders are more likely than others to have been born on the Island of Ireland or have parents who were born there.)[1]

A third potential response is to say that the example of the Durham teacher does not actually meet my definition of a stereotype because the association about where someone works is not a social attitude. Take the following definitions of social attitudes: "a psychological tendency that is expressed by evaluating a particular entity with some degree of favour or disfavour" (Eagly and Chaiken 1993, 1). Or the definition of social attitude given by the *APA Dictionary of Psychology*:

1. A person's general outlook on social issues and approach to his or her responsibilities
2. A person's general disposition or manner towards other people (e.g. friendly or hostile)
3. An opinion shared by a social group (2023).

Social attitude, as I intend to use the terminology, is closest to 2 and 3. The attitude involves having a disposition towards people or an opinion of them, and it is a tendency to view people in a way that might favour or disfavour them by associating them with certain characteristics. It might be said that believing that someone is likely (or more likely than others) to

[1] If this option is taken, it will be important to think carefully about what is taken to be definition-like. For example, we would want to leave room for certain social attitudes that some people take to be definitional or definition-like to be stereotypes. For example, some people might think that women are by definition nurturing, but it does not seem that their viewing this attitude as definitional excludes it from being a stereotype. Consequently, it seems we should not like to say that what counts as a definition or definition-like is down to the beholder.

work in the North of England due to their working in Durham is not a disposition towards them, an opinion of them, or a tendency to view them in a way that may favour or disfavour them. Merely associating someone with working in a general geographical region based on knowledge that they teach in a more specific geographical region is not in and of itself having a disposition or an opinion towards them that might favour or disfavour them. This belief might lead to further opinions and positive or negative attitudes downstream, but the association itself does not represent an opinion or positive or negative attitude. Thus, it might be said that the association between teaching in Durham and working in the North of England is not a stereotype because it is not an expression of a disposition towards a person or people, an opinion of them, or a tendency to view them in a way that favours or disfavours them.

I am inclined to opt for a combination of the second and third of these responses: the association of Durham teachers with working in the North of England is not an expression of an opinion, a disposition towards a person or people, or a tendency to view a person or people in a way that will disfavour or favour them. Instead, it is a definition-like statement of a perceived evaluatively neutral fact. To be in the category of teachers who work in Durham one has, almost by definition, to work in the North of England. What consequence does this have for my definition of stereotypes? It may be altered as suggested above to stipulate that definitions and definition-like propositions are not stereotypes, but a large amount of the work of distinguishing cases of stereotyping and non-stereotyping comes from the stipulation that stereotypes are social attitudes.

2. Single/dual factor view

In HSDU I defend a multifactorial view of stereotyping, according to which there are multiple features of any act of stereotyping that determine whether the application of a stereotype is likely to increase or decrease the chance of an accurate judgement being made. In my defence of the multifactorial view, I reject two alternative views of stereotyping: the single and the dual factor view. According to the single factor view, a stereotype is likely to increase the chance of an accurate judgement being made when it is applied so long as it reflects an aspect of social reality. According to the dual factor view, the accuracy of the stereotype and the availability or lack thereof of unambiguous evidence determines whether the application of the stereotype increases or decreases the chance of an accurate judgement being made. My challenge to these views involves arguing that there are various other factors that determine whether an act

of stereotyping increases or decreases the chance of an accurate judgement being made. These factors include whether the stereotype is relevant in the context in which it is applied, and many other factors relating to how the person stereotyping accesses and processes case specific information (ambiguous or not) when engaging in stereotyping.

In this section, I am going to respond to some worries raised by Picinali and Arena in response to my critique of the single and dual factor views.

2.1 Critiquing the single factor view

When it comes to the single factor view, Picinali argues that this is so implausible as to be a straw man, and that no theorist would claim that stereotyping is likely to increase the chance of an accurate judgement being made. For example, it is a known fact that in racist societies characterised by racist police force Black people are overrepresented among those arrested for crimes like drug crimes, and this known fact means that people will be aware that a stereotype might reflect an aspect of social reality (i.e. arrest rates) but be unlikely to lead to a correct judgement. While I absolutely agree that there are facts like those gestured towards by Picinali that suggest that the single factor view is incorrect, it is nonetheless important to spell out a view like this and to challenge it. This is because it is likely that many people would defend specific stereotypes and acts of stereotyping precisely on the basis that there is a link between the stereotype and some aspect of social reality. When working on this type of issue, it is not enough to engage only with those who would develop well-formulated arguments with respect to stereotyping, it is important to engage with all those arguments that are likely powerful and widely endorsed. Picinali is right that the single factor view is weak and implausible, but to persuade me it is not worth considering it would be necessary to convince me that the argument has little or no power in society. Sadly, I suspect that this is untrue. Instead, my suspicion is that this type of position implicitly underlies much thought about stereotyping. As such, it is crucial to consider and reject it. In addition to this, the single factor view is used strategically as a jumping off point to develop the more plausible multifactorial view.

2.2 Critiquing the dual factor view

In discussion of the dual factor view, Picinali critiques my attempt to define what it is to be influenced by an accurate stereotype. I speak about a stereotype being accurate if it makes someone respond in a way that is fitting with accurate statistical information. For Picinali this reads as saying that the accuracy lies not in the generalisation but in the judgement

produced. Given that the dual factor view (and my general project at this point in HSDU) is to say when stereotypes are likely to lead to accurate and inaccurate judgements, a definition of an accurate stereotype as one that produces an accurate judgement would turn my claim that accurate judgements are more likely to be produced when accurate stereotypes are applied into a tautology.

To clarify, I do not mean to define accurate stereotypes as those that produce accurate judgements, but instead as those that reflect statistical information in a way that is accurate (for an understanding of what this might look like, see the work of Jussim and colleagues (2012) who argue that nearly all stereotypes reflect statistical information), leading people to make judgements that are reflective of that information. On my view a judgement could reflect background statistical information, via a stereotype, without being an accurate judgement. This is one of the main claims of the book: that even a person who makes a judgement that is in line with background statistical information (i.e. base rates) can nonetheless make errors, in particular, errors that are the result of their judgements being shaped by the background statistical information.

A second criticism that Picinali raises against the dual factor view relates to claims about relevance. One of my claims is that neither the single or dual factor views gives recognition to the ways that stereotypes can be, and seem likely to frequently be, applied when they are not relevant. Picinali criticises my work for not giving a definition of relevance. I was working with a commonsense understanding of relevance as meaning, roughly, pertaining to the truth of a judgement. Picinali argues that a common sense understanding of relevance does not seem to apply to at least one of my key examples of how stereotypes can be applied when not relevant:

> A police officer approaches the car of a Black male, which has been pulled over for a minor traffic violation, e.g. one of his headlights is not working. The police officer asks the man to step out of the vehicle but he responds slowly and cautiously to the command. The police officer is offended at what he takes to be a threat to his authority. This triggers a stereotype associating the innocent man with crime; the police officer evaluates the man as a criminal and treats him with hostility; and this leads to an escalation of tension and hostility between the two individuals. The stereotype associating Black people with crime is triggered although the Black man has not committed a crime, only a minor traffic violation (HSDU, 46)

Picinali argues that the stereotype could be relevant in this case because after stopping the car the police officer may become suspicious that the driver is or will commit a crime. Then the stereotype about social group and criminality would be relevant. I put aside the latter point about whether the stereotype becomes relevant if a suspicion arises about there being a high chance that the driver would engage in criminality. I have serious doubts about this: why would the stereotype be relevant when the police officer has access to case specific information about the driver? I do not need to develop this doubt any further, however, because there is another problem with this critique: it requires adding detail to the example and thereby changing it. The example is designed to show that there are likely to be many cases in which a stereotype is triggered when not relevant. This point is backed up by psychological findings speaking to the question of when stereotypes tend to be triggered (i.e. when there is a challenge to the status quo or someone has a wounded ego) and the example, as presented. In the example, a Black person is stopped for a traffic violation. The stereotype associating Black people more strongly than white people with criminality does not pertain to the truth of the judgement of the driver's situation. It is not relevant. We can try to construct similar cases, adding detail to show how in different cases stereotypes might be relevant. However, this does not undermine the point that there are conditions, seemingly many, in which a stereotype is likely to be triggered but not relevant on any commonsense understanding of the term.

Arena raises a similar, but importantly distinct, worry about the idea that relevance is a factor that determines if stereotyping helps or hinders judgement that is not captured by the single or dual factor view. Arena provides three potential definitions of relevance:

> (1) Relevance as meaning that the individual is within the scope of the stereotype.
> (2) Relevance as "non-spurious statistical correlation".
> (3) Relevance as "argumentative force".

Arena suggests that the dual factor view captures (1) and (2). So if by focusing on relevance I meant to show that stereotypes can be applied to people outside the scope of the stereotype, or when there is at best a spurious statistical correlation between group membership and a target characteristic then I would not be presenting a challenge to the dual factor view. In fact, I did not mean to refer to (1) or (2), but instead, as mentioned above, something like "pertains to the truth of". My point is that a person may fall within the scope of the stereotype, in the sense that they have the social identity that the stereotype makes an association with, and the stereotype can still lead the person stereotyping astray independently of

whether there is a non-spurious correlation between being a group member and having the characteristic. Even if there were a non-spurious correlation between being member of group G and having trait T, and person x falls into category G, the stereotype associating G with T may be applied to x under conditions where the stereotype is uninformative, that is, the stereotype does not pertain to the truth of whether G is displaying characteristic T. What I am getting to is closest to (3): the idea that the stereotype does not have force in the context in which it is applied. But it is not argumentative force, as in persuasive force, but instead something more like evidential force, whether evidence pertains to the truth of the matter at hand.

One final point on the dual factor view. I argue that there are various ways that stereotypes can deceive us by leading to a distorted response to case specific information. In the terminology I use later in the book: there are various ways that stereotypes dispose us to respond poorly to information that is available in our environment. Arena suggests that the dual factor view can accommodate this thought:

> [G]iven the emphasis that the Dual factor view puts on case-specific information, (…) the cases of stereotyping pointed out by the examples would be considered by the Dual factor view as lacking epistemic quality, given that the ambiguity of individual information was not adequately ruled out. (Arena this issue, 18)

While I appreciate that the effects that I describe are ways that the information about an individual is not adequately assessed when stereotyping occurs, my aim in presenting the multifactorial view as an alternative to the dual factor view is to emphasise that even if one's external environment provides high quality information that information might be hard to access and properly process precisely because one harbours a stereotype.

3. Epistemic benefits of egalitarian beliefs

One of the main goals of the book is to outline how stereotypes can lead us to misperceive and misunderstand people and events involving people. As such, via the multifactorial view, the book outlines various epistemic pitfalls that are associated with stereotyping. Building on this picture of the various epistemic pitfalls that are associated with harbouring stereotypes, I argue that having false egalitarian beliefs can bring overall epistemic benefits, outweighing the epistemic costs of their falseness. The egalitarian

beliefs bring epistemic benefits because they prevent people from being subject to the pitfalls associated with stereotyping. Picinali and Arena both object to the idea that egalitarian beliefs guarantee epistemic benefits, and the avoidance of the epistemic costs. Picinali gives the example of someone who endorses an attitude according to which both men and women are unlikely to have scientific experience. The attitude may lead the person to make epistemic errors, e.g., misinterpreting ambiguous information, misremembering scientific achievements, etc. For Arena,

> The point here is that the distortion of case-specific information is a consequence of basing the assessment of individual traits on a generalisation, regardless of whether it has an egalitarian or non-egalitarian content. (Arena this issue, 20)

These are interesting and important points. In reply, I would first re-iterate the strength of the claim that I defend. It is that *often*, and likely *more often than not*, the epistemic benefits of a false egalitarian attitude will outweigh the epistemic costs of the falsity of the attitude. For example, for some people it may be better from an epistemic perspective to falsely believe that men and women are equally likely to have scientific expertise. I think that this is plausible because stereotyping brings the risk of a raft of epistemic errors, including memory errors, misinterpreting ambiguous evidence, assuming similarities when they do not exist, failing to notice similarities when they do, failing to give adequate credibility to testimony, etc. Having an egalitarian attitude can reduce the risk of at least some of these errors, thereby increasing the chance of true beliefs, understanding, and so forth. What I do not claim is that egalitarian attitudes guarantee that every one of the epistemic errors will be avoided, or that the egalitarian attitudes cannot bring similar errors. So, it is compatible with my view that egalitarian attitudes can bring epistemic costs and that the egalitarian attitudes do not guarantee the avoidance of epistemic error.

Nonetheless, it would be remiss not to acknowledge how Picinali and Arena's comments have challenged me to think further about how egalitarian attitudes can bring epistemic costs. It certainly seems right that if you are a committed egalitarian, you may be prone to interpreting ambiguous evidence in a way that is consistent with your egalitarian beliefs, for example.

However, when considering other epistemic pitfalls associated with stereotyping, it seems less plausible that egalitarian beliefs will have the same negative epistemic effects as stereotypes. This is because the negative effects of the stereotyping are specifically associated with the process of categorisation of individuals into different social groups. An

egalitarian judgement that does not encourage seeing members of the groups differently seems less likely to have the same effects. Take, for example, the way that classifying someone as a member of a minority group can lead them to be viewed as more similar than they really are to other members of the same group (Bartsch and Judd 1993; Hewstone, Crisp, and Turner 2011). Or take the way that those classifying individuals as a member of one group (e.g. women scientists) may fail to notice similarities between those individuals and members of other groups (e.g. men scientists) (e.g. Tajfel 1981). It seems unlikely that there will be similar effects that occur because of egalitarian beliefs that emphasise similarities, or include claims that apply to all relevant groups (e.g. both men and women who are scientists).

Something similar can be said about the relationship between stereotyping and memory (cf. Arena this issue). As I discuss in my HSDU, psychological studies suggest that people often remember information consistent with a stereotype better than other information, and this effect is explained in terms of the role of social schemas or expectancies (e.g. Rothbart et al. 1979; Fiske and Linville 1980). It is argued that information that is consistent with a social schema is often more easily stored and retrieved from memory. Where there is a stereotype relating women to being less likely to have scientific expertise than men this will be a part of the social expectancies or social schema "WOMAN". This social schema is likely to be activated in response to women scientists, and to shape the way that information about them is processed and stored. It is far from clear, in contrast, that if someone has the egalitarian attitude that both men and women are likely to lack scientific expertise, that lacking scientific expertise will be a part of the social schema for either men or women. Or that the social schema of "MAN" or "WOMAN" will be activated in response to any individual scientist, leading to schema-consistent memory effects. What seems more likely to happen is that all scientists will be viewed as exceptions to the general rule that men and women are unlikely to be good at science. For one final example, in HSDU, I discuss Kristie Dotson's (2011) work on testimonial smothering, that is, on the way that people may choose to suppress risky testimony, especially if they fear that it will sustain or compound existing stereotypes that they take others to possess. There seems to be good reason to think that a woman is likely to suppress risky testimony about a scientific matter to prevent themselves from compounding the stereotype that women are less likely to have scientific expertise than men. In contrast, there seems to be significantly less risk of someone choosing not to speak about a scientific matter for fear of compounding the stereotype that men and women are both unlikely to have scientific expertise.

In sum, then, egalitarian attitudes do not appear to bring the same risk of epistemic pitfalls as stereotypes because many of the pitfalls of stereotyping seem to be closely tied to features of social categorisation. Because there are some epistemic pitfalls more strongly associated with stereotyping than egalitarian beliefs, false egalitarian beliefs are likely to often bring more epistemic benefits than stereotypes that they would replace. This is all that is needed to support my claim that *often,* and likely *more often than not*, having a false egalitarian belief can be best from an epistemic perspective.

4. Beyond stereotyping beliefs

Although I maintain that false egalitarian beliefs can be better from an epistemic perspective because they can avoid significant epistemic costs, the idea that egalitarian beliefs can also dispose people to fail to respond appropriately to information that they encounter downstream brings me to my next point.

In HSDU I present a new approach to evaluating stereotypes and stereotyping along the epistemic dimension: evaluative dispositionalism. Evaluative dispositionalism encourages people to focus on the epistemic dispositions, that is, dispositions to respond one way or another to evidence, that a person has displayed when acquiring a stereotype *and* those epistemic dispositions that a person has due to harbouring the stereotype. But, as pointed out by Picinali, many of the examples that I use in HSDU to illustrate the nature of an epistemic disposition are unrelated to stereotyping (e.g. I mention how a person's beliefs about a football team can shape their responses to evidence about their performances). Picinali's observation points towards a broader ambition that I have: to apply evaluative dispositionalism to other, non-stereotyping, beliefs. For example, I think that the evaluative dispositionalist approach could be fruitfully applied to the core beliefs held by members of echo chambers; that it would be valuable to evaluate both the epistemic dispositions that people have displayed in entering the echo chamber and the way that they are disposed to respond to evidence once they have the core echo chamber beliefs. The evaluative dispositionalist approach could also be applied to egalitarian beliefs, where those dispose people to respond poorly to evidence.

Considering the issues raised by the commentators on my book in this symposium (see sections 5 and 6 below), it would be interesting to also consider whether we should move away from evaluating the epistemic standing of individual beliefs, towards considering how individual beliefs

interact with other beliefs, situational factors, and personality types to dispose people to respond well or poorly to evidence.

5. Which disposition?

In my defence of evaluative dispositionalism in HSDU I suggest that it would be useful advice to "check your dispositions" in relationship to stereotyping—rather than merely considering how a stereotype is formed, one ought to consider how one is disposed, due to possessing the stereotype, to respond to evidence about individual people or events. This is because people can come to harbour stereotypes in better or worse ways—displaying poor dispositions or merely experiencing the misfortune of being in a hostile environment—but people can also be disposed to respond in better or worse ways to relevant information once they harbour the stereotype.

Saul suggests that it will be difficult to apportion praise or blame to people on this type of approach because we cannot possibly know what dispositions a person has. This problem is especially acute given how many dispositions people have in relation to any specific belief. Saul calls this *the very many dispositions argument*. The very many dispositions argument highlights a serious issue. It will not always, or perhaps often, be possible to be sure about all dispositions that someone has because of believing any proposition (or harbouring any implicit stereotype non-propositionally). It will be difficult in any specific case to identify how a person is disposed to respond due to the stereotypes that they harbour, especially to the level of certainty that might be required for justified praise or blame. Nonetheless, I believe that there are good reasons for thinking that it is still valuable to focus on dispositions.

First, if you are aware of your own tendency to stereotype, or that you are likely to stereotype, you can make efforts to reflect on whether stereotyping is likely to be leading you astray in *any specific context*. It is not necessary to be able to identify all dispositions that might be manifest in any context to do this, so the very many dispositions argument does not come into force. Notwithstanding the shortcomings in people's awareness of what stereotypes they harbour and when stereotypes are likely to be triggered, people who are informed by relevant psychological findings, theoretical accounts of stereotyping, personal accounts of being on the receiving end of stereotyping, etc.—the types of information discussed in HSDU—can reflect on the likely effect of stereotypes on the dispositions that they are may display in a particular context. For example, they can reflect on why they noticed certain information, whether there was other

information they did not notice, whether there are gaps in their memories, and so forth. My view encourages and supports people to engage in this type of reflection on the dispositions that they might display by highlighting the types of dispositions that stereotypes bring.

Second, a focus on the dispositions that *other people* are likely to display in any specific context due to stereotyping can lead to appropriate skepticism towards the beliefs that they articulate relating to groups that they are likely to stereotype. One does not have to know for certain which dispositions a person holds to factor in that they might be misled by stereotypes, and that it is therefore worth seeking out evidence that could confirm or disconfirm whether they are being misled. It can be useful to be aware of the types of dispositions that a person might have due to stereotyping, to probe them to test for likely effects. For instance, let us say that someone claims that a member of a minority group in your workplace has not been contributing as they should. One might seek out information about what specifically that person has noticed and remembers in relation to the minority group member, the context in which they have interacted, and in which their memories about the minority group member were formed. This information can provide guidance about whether the person is likely to be displaying a disposition not to properly encode, process and recollect information about the member of the minority group, due to stereotyping.

The argument here has similarities to one that Saul (2013) provided when defending what has become known as "Saulish skepticism" (Antony 2016) in relation to implicit bias. Saul argues that evidence of implicit bias provides reason to adopt a skeptical attitude towards beliefs about social actors and objects because those judgements could easily, unbeknownst to the person judging, be influenced by irrelevant and distorting implicit stereotyping. I, like many others (e.g. Fricker 2007; Antony 2016), emphasise that stereotypes can operate, either explicitly or implicitly, to supply information that is relevant to a judgement. However, I also argue that stereotypes—even those that reflect reality—can dispose us to respond poorly to evidence. This suggests that we ought to adopt a skeptical stance towards beliefs that may have been shaped by stereotypes, carefully considering how stereotypes might have operated to influence how the beliefs have been formed by shaping the dispositions of the believer.

Third, while it will be difficult to discern for any individual which dispositions they possess due to stereotyping, there is an important lesson to be learnt from this problem. The lesson is this: it is important to be aware that information is not equally safe in anybody's hands. Certain information about social groups can be useful and informative in the hands

of some people while damaging in the hands of others. The main difference-maker could be the dispositions that those people possess due to harbouring the stereotype. Not only this, it may be difficult or impossible to know all of the dispositions that a person could harbour due to stereotyping. It is therefore important to be cautious about whom to trust in relation to specific information—for example, information about a person's mental health condition—because one cannot be sure how the information may lead any specific person to be disposed to respond to evidence that they might encounter.

6. More radical implications of evaluative dispositionalism

As mentioned in sections 4 and 5, according to the view proposed in chapter 8 of HSDU, evaluative dispositionalism, one ought to evaluate a stereotyping belief by considering both the dispositions displayed forming the belief and the dispositions possessed due to harbouring the belief. This pluralistic approach involves considering both the causal history of the stereotyping belief (the dispositions displayed when coming to possess a belief), and the consequences of believing (the dispositions held as a result).

Saul does not object to the proposal that it is worthwhile considering the causal history and consequences of stereotyping, but suggests that my argument implies something more radical than evaluative dispositionalism, which focused on the causal history and consequences *of a particular stereotyping belief*. For Saul my arguments suggest a more intriguing possibility: that we should not be epistemically evaluating any single belief on its own.

Saul gives the example of two individuals, Betty and Caleb, both of whom take a training session on bias. While Caleb is attentive, takes a handout, reads it carefully, and attempts to prevent a specific stereotyping belief from influencing his cognition downstream, Betty does not even look the handout. For Saul, it is not any stereotyping belief alone that is an apt object of epistemic appraisal. Other beliefs, such as the belief about whether it is worthwhile looking at the handout, are also apt, and perhaps even more apt, for praise or criticism. Saul also gives the example of Edith and Dorinda, who both believe that if a man and woman seem equally qualified on paper the woman is in fact better qualified because she will have achieved what she has despite facing barriers that the man has not faced. Although Edith and Dorinda are similar in this way, Edith but not Dorinda understands the barriers faced by black men. This leads Edith but not Dorinda to factor in a man's race when considering the barriers faced

by the two individuals. Saul takes both these examples to show that "what dispositions people have arising from any belief depends on many other facts about them—including, crucially, what other things they believe" (Saul this issue, 64), and, furthermore, to suggest that when it comes to stereotyping it is not just a specific stereotyping belief that ought to be evaluated.

A similar thought emerges from Picinali's contribution. Picinali takes one of the examples from the book, of Nora and Ned, to show that the focus on any specific stereotyping belief seems to narrow. In the example, both Nora and Ned harbour the stereotyping belief that men are more likely than women to have scientific expertise. However, only Ned makes a catalogue of errors due to the stereotype: e.g. memory errors, misinterpreting ambiguous information, and so forth. Because both individuals harbour the stereotype, Picinali argues, the stereotype itself can only have a limited causal role in determining the dispositions, otherwise both Nora and Ned would display the same dispositions. The causal story is more complex than evaluative dispositionalism suggests.

These comments have been extremely helpful in clarifying my thinking about stereotypes and dispositions. One of the main motivations behind the HSDU project was to show that a simplistic picture of stereotyping cannot work because of the variety of ways that believing a stereotype can shape our cognition as well as our action. There will be individual differences in the way that stereotypes shape responses to evidence. Some of these differences are due to situational factors, e.g. time, cognitive load, amount of information to process, others relate to personality, and others further to the wider set of beliefs that are held. Given this, it seems right, as Saul and Picinali suggest, that stereotyping beliefs should not be considered and evaluated in isolation. An epistemic evaluation of an act of stereotyping should consider how the stereotyping belief is likely to lead a specific person to be disposed to respond to information given other facts about them, including the other beliefs that they hold. This is not to diminish evaluative dispositionalism as an approach to stereotyping. But it is to acknowledge that it may be more important than HSDU has emphasised to consider other factors that have a causal role in determining how stereotypes deceive us, and what they mean for how we might be disposed to respond to evidence. As Saul suggests, my argument for evaluative dispositionalism provides some support for the radical idea that when epistemically evaluating people's stereotyping, we ought not to only focus on any specific stereotyping belief, but a far wider range of phenomena.

7. Stereotypes, reality and testimonial injustice

Leonie Smith's contribution encourages us to delve deeper into the relationship between stereotypes that appear to reflect reality and testimonial injustice. The terminology of testimonial injustice was introduced by Miranda Fricker (2007) to capture cases where identity prejudice leads the testimony of members of marginalised groups to be systematically given reduced credibility. Fricker describes prejudice as involving an epistemically culpable failure and affective investment. In HSDU, I describe cases where people suffer harm because they are given less credibility than they deserve due to stereotypes relating to their social identity, but where the stereotyping does not seem to meet Fricker's definition of prejudice because it does not involve affectively invested epistemic culpability: the stereotype is acquired in response to the information available in one's environment. I often focus on the stereotype associating scientific expertise more strongly with men than women. A person could acquire this stereotype by being well-informed about levels of training across gender groups rather than due to any affective investment in the stereotype. In commentary on these ideas, however, Smith encourages us to think further about the relationship between stereotypes like the science stereotype and testimonial injustice. Smith suggests that, properly understood, prejudice can be viewed as having an important role in the types of case I focus on.

Smith describes several ways that these types of case could involve prejudice and/or epistemic culpability, and so be classified as testimonial injustices.

(1) The cases may be said to involve prejudice because prejudice shapes the social backdrop, e.g. levels of expertise among men and women, that makes the stereotype accurate.
(2) The people in the cases may have internalised prejudiced beliefs due to the prejudiced social backdrop, so only appear not to be prejudiced.
(3) Prejudice could be present in the inappropriate application of stereotypes in contexts where evidence suggests that they do not apply.
(4) The cases could involve prejudice because although there appears to be an accurate and well-supported stereotype in operation there is in fact an inaccurate one that is poorly supported by the evidence.

This analysis provides a more fine-grained taxonomy of cases than is provided in HSDU. It presents several distinctive ways of conceptualising

prejudice and its role in credibility assessments. It thereby points towards different ways that the condition that testimonial injustice involves prejudice could be met. However, I would add the following observations to Smith's suggestions.

First, while it is important to recognise that prejudice can form the backdrop to people having stereotypes that are accurate, there is still value in distinguishing between cases where people dismiss other people as lacking credibility due to being prejudiced, and those where people dismiss others as lacking credibility because they grew up in societies shaped by prejudice. One way to make this distinction is to stipulate that there is only testimonial injustice when prejudice directly causes the credibility deficit and not where prejudice only indirectly influences the credibility judgement by shaping society so certain stereotypes are accurate. If, on the contrary, we were to say that there is testimonial injustice in all cases where there is a backdrop of prejudice then this important distinction will be lost.

Second, while it is highly plausible that many people in prejudiced societies internalise prejudice, this does not mean that there are not also people who engage in stereotyping, and subsequently harm others by giving them less credibility, without having internalised prejudice. As discussed in HSDU, psychological evidence strongly suggests that stereotypes impact how people respond to evidence, including about people's credibility, regardless of whether those stereotyping endorse prejudicial beliefs or have any affective investment in them. It may be that these responses to evidence are indirectly due to prejudice in society, reflecting, for example, prejudicial social structures, but this does not equate to them being due to internalised prejudice. In HSDU I encourage readers to recognise the epistemic costs that can follow from this type of stereotyping, absent the direct role of prejudice. I would therefore stress the importance of not conflating the claim that some people are likely to internalise prejudice in prejudicially structured societies with the claim that all harmful credibility deficits are due to internalised prejudice.

Third, it is important to distinguish two claims: (i) prejudiced attitudes can be *reflected in* the way that people apply a stereotype, i.e. whether they apply it out of context, (ii) prejudice can be *constituted by* people applying a stereotype out of context. (i) certainly seems to be true. Prejudice can lead people to apply generalisations more broadly than they should, for example, applying the stereotype that scientific expertise is more common among men than women when it should not be applied, to a trained woman scientist. But this observation only establishes that some cases in which people apply stereotypes out of context, or when they are irrelevant, are ones where prejudice has been operational. It does not establish that in all

cases where people apply stereotypes where they are irrelevant this is due to prejudice. It therefore leaves open the possibility that some cases in which people apply a stereotype when it is irrelevant occur in the absence of prejudice having a direct role.

On the other hand, it might be argued that when people apply stereotypes out of context this *constitutes* prejudice. This would suggest that wherever stereotypes are applied out of context, like they are to the woman scientist, there is prejudice. So at least some of the examples that I use to support the claim that there can be credibility deficits in the absence of prejudice having a direct role—i.e. those where stereotypes lead to credibility deficit only because they are applied out of context—could not be used in this way. They would be cases where prejudice has a direct role. This would weaken my case in support of the claim that there are credibility deficits due to stereotyping that do not constitute testimonial injustice because there is no prejudice involved. However, making this move would also involve significantly revising any conception of prejudice that does not define it in terms of stereotypes being applied out of context, which I take would include many commonsense conceptions.

Finally, I agree there will be cases where judgements appear to be underpinned by accurate stereotypes, but inaccurate stereotypes are in fact operational. One general challenge associated with evaluating someone's acts of stereotyping is to pin down the content of the generalisation that they are applying. However, I take it that rather than counting against my position in HSDU, these observations provide additional reason to take evaluative dispositionalism seriously. Evaluative dispositionalism provides a framework that can be applied to evaluate acts of stereotyping even when it is difficult to establish beyond doubt the content of the stereotype, or whether the content accurately reflects reality. You can seek evidence about whether the person engaging in stereotyping seems to be displaying the dispositions that would be associated with stereotyping of the type that is suspected.

8. Normative stereotypes

Arena points towards an important distinction that is not covered in my book, but which is of importance to discussion of stereotyping. This is the distinction between descriptive and normative stereotypes:

> In the case of Price Waterhouse v. Hopkins, the American Psychological Association's amicus curiae noted the importance of distinguishing between descriptive and

> normative stereotypes about women. The authors of the amicus curiae claimed that: "descriptive stereotypes characterize women in a way that undermines their competences and effectiveness; normative stereotypes label women whose behaviour is inappropriately masculine as deviant". (Arena this issue, 21)

Normative stereotypes are stereotypes that dictate how people who are categorized under them ought to behave, characterising ways that it is deemed appropriate for them to behave. For example, a normative stereotype might dictate that it is appropriate for women to be co-operative. Arena points out how normative stereotypes attempt to shape behaviour and can shape the social world to fit the stereotypes. Normative stereotypes, for Arena, can have dangerous epistemic effects, as people may abandon their epistemic goals, focusing on disapproving of people who do not conform, and "constructing the facts in such a way as to make it possible to inflict some type of punishment" (ibid., 21), e.g. denying that a victim is a victim on the basis that she did not conform to stereotypical behaviour. On a similar note, Wade Munroe (2016) has argued that there can be "prescriptive credibility deficits" that occur when a speaker who fails to meet a relevant prescriptive stereotype (e.g. a stereotypically feminine speaking style) are assigned a lower level of credibility than they would be otherwise. The prescriptive aspect of stereotyping is not an aspect that I explore in the book, but I agree that it is an important phenomenon.

9. Moral encroachment

In a brief section in chapter 5 of HSDU, I compare my argument to moral encroachment views. The basic claim of moral encroachment views is that moral factors can determine what one is justified in believing, can believe rationally, or knows. Some moral encroachment theorists argue that the moral encroachment view can dissolve an epistemic-ethical dilemma that has been argued to be posed by stereotypes and stereotyping (Basu 2019a, 2020; Basu and Schroeder 2018). The purported epistemic-ethical dilemma is that where stereotypes reflect aspects of social reality—e.g. base-rate information about rates of arrest across different social groups—it can be epistemically beneficial to apply these stereotypes, but at the same time unethical (Gendler 2011; Mugg 2013). One faces a dilemma between achieving one's epistemic goals or one's ethical goals. Some moral encroachment theorists have argued that a dilemma like this does not emerge when it comes to (at least many) stereotypes: the stakes involved in situations in which people might stereotype raise the evidentiary standards, so a high level of evidence is required to be justified in

believing, to be rational, or to know. However, those engaging in stereotyping will fail to meet these standards. They will therefore not be justified because the moral stakes of the situation have raised the evidentiary standards beyond those attained by the believer. The correct thing from the epistemic perspective will be the ethical thing, i.e. not stereotyping.

In HSDU, I present an alternative approach to the epistemic-ethical dilemma. I argue that the situation that people face in relation to stereotyping is far more complex than the simplistic description of the epistemic-ethical dilemma suggests. Sometimes applying stereotypes that reflect reality can be epistemically costly rather than beneficial, because of the epistemic pitfalls associated with stereotyping. Sometimes it can be ethically required, for example, if the stereotype associates members of a particular group more strongly than others with certain medical or social conditions that they require help with. Sometimes epistemic and ethical goals conflict, but sometimes they concur, with both epistemic and ethical goals being achieved either by stereotyping or not doing so. I argue that this analysis of the complex interplay of epistemic and ethical demands of stereotyping provides things that are missed by moral encroachment views that focus narrowly on the idea that some stereotyping can never meet the high evidentiary standards set by the stakes of social situations in which stereotyping occurs.

In Picinali's contribution to the symposium he focuses on one specific idea relating to this discussion of moral encroachment. He argues that moral encroachment views can explain how some stereotyping may be epistemically acceptable, taking this to undermine the claim that my approach is preferable. Picinali applies a formal framework, initially proposed to model pragmatic encroachment views, to argue that

> [T]here will be situations in which the false negative has such a high moral cost that the threshold for outright belief (and, hence, for knowledge) will be relatively low; sufficiently low to be satisfied by stereotyping. In the medical example in which a false negative (missed diagnosis) leads to a quick death and a false positive (false diagnosis) leads to some health benefits and mild side effects, the doctor may well be morally warranted to follow a stereotype linking members of a group to which the patient belongs with the medical condition at issue (more strongly than non-members). (Picinali this issue, 51-52)

This is not the place to dig into the details of the framework Picinali proposes. It may be possible to develop a formal model that assigns a value

to different outcomes (e.g. correct diagnosis, incorrect diagnosis), reflecting the morality of the outcomes, and sets a probability threshold that determines when it is justified to act as if a proposition, i.e. a stereotype, is true. It may be that at times the values are set such that one could be said to be justified in acting as if the stereotype is true, given, for instance, the strong moral demand to achieve an outcome that this would facilitate. However, the development of this model would only get us so far in understanding the costs and benefits of stereotyping, and how we can do our best in relation to stereotyping. What my approach suggests is that acting as if a stereotype is true can at the same time bring benefits and very significant costs. Adopting a fine-grained approach to understanding stereotyping, like the one proposed in HSDU, allows us to focus on both the costs and the benefits, recognising this complexity. Ideally, it would enable people to reflect on their practices, to harness the benefits while avoiding at least some of the costs of stereotyping. A formal model that simply delivers a result that either people are or are not warranted to act as if a stereotype is true would obscure the complexity of the situation, which, I argue, needs to be faced head on.

Acknowledgments

I'm very grateful to Marina Trakas for organising this symposium and to all the contributors for their thoughtful comments on my book.

REFERENCES

Antony, Louise M. 2016. "Bias: Friend or Foe? Reflections on Saulish Skepticism." In *Implicit bias and philosophy*. *Volume 1: Metaphysics and Epistemology*, edited by Michael Brownstein and Jennifer Saul, 157-190. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198713241.003.0007.

APA Dictionary of Psychology. 2023. "Social Attitude." URL: APA Dictionary of Psychology.

Bartlett, Frederic Charles. 1995. *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.

Bartsch, Robert A., and Charles M. Judd. 1993. "Majority—Minority Status and Perceived Ingroup Variability Revisited." *European Journal of Social Psychology* 23 (5): 471-483. https://doi.org/10.1002/ejsp.2420230505.

Basu, Rima. 2019. "The Wrongs of Racist Beliefs." *Philosophical Studies* 176 (9): 2497-2515. https://doi.org/10.1007/s11098-018-1137-0.

Basu, Rima 2020. "The Specter of Normative Conflict: Does Fairness Require Inaccuracy?" In *An Introduction to Implicit Bias: Knowledge, Justice, and the Social Mind*, edited by Erin Beeghly and Alex Madva, 191-210. New York, NY, USA: Routledge.

Basu, Rima, and Mark Schroeder. 2018. "Doxastic Wronging." In *Pragmatic Encroachment in Epistemology*, edited by Brian Kim and Matthew McGrath, 181-205. New York, NY, USA: Routledge.

Beeghly, Erin. 2015. "What is a Stereotype? What is Stereotyping?" *Hypatia*, 30 (4): 675-691. https://doi.org/10.1111/hypa.12170.

Blum, Lawrence. 2004. "Stereotypes and Stereotyping: A Moral Analysis." *Philosophical papers* 33 (3): 251-289. https://doi.org/10.1080/05568640409485143.

Dotson, Kristie. 2011. "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia* 26 (2): 236-257. https://doi.org/10.1111/j.1527-2001.2011.01177.x .

Eagly, Alice H., and Shelly Chaiken. 1993. *The Psychology of Attitudes.* Harcourt Brace Jovanovich College Publishers.

Fiske, Susan T., and Patricia W. Linville. 1980. "What does the Schema Concept Buy Us?" *Personality and Social Psychology Bulletin* 6 (4): 543-557. https://doi.org/10.1177/0146167280640.

Miranda Fricker. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Gendler, Tamar Szabó. 2011. "On the epistemic costs of implicit bias." *Philosophical Studies* 156: 33-63. https://doi.org/10.1007/s11098-011-9801-7.

Hewstone, Miles, Richard J. Crisp, and Rhiannon N. Turner. 2011. "Perceptions of Gender Group Variability in Majority and Minority Contexts." *Social Psychology* 42 (2): 135-143. https://psycnet.apa.org/doi/10.1027/1864-9335/a000056.

Jussim, Lee. 2012. *Social Perception and Social Reality: Why Accuracy Dominates Bias and Self-fulfilling Prophecy*. New York: Oxford University Press.

Kahneman, Daniel. 2011. "Thinking, Fast and Slow." *Farrar, Straus and Giroux*.

Mugg, Joshua. 2020. "How not to Deal with the Tragic Dilemma." *Social Epistemology* 34 (3): 253-264. https://doi.org/10.1080/02691728.2019.1705935

Munroe, Wade. 2016. "Testimonial Injustice and Prescriptive Credibility Deficits." *Canadian Journal of Philosophy* 46 (6): 924–47. https://doi.org/10.1080/00455091.2016.1206791

Rothbart, Myron, Mark Evans, and Solomon Fulero. 1979. "Recall for Confirming Events: Memory Processes and the Maintenance of

Social Stereotypes." *Journal of Experimental Social Psychology* 15 (4): 343-355. https://doi.org/10.1016/0022-1031(79)90043-X.
Saul, Jennifer. 2013. "Scepticism and Implicit Bias." *Disputatio* 5 (37): 243-263. https://doi.org/10.2478/disp-2013-0019
Tajfel, Henri. 1981. *Human Groups and Social Categories: Studies in Social Psychology*. Cambridge: Cambridge University Press.

European Journal of Analytic Philosophy EuJAP | 2025 | Vol. 21 | No. 1 | 91-93

BOOK REVIEW

Lisa Bortolotti
THE EPISTEMIC INNOCENCE OF IRRATIONAL BELIEFS
Oxford University Press, 2020
ISBN-10: 0198863985
Hardcover: € 57
e-book: € 45.80

Dino Jelčić[1]

[1] University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

The problem of determining whether epistemically irrational beliefs can contribute to psychological or biological adaptiveness has been extensively explored. However, a similar question has until recently remained unanswered: Can a belief that is epistemically *irrational* nevertheless contribute to the pursuit and attainment of epistemic goals? Or, put differently, can epistemically irrational beliefs have a positive impact on epistemic performance*?*

In her book, *The Epistemic Innocence of Irrational Beliefs,* Lisa Bortolotti reviews a body of psychological, psychiatric and philosophical evidence that suggests a positive answer to this question. She argues that at least some epistemically irrational beliefs can be useful for promoting epistemic agency.

Epistemically irrational beliefs are defined as beliefs that are either poorly grounded in relevant evidence or resistant to best available counterevidence—including cases where they exhibit logical incoherence or empirical implausibility.

In the existing literature on this subject, irrational beliefs are rarely viewed charitably in light of their epistemic shortcomings. However, Bortolotti challenges this standard picture and makes a compelling case that epistemic irrationality should, in some cases, be tolerated—if it leads to otherwise unattainable improvements to epistemic functionality or psychological well-being. If an agent's overall epistemic or psychological condition benefits from holding such beliefs, then they should be viewed as being *epistemically innocent.*

Correspondence: djelcic@uniri.hr

For a belief to qualify as epistemically innocent, however, another condition must also be met: better epistemic alternatives must either be nonexistent or simply unavailable to the subject. This means that an irrational belief should only be rejected if a rational alternative exists that provides comparable epistemic benefits.

At this point, it is useful to illustrate Bortolotti's argument with a concrete example. Consider, for instance, distorted memory beliefs in patients with amnesia or dementia—such cases are discussed in various parts of the book. These beliefs, though inaccurate or incomplete, help individuals keep certain key bits of autobiographical information intact, thus maintaining a coherent sense of self. Moreover, these beliefs can aid in emotional regulation and social communication with peers, leading to a richer, more stable, and well-integrated understanding of one's place in the world.

The book is divided into seven chapters, each (apart from the introductory and the final, concluding chapter) dedicated to an exploration of a specific type of irrational belief. After discussing distorted memory beliefs in the second chapter, the following chapters examine confabulated explanations, elaborated and motivated delusional beliefs, and optimistically biased beliefs.

Each chapter begins with introductory remarks and conceptual groundwork, followed by a detailed argument demonstrating that the belief in question—despite being irrational—offers epistemic benefits. This is where the book really shines. Bortolotti presents a wealth of empirical and conceptual evidence, showing that various forms of epistemic irrationality have adaptive features that either directly improve epistemic functionality, or indirectly support it through emotional, psychological or social well-being. This idea serves as the central thread that weaves together various topics discussed in the book, even though the argument develops in important aspects to accommodate the specifics of each case.

For example, consider optimistically biased beliefs. A person with an overly optimistic view of their ability—whether in mathematics, language learning, or another skill—may demonstrate higher motivation and resilience when faced with setbacks. Even though their self-assessment is objectively inaccurate, it opens up new epistemic possibilities that would be inaccessible from a more modest (yet accurate) perception of their abilities. On the other hand, in psychiatric conditions such as schizophrenia, irrational beliefs often emerge in response to deeply distressing cognitive or affective symptoms. Take the example of the Cotard delusion, in which a person believes they are dead or do not exist. Bortolotti argues that this

belief, while highly disruptive, can also be epistemically beneficial if it provides the best explanation available for an otherwise anomalous and deeply puzzling experience. If a person says that she cannot feel anything and that she believes she is dead, we should recognize this as an attempt to structure an incomprehensible experience, rather than dismissing it as mere irrationality. If a more rational explanation were less effective in preserving psychological stability, then the irrational belief—while flawed—might still be the more epistemically functional one.

One of Bortolotti's most important contributions to the debate on epistemic irrationality is her challenge to the traditional view that irrational beliefs are deviations from epistemic norms (truth, justifiability, coherence, etc.), and should therefore be discarded as epistemically worthless. After reading her book, it becomes difficult to maintain the conviction that irrational beliefs are intrinsically defective. Her argument encourages deeper and broader exploration of the boundary between rational and irrational beliefs—one that moves beyond the standard, conventional distinctions.

Beyond its theoretical implications, Bortolotti's book also has practical significance. Her account challenges systematic exclusion of psychiatric patients, particularly those with schizophrenia. By recognizing their epistemically innocent beliefs as strategies for coping with abnormal experiences, she argues that their perspective deserves serious attention, rather than outright dismissal. This has direct implications for psychiatric practice, offering a framework that acknowledges hermeneutic injustice, while also building a conceptual bridge between clinical and non-clinical, everyday occurrences of irrational behavior, which has an important role in cultivating a more balanced approach to this important area of research.

In *The Epistemic Innocence of Irrational Beliefs* Lisa Bortolotti invites us to reconsider what it means for a belief to be epistemically valuable. It offers a powerful challenge to traditional epistemology that sees irrational beliefs as serving no epistemic functions at all. This impressive book expands our understanding of rationality, mental health and epistemic agency, while simultaneously serving as an appeal to a more compassionate and sympathetic treatment of people under psychiatric care. As such, it is an essential reading for anyone interested in epistemology, philosophy of mind, philosophy of psychiatry, or ethics.

ABSTRACTS (SAŽECI)

On the Discomfort with the Accuracy Criterion: Reassessing Puddifoot's Criticism of the Dual Factor View

Federico José Arena
Universitat de Girona, Spain and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

ABSTRACT

Given the pervasive presence of stereotypes' negative effects, there is a widely shared view according to which almost all stereotypes are harmful. However, some scholars have claimed that we should distinguish between stereotypes: those with and those without statistical support. In her book, Katherine Puddifoot claims that the statistical criterion falls short of what we need to develop a full theory of the epistemology of stereotyping, so she advocates for a Multifactorial view. While I share Puddifoot's discomfort with the accuracy criterion, in these comments I will put forward some critical considerations. Firstly, I will introduce a methodological concern regarding the discussion about the normative versus non-normative conception of stereotypes. Secondly, I will introduce some doubts about the extent to which the additional factors pointed out by Puddifoot's Multifactor view, are actually a challenge to the accuracy criterion. Thirdly, I will also critically comment on the argument that not having stereotypes or having egalitarian stereotypes that do not reflect some aspect of social reality improves our chances of attaining certain epistemic ends regarding the perceptions of individuals. Finally, I will briefly introduce a further factor of deception: the normativity of stereotypes, which is not considered within Puddifoot's proposal.

Keywords: conceptual analysis; accuracy criterion; generalizations; normative stereotypes.

Nelagode s kriterijem točnosti: ponovna procjena Puddifootine kritike dvofaktorskog stajališta

Federico José Arena
Universitat de Girona, Spain and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

SAŽETAK

S obzirom na sveprisutne negativne učinke stereotipa, široko je prihvaćeno gledište prema kojem su gotovo svi stereotipi štetni. Međutim, neki su istraživači tvrdili da bismo trebali razlikovati stereotipe: one koji imaju statističku potporu i one koji je nemaju. U svojoj knjizi Katherine Puddifoot tvrdi da statistički kriterij nije dovoljan za razvoj cjelovite epistemološke teorije stereotipizacije te stoga zagovara multifaktorski pristup. Iako dijelim Puddifootinu nelagodu s kriterijem točnosti, u ovom ću radu iznijeti nekoliko kritičkih razmatranja. Prvo, iznijet ću metodološku zabrinutost u vezi s raspravom o normativnoj naspram nenormativne koncepcije stereotipa. Drugo, izrazit ću određene sumnje u to do koje mjere dodatni čimbenici koje naglašava Puddifootin multifaktorski pristup zapravo predstavljaju izazov kriteriju točnosti. Treće, kritički ću se osvrnuti na argument prema kojem neimanje stereotipa ili posjedovanje egalitarnih stereotipa koji ne odražavaju određene aspekte društvene stvarnosti poboljšava naše šanse za postizanje određenih epistemičkih ciljeva u percepciji pojedinaca. Na kraju, ukratko ću predstaviti još jedan faktor obmane: normativnost stereotipa, koji nije razmotren u Puddifootinoj teoriji.

Ključne riječi: pojmovna analiza; kriterij točnosti; generalizacije; normativni stereotipi.

Accurate Stereotypes and Testimonial Injustice

Leonie Smith
Lancaster University, United Kingdom

ABSTRACT

In *How Stereotypes Deceive Us*, Katherine Puddifoot provides a convincing non-normative account of what stereotypes are, and of the conditions under which we appropriately rely on them in achieving our epistemic and ethical goals. In this paper, I focus on Puddifoot's discussion of what she takes to be the non-prejudicial use of accurate stereotypes and their role in causing or perpetuating harm. Such use can cause harm but does not, on the face of it, appear to be wrongful in the way that ordinary cases of prejudicially motivated use of stereotypes are. This raises a challenge for identifying when our use of such stereotypes might be unjust or wrongful (and why). In response, I first suggest that prejudice might be located within the context in which one uses a stereotype, rather than

within the content of the stereotype itself. In this way, we can indeed distinguish prejudicial (and therefore wrongful) use of accurate stereotypes from non-prejudicial (innocent) use of accurate stereotypes. And second, I suggest that we also ought to question whether the stereotypes being invoked in all cases really are accurate, given the context and scope of application.

Keywords: testimonial injustice; stereotypes; context; epistemic injustice.

Točni stereotipi i svjedočanska nepravda

Leonie Smith
Lancaster University, Ujedinjeno Kraljevstvo

SAŽETAK

U knjizi *Kako nas stereotipi obmanjuju*, Katherine Puddifoot pruža uvjerljivo nenormativno objašnjenje prirode stereotipa i uvjeta pod kojima se na njih možemo prikladno oslanjati u ostvarivanju naših epistemičkih i etičkih ciljeva. U ovom radu fokusiram se na Puddifootinu raspravu o onome što ona smatra nepredrasudnom upotrebom točnih stereotipa i njihovoj ulozi u uzrokovanju ili održavanju štete. Takva upotreba može prouzročiti štetu, ali na prvi pogled ne djeluje kao moralno pogrešna na isti način kao uobičajeni slučajevi upotrebe stereotipa motivirane predrasudama. To otvara pitanje kako prepoznati kada je naša upotreba takvih stereotipa nepravedna ili pogrešna (i zašto). Kao odgovor, najprije predlažem da se predrasude možda mogu smjestiti u kontekst u kojem se stereotip koristi, a ne nužno u sam sadržaj stereotipa. Na taj način možemo razlikovati predrasudnu (i stoga pogrešnu) upotrebu točnih stereotipa od nepredrasudne (nevine) upotrebe točnih stereotipa. Drugo, sugeriram da bismo također trebali preispitati jesu li stereotipi koji se koriste u svim slučajevima doista točni, uzimajući u obzir kontekst i opseg njihove primjene.

Ključne riječi: svjedočanska nepravda; stereotipi; kontekst; epistemička nepravda.

Some Critical Thoughts on "How Stereotypes Deceive Us"

Federico Picinali
London School of Economics and Political Science, United Kingdom

ABSTRACT

The goal of this paper is to identify and discuss the weaker aspects of some of the arguments in Kathy Puddifoot's fascinating and thought-provoking book. Section 1 deals with Puddifoot's treatment of the "single factor view" and the "dual factor view" of stereotyping. Section 2 deals with Puddifoot's treatment of egalitarian attitudes. Section 3 deals with Puddifoot's treatment of the moral encroachment approach to stereotyping. Finally, section 4 deals with Puddifoot's theory of evaluative dispositionalism. The sections can be read independently.

Keywords: stereotyping; egalitarian attitudes; moral encroachment; pragmatic encroachment; evaluative dispositionalism.

Neke kritičke misli o djelu „Kako nas stereotipi zavaravaju"

Federico Picinali
London School of Economics and Political Science, Ujedinjeno Kraljevstvo

SAŽETAK

Cilj ovog rada je identificirati i raspraviti slabije aspekte nekih argumenata iz fascinantne i poticajne knjige Kathy Puddifoot. Prvi dio bavi se Puddifootinim razmatranjem „jednofaktorskog" i „dvofaktorskog" gledišta na stereotipiziranje. Drugi dio analizira njezin pristup egalitarnim stavovima. Treći dio raspravlja njezin tretman moralnog zadiranja u kontekstu stereotipiziranja. Konačno, četvrti dio obrađuje Puddifootinu teoriju evaluativnog dispozicionalizma.

Ključne riječi: stereotipizacija; egalitarni stavovi; moralno zadiranje; pragmatično zadiranje; evaluativni dispozicionalizam.

Some Startling Consequences of How Stereotypes Deceive Us

Jennifer Saul
University of Waterloo, Canada

ABSTRACT

This paper argues that Puddifoot's arguments in How Stereotypes Deceive Us have more radical consequences than those argued for in the book. It does this by pointing out two problems for evaluating stereotypes via Evaluative Dispositionalism, Puddifoot's view. The first problem concerns the very large number of dispositions associated with any stereotype, and the second the difficulty of evaluating a stereotype in isolation from other elements of a person's psychology. The paper suggests that, when we take seriously Puddifoot's arguments, we'll end up concluding that it's not possible to assess the epistemic worth of any individual belief or stereotype. We could still discuss the epistemic merits of how it was formed, and discuss the epistemic consequences that this belief or stereotype has for a particular person in a particular situation. But overall epistemic evaluations of stereotypes, or even acts of stereotyping, would be something that we should try to avoid.

Keywords: stereotypes; dispositions; beliefs; epistemic evaluation.

Neke iznenađujuće posljedice toga kako nas stereotipi obmanjuju

Jennifer Saul
University of Waterloo, Kanada

SAŽETAK

Ovaj rad tvrdi da argumenti koje Puddifoot iznosi u knjizi *Kako nas stereotipi obmanjuju* imaju radikalnije posljedice od onih za koje se u knjizi eksplicitno zalaže. To pokazujemo pozivanjem na dva problema u evaluaciji stereotipa putem Evaluativnog dispozicionalizma, gledišta kojeg zastupa Puddifoot. Prvi problem odnosi se na iznimno velik broj dispozicija povezanih s bilo kojim stereotipom, a drugi na teškoću evaluacije stereotipa u izolaciji od drugih elemenata nečije psihologije. Rad sugerira da, ako ozbiljno shvatimo Puddifootine argumente, na kraju moramo zaključiti da nije moguće procijeniti epistemičku vrijednost bilo kojeg pojedinačnog vjerovanja ili stereotipa. I dalje bismo mogli raspravljati o epistemičkim zaslugama načina na koji je neko vjerovanje ili

stereotip formiran te o epistemičkim posljedicama koje ono ima za određenu osobu u određenoj situaciji. No, cjelovite epistemičke evaluacije stereotipa ili čak samih činova stereotipizacije bile bi nešto što bismo trebali nastojati izbjeći.

Ključne riječi: stereotipi; dispozicije; vjerovanja; epistemička procjena.

Replies to Contributors

Katherine Puddifoot
Durham University, United Kingdom

ABSTRACT

This paper provides responses to the 4 commentaries by Federico José Arena, Leonie Smith, Federico Picinali, and Jennifer Saul under the main headings: “Definition of stereotypes”; “Single/dual factor view”, “Epistemic benefits of egalitarian beliefs”, “Beyond stereotyping beliefs”, “Which disposition?”, “More radical implications of evaluative dispositionalism”, “Stereotypes, reality and testimonial injustice”, “Normative stereotypes”, and finally “Moral encroachment”.

Keywords: epistemic benefits; dispositions; stereotypes; single/dual factor view.

Odgovori komentatorima

Katherine Puddifoot
Durham University, Ujedinjeno Kraljevstvo

SAŽETAK

Ovaj rad pruža odgovore na četiri komentara Federica Joséa Arene, Leonie Smith, Federica Picinalija i Jennifer Saul pod naslovima: „Definicija stereotipa”, „Jednofaktorsko i dvofaktorsko gledište”, „Epistemičke prednosti egalitarnih uvjerenja”, „Iza stereotipnih uvjerenja”, „Koja dispozicija?”, „Radikalnije implikacije evaluativnog dispozicionalizma”, „Stereotipi, stvarnost i svjedočanstvena nepravda”, „Normativni stereotipi” i, konačno, „Moralna interferencija”.

Ključne riječi: epistemičke prednosti; dispozicije; stereotipi; jednofaktorsko/dvofaktorsko gledište.

Translated by Marko Jurjako (Rijeka)

Papers in this issue of EuJAP have been formatted by Marko Jurjako

AUTHOR GUIDELINES

Publication ethics

EuJAP subscribes to the publication principles and ethical guidelines of the Committee on Publication Ethics (COPE).

Submitted manuscripts ought to:

- be unpublished, either completely or in their essential content, in English or other languages, and not under consideration for publication elsewhere;
- be approved by all co-Authors;
- contain citations and references to avoid plagiarism, self-plagiarism, and illegitimate duplication of texts, figures, etc. Moreover, Authors should obtain permission to use any third-party images, figures and the like from the respective copyright holders. The pre-reviewing process includes screening for plagiarism and self-plagiarism by means of internet browsing and software Turnitin;
- be sent exclusively electronically to the Editors (eujap@ffri.uniri.hr) (or to the Guest editors in the case of a special issue) in a Word compatible format;
- be prepared for blind refereeing: authors' names and their institutional affiliations should not appear on the manuscript. Moreover, “identifiers” in MS Word Properties should be removed;
- be accompanied by a separate file containing the title of the manuscript, a short abstract (not exceeding 300 words), keywords, academic affiliation and full address for correspondence including e-mail address, and, if needed, a disclosure of the Authors' potential conflict of interest that might affect the conclusions, interpretation, and evaluation of the relevant work under consideration;
- be in American or British English;
- be no longer than 9000 words, including references (for Original and Review Articles).
- be between 2000 and 5000 words, including footnotes and references (for Discussions and Critical notices)

We ask authors to submit only one manuscript at a time. A second submission by the same author is allowed only after a final decision has been made on their previously submitted manuscript.

Norms for publishing with AI

The Journal does not exclude the use of AI generated text. However, all authors (including reviewers and editors) take full responsibility for its factual accuracy and the proper acknowledgement of sources. In the acknowledgement section of your manuscript or the title page (depending on the submission/publication stage) or in other kind of reports you must identify the AI that was used, and the extent of the contribution. For instance, ChatGPT (version or the date when the AI was used).

The contribution level of the AI can be defined as follows:

- negligible – means the AI only made minor changes to the manuscript's style or grammar (this includes using AI for copyediting and similar services);
- modest – means the AI made important suggestions but was not the primary driver of the research or had an essential role in writing the manuscript;
- substantial – means the AI made several crucial suggestions that shaped the research and the manuscript could not have been completed without it.

If the contribution of the AI is "negligible", there is no requirement to mention its usage during the submission or review and publication processes. However, for any other level of contribution, it is expected that authors will report the extent of AI usage. In cases where the AI contribution is "substantial", authors, reviewers, and editors should provide a comprehensive description of the AI usage and its contributions in a narrative format.

Initial submission

When first submitting a manuscript, it is not required that the manuscript conforms to EuJAP's style guidelines. Only after a manuscript has been accepted for publication, we expect the authors to format the manuscript in accordance with EuJAP's style guidelines.

Submitting revised manuscripts

When submitting a revised manuscript, please include also a separate document where it is explained how revisions were made in response to reviewers' comments.

Policy for submitted manuscripts

If the submitted manuscript is authored by more than one person, there should be a brief explanation in the title page of the contribution of each Author with respect to the conception and design of the argument, study, etc. and writing of the paper.

To preserve the anonymous status of the review process, we prefer (but do not require) that submitted versions of manuscripts are not deposited in open access article repositories.

Policy for accepted and published manuscripts

Accepted and published versions of the manuscript can be deposited in institutional or personal repositories without an embargo period. In case of published manuscripts, a link (with DOI) to the journal's web pages and/or HRCAK should be added.

Malpractice statement

If the manuscript does not match the scope and aims of EuJAP, the Editors reserve the right to reject the manuscript without sending it out to external reviewers. Moreover, the Editors reserve the right to reject submissions that do not satisfy any of the previous conditions.

If, due to the authors' failure to inform the Editors, already published material will appear in EuJAP, the Editors will report the authors' unethical behaviour in the next issue and remove the publication from EuJAP web site and the repository HRČAK.

In any case, the Editors and the publisher will not be held legally responsible should there be any claims for compensation following from copyright infringements by the authors.

For additional comments, please visit our web site and read our Publication ethics statement (https://eujap.uniri.hr/publication-ethics/). To get a sense of the review process and how the referee report ought to look like, the prospective Authors are directed to visit the *For Reviewers* page on our web site (https://eujap.uniri.hr/instructions-for-reviewers/).

Style

Accepted manuscripts should:

- follow the guidelines of the most recent Chicago Manual of Style
- contain footnotes and no endnotes
- contain references in accordance with the author-date Chicago style, here illustrated for the main common types of publications (T = in text citation, R = reference list entry)

Book
T: (Nozick 1981, 203)
R: Nozick, R. 1981. *Philosophical Explanations.* Cambridge: Harvard University Press.

Book with multiple authors

T: (Hirstein, Sifferd, and Fagan 2018, 100)

R: Hirstein, William, Katrina Sifferd, and Tyler Fagan. 2018. *Responsible Brains: Neuroscience, Law, and Human Culpability*. Cambridge, Massachusetts: The MIT Press.

Chapter or other part of a book
T: (Fumerton 2006, 77-9)
R: Fumerton, Richard. 2006. 'The Epistemic Role of Testimony: Internalist and Externalist Perspectives'. In *The Epistemology of Testimony*, edited by Jennifer Lackey and Ernest Sosa, 77–91. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199276011.003.0004.

Edited collections
T: (Lackey and Sosa 2006)
R: Lackey, Jennifer, and Ernest Sosa, eds. 2006. *The Epistemology of Testimony*. Oxford: Oxford University Press.

Article in a print journal
T: (Broome 1999, 414-9)
R: Broome, J. 1999. "Normative requirements." *Ratio* 12: 398-419.

Electronic books or journals
T: (Skorupski 2010)

R: Skorupski, John. 2010. “Sentimentalism: Its Scope and Limits.” *Ethical Theory and Moral Practice* 13 (2): 125–36. https://doi.org/10.1007/s10677-009-9210-6.

Article with multiple authors in a journal
T: (Churchland and Sejnowski 1990)
R: Churchland, Patricia S., and Terrence J. Sejnowski. 1990. “Neural Representation and Neural Computation.” *Philosophical Perspectives 4*. https://doi.org/10.2307/2214198

T: (Dardashti, Thébault, and Eric Winsberg 2017)
R: “Dardashti, Radin, Karim P. Y. Thébault, and Eric Winsberg. 2017. Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity.” *The British Journal for the Philosophy of Science* 68 (1): 55–89. https://doi.org/10.1093/bjps/axv010

Website content
T: (Brandon 2008)
R: Brandon, R. 2008. Natural Selection. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed September 26, 2013. http://plato.stanford.edu/archives/fall2010/entries/natural-selection

Forthcoming
For all types of publications followed should be the above guideline style with exception of placing ‘forthcoming’ instead of date of publication. For example, in case of a book:
T: (Recanati forthcoming)
R: Recanati, François. forthcoming. *Mental Files*. Oxford: Oxford University Press.

Unpublished material
T: (Gödel 1951)
R: Gödel, Kurt. 1951. *Some basic theorems on the foundations of mathematics and their philosophical implications*. Unpublished manuscript, last modified August 3, 1951.

Final proofreading

Authors are responsible for correcting proofs.

Copyrights

The journal allows the author(s) to hold the copyright without restrictions. In the reprints, the original publication of the text in EuJAP must be acknowledged by mentioning the name of the journal, the year of the publication, the volume and the issue numbers and the article pages.

Archiving rights

The papers published in EuJAP can be deposited and self-archived in the institutional and thematic repositories providing the link to the journal's web pages and HRČAK.

Subscriptions

A subscription comprises two issues. All prices include postage.

Annual subscription:

International:

individuals € 50

institutions € 100

Croatia:

individuals € 30

institutions € 60

Bank: Zagrebačka banka d.d. Zagreb

SWIFT: ZABAHR 2X
IBAN: HR9123600001101536455

Only for subscribers from Croatia,

please add: "poziv na broj": 0015-03368491

European Journal of Analytic Philosophy is published twice per year.

The articles published in the *European Journal of Analytic Philosophy* are indexed and abstracted in SCOPUS, SCImago, Web of Science (Emerging Sources), The Philosopher's Index, European Reference Index for the Humanities (ERIH PLUS), Dimensions, Directory of Open Access Journals (DOAJ), PhilPapers, Portal of Scientific Journals of Croatia (HRČAK), Sherpa Romeo (now Jisc's open policy finder), ANVUR (Italy)