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

ABSTRACT

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

Keywords: epistemic benefits; epistemic costs; stereotypes.

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.