Routledge
Taylor & Francis Group

INQU Inquiry
An Interdisciplinary Journal of Philosophy

INQUIRY
An Interdisciplinary Journal of Philosophy

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/sinq20

Implicit bias and qualiefs

Martina Fürst

To cite this article: Martina Fürst (2023): Implicit bias and qualiefs, Inquiry, DOI: 10.1080/0020174X.2023.2217561

To link to this article: https://doi.org/10.1080/0020174X.2023.2217561

Published online: 05 Jun 2023.

INQUIRY
https://doi.org/10.1080/0020174X.2023.2217561

Routledge
Taylor & Francis Group

Implicit bias and qualiefs

Martina Fürst

Department of Philosophy, University of Graz, Graz, Austria

ABSTRACT

In analyzing implicit bias, one key issue is to clarify its metaphysical nature. In this paper, I develop a novel account of implicit bias by highlighting a particular kind of belief-like state that is partly constituted by phenomenal experiences. I call these states 'qualiefs' for three reasons: qualiefs draw upon qualitative experiences of what an object seems like to attribute a property to this very object, they share some of the distinctive features of proper beliefs, and they also share some characteristics of what Gendler calls 'aliefs'. I proceed as follows: First, I develop a general theory of qualiefs. Second, I argue that implicit bias involves *generic* qualiefs that involve experiences that have been shaped by stereotypes. Elaborating on the particular content of a generic qualief, I explain why we are unaware of the bias even though it involves an experience. Third, I demonstrate that the qualief-model best explains the key features of implicit bias: it accounts for the biases´ implicitness and automaticity. Moreover, it elucidates how implicit bias can be insensitive to logical form and evidence, but at the same time it can serve as propositional input to further mental states.

ARTICLE HISTORY Received 3 February 2022; Accepted 20 March 2023

KEYWORDS Implicit bias; phenomenal experiences; phenomenal concepts; generics; beliefs; aliefs

1. Introduction

Over the last two decades, the psychological literature on implicit bias has been flourishing. More recently the phenomenon has gained increased interest among philosophers as well.[1] So, what is implicit bias? As a first approximation, we can say that the notion of 'implicit bias' aims to capture implicit mental states that influence our behavior and attitudes

CONTACT Martina Fürst martina.fuerst@uni-graz.at Department of Philosophy, University of Graz, Heinrichstrasse 26/5, 8010Graz, Austria

[1]For a collection of philosophical papers on the topic, see Brownstein and Saul 2016.

when social categories are in play.[2] The 'implicitness' of these states can be interpreted as these states being unconscious (or unaware), unendorsed, uncontrollable, or revealed by indirect measures (for a helpful taxonomy of these readings, see Holroyd, Scaife, and Stafford 2017). Here I adopt the widely shared reading of implicitness as *unawareness*.

Implicit bias leads to discriminatory behavior and to unfair judgments (e.g. about the qualification of a job applicant (Dovidio and Gaertner 2000; Bertrand and Mullainathan 2004; Rooth 2013). Moreover, split second decisions seem to be particularly open to the influence of implicit bias. For example, when a subject is primed with the picture of a Black person, a harmless object is more likely to be misidentified as a gun (Payne 2006). The high relevance of these experimental findings concerning the consequences of implicit bias is obvious and motivates a deeper analysis of the phenomenon.

It is worrisome that in most cases subjects are unaware of their implicit biases. This unawareness is often held to be a key feature of implicit bias (Saul 2013).[3] However, some studies question the introspective inaccessibility of implicit bias. For example, bogus pipeline experiments (Nier 2001) point towards the accessibility of implicit bias and studies by Hahn et al. (2014) and Hahn and Gawronski (2019) suggest that individuals, when reflecting carefully, can predict to some extent their behavior and judgments influenced by implicit bias. According to their hypothesis what is introspectively accessible are gut feelings from which one *infers* the implicit bias. (However, if the emotional aspect is not a constitutive part of the bias, the implicit bias itself would still be *directly* introspective inaccessible.) In his (2019), Gawronksi argues against the hypothesis that people are unaware of the contents of their implicit bias as well (but he holds that people might still be unaware of the origin or the effects of their implicit biases.) In the light of these studies, I do not assume that implicit bias is completely inaccessible. I confine myself to the weaker claim that the contents of implicit bias are *not easily introspectively accessible*. Moreover, even if one becomes aware of one´s implicit biases, they are hard to change.[4] One powerful motive for overcoming one´s implicit bias is the insight that it conflicts with one´s explicit egalitarian views and anti-discriminatory commitments. Accordingly, in the literature, cases in

[2]'Implicit bias' can be seen as a notion that refers to implicit mental states or to a higher-order, normative, phenomenon that is realized by underlying mental states. Here I use the notion to refer to the implicit mental states that influence our behavior and judgments.

[3]Subjects can become indirectly aware of their implicit bias due to reading about implicit bias, measurements such as the Implicit Association Test, or by reflecting on their behavior.

[4]For a thorough analysis of debiasing experiments, see Byrd 2019.

which implicit bias is in tension with explicitly held beliefs receive much attention (e.g. Levy 2015; Gendler 2008a, 2008b).[5]

In analyzing the phenomenon of implicit bias, the key issue to begin with is to clarify its metaphysical nature. There is significant controversy about how to characterize implicit bias. We can roughly discern two competing views on the issue: on the *associative* view, implicit bias is best characterized in terms of *associations* (Gawronski and Bodenhausen 2006[6]; Olson and Fazio 2006). On a wide reading of the associative view, the sui generis state Gendler (2008a, 2008b) calls 'alief' can be subsumed under the associative view as well, since **'**a paradigmatic alief is a mental state with associatively linked content that is representational, affective, and behavioral, and that is activated – consciously or non-consciously – by features of the subject´s internal or ambient environment.' (2010, 263). Notably, none of the associatively linked aspects need to be propositional.

The alternative *propositional view* has it that implicit bias is best analyzed as beliefs or belief-like attitudes (e.g. De Houwer 2014; Egan 2011; Mandelbaum 2013, 2016; Mitchell, De Houwer, and Lovibond 2021). Levy (2015) also thinks that implicit bias has a propositional structure, although he does not think that it qualifies as a proper belief but rather as a 'patchy endorsement'.

It is important to clarify whether implicit bias has an associative or a propositional structure, since the competing views have different consequences for a deeper understanding of the phenomenon. For example, depending on which view one endorses, different methods for changing implicit bias appear promising. So, if we can elucidate the metaphysical structure of implicit bias, we have a fixed point from which to explore further important questions concerning the phenomenon. Accordingly, the goal of this paper is to clarify the metaphysical nature of implicit bias.

I proceed as follows. In Section 2, I survey the challenges faced by the associative models and by the propositional models and motivate the search for an alternative model. In Section 3, I introduce a special, belief-like, mental state that essentially involves experiences. I call these states 'qualiefs' since they use *qualitative* experiences to think about external objects. In Section 4, I develop a novel account of implicit bias as involving qualiefs. In particular, I argue that implicit bias is best

[5]An analysis of implicit bias that is *in alignment* with explicit beliefs offers fruitful insights of the phenomenon as well (Holroyd 2016). Here I follow the main focus in the literature and expose my view by investigating *conflict* cases between explicit beliefs and implicit bias.

[6]They updated their view to a *not purely* associative account in Gawronski and Bodenhausen 2014.

analyzed as a *generic* qualief that involves experiences that have been shaped by stereotypes. Elaborating on the particular content of a generic qualief, I explain why we are unaware of the bias even though it involves an experience. Section 5 is dedicated to the explanatory power of the proposed account. I demonstrate that generic qualiefs best explain the key features of implicit bias: its implicitness, automaticity, and insensitivity to evidence. Finally, I show how the proposed account elucidates a surprising characteristic of implicit bias that I label its 'asymmetric inferential profile'.

2. The structure of implicit bias: associative or propositional?

The debate about how best to characterize implicit bias turns mainly on the question of whether its structure is associative or propositional.[7] Defenders of the former view think that bias consists in mere *associations* between paired representations, e.g. between two concepts, or a concept and a valence. (The associative *structure* of the bias – i.e. the specific *causal relation* between mental representations – does not preclude that propositional elements can also be related associatively. What is important is that the elements need not be propositional and that the relation between the elements is essentially associative).[8]

The associative view fits nicely with dual-system theories which have it that human cognition is divided into two systems: an automatic, associative System1 and a more reflective, rule-based System2. (For an overview, see e.g. Evans and Stanovich 2013 and Evans and Frankish 2009). In particular, research in the heuristics and biases tradition (e.g. Kahneman 2011) points towards implicit bias as realized by processes underlying System1 that are associative, fast, automatic, and unconscious (whereas System2 involves slow, reflective, conscious processes that follow logical norms).[9]

[7]This categorization does not capture all views on the issue. For example, on Machery´s (2016) view, attitudes are *traits* which cannot be characterized as implicit or explicit at all.

[8]Thanks to an anonymous referee for drawing my attention to this point.

[9]The dual-system theories we find in the literature often differ in terms of what they consider to be the key features of the two systems. For example, Sloman (1996) focuses on the associative character of System1 and the rule-based character of System2, whereas Evans and Over (2004) focus on the implicitness of System1 and the explicitness of System2. Moreover, they link the former system to instrumental rationality and the latter system to normative rationality. Gawronski and Bodenhausen (2006) concentrate on the interplay of the two kinds of processes and the implications of this interplay for methods to change implicit attitudes. What is important for our present purposes is that most theorists agree that the processing of System1 is essentially associative, automatic, and unconscious, which fits nicely with many of the key features of implicit bias.

Gendler´s model of 'aliefs' (2008a, 2008b) can be subsumed under the associative view as well, insofar as it invokes a single, sui generis mental state, consisting of a tightly, associatively, connected triad of representational, associative and emotional aspects. Importantly, none of these aspects need to be propositional and their connection is essentially associative.[10] If implicit bias is characterized as automatically linked associations, it is not open to revision by reasoning. This precludes some methods for mitigating it, but opens the door to new strategies for combatting it, e.g. via counter-conditioning and extinction.[11]

Opponents of the associative model think that this model suffers from the weakness that the associative contents are not truth-apt, since they lack the right structure for having accuracy conditions (Mandelbaum 2013; Levy 2015). As a consequence, implicit bias would be neither open to reasoning nor able to play a role in inferences. This stands in tension with research findings that show that new counter-attitudinal information can change implicit bias (Van Dessel, Ye, and De Houwer 2000) and that implicit bias partakes in propositional reasoning (Gawronski, Hofmann, and Wilbur 2006). Gendler´s 'alief' model faces an additional challenge, namely to motivate the need for an additional *sui generis* kind of state to account for implicit bias (Cimpian and Erickson 2012; Egan 2011). I return to this worry in Section 3.

On the alternative view, implicit bias has a *propositional* structure. In support of the propositional view, its defenders cite studies showing that implicit bias can function inferentially. For example, Mandelbaum (2016) points towards findings by Gawronski and colleagues (2006) that subjects harboring a negative implicit attitude towards a person A, when told that A dislikes B, develop a positive implicit attitude towards B. These findings – that a subject harboring implicit bias seems to subscribe to the reasoning 'the enemy of my enemy is my friend' – make sense on the cognitive balance theory (which assumes inferences between propositions), but are hard to explain on the associative

[10] Aliefs are supposed to explain a variety of psychological phenomena and implicit bias is only one of them.

[11] Philosophers disagree about how successful counterconditioning is in changing implicit bias. Studies by Dasgupta and Greenwald (2001), Olson and Fazio (2006) and Hu et al. (2017) underpin the positive effect of counterconditioning. Gendler argues that aliefs can be successfully changed via counterconditioning (2008, 572-576). In contrast, Mandelbaum holds that in some cases 'the logical intervention, being told that what they previously had learned was in fact backwards, was more effective than intensive counterconditioning' (2016, 17). Kurdi and Banaji (2019) showed that verbal information shifted implicit bias more effectively than repeated evaluative pairings but the effect decayed quickly. The adequate interpretation of these studies is disputed. Thus, I confine myself to the claim that in *many* cases implicit bias is not responsive to evidence but to counterconditioning.

model, so Mandelbaum says: 'If you find two negatives making a positive, what you've found is a propositional, and not an associative, process.' (Mandelbaum 2016, 18). From its alleged propositional structure, many philosophers conclude that implicit bias is best characterized as a belief. The implicitness of the phenomenon is then explained by adding that this belief is either 'unconscious' or 'fragmented' (Mandelbaum 2016; Egan 2011).[12] With regard to the methods for overcoming implicit bias, propositional model suggests the standard methods of belief revision and rational argument.

Propositional models face the challenge to explain why implicit bias is insensitive to logical form (Madva 2016) and to evidence (Gregg, Seibt, and Banaji 2006) – even though it can function as input to further beliefs.[13] Moreover, the propositional models have difficulties in accounting for the affective or phenomenal aspects involved in implicit bias. Acknowledging these aspects of implicit bias is a desideratum, since it offers an explanation of the biases´ insensitivity to evidence and of why in some cases subjects can predict their IAT scores. As Hahn and Gawronski (2019) argue, this prediction of implicit bias is plausibly guided by implicit evaluations that are consciously experienced as affective reactions.

This is just a rough sketch of the lively debate about the structure of implicit bias. What matters for the present purposes is that, on the one hand, associative models fare well in explaining the affective and phenomenal aspect of implicit bias, whereas the propositional models fail to do justice to the importance of these aspects. As I will show, acknowledging the phenomenal aspect helps to explain why bias is not under our control, insensitive to evidence, and why it can be mitigated by counterconditioning. On the other hand, research findings about the inferential role of implicit bias speak in favor of propositional models and are hard to make sense of on the associative views. Moreover, studies on how to change implicit bias often disagree in their findings. For example, Kurdi and Banaji (2019) found that verbal information can mitigate implicit bias, whereas other studies showed that associative debiasing manipulations are successful (Byrd 2019). All this suggests that implicit bias has a *heterogeneous character:* it might involve

[12]Another view, that can be subsumed under the propositional model, analyzes implicit bias as a sui generis state with a propositional structure called a 'patchy endorsement' (see Levy 2015).

[13]Mandelbaum (2016) discusses studies (Brinol et al. 2009) that suggest that some instances of implicit bias *are* sensitive to arguments. What matters for the present purposes is that implicit bias is at least *not as easily* revisable in the face of evidence as standard beliefs are.

different kinds of mental states and processes (Holryod and Sweetman 2016). Explaining the heterogenous character of implicit bias does not easily fit with either the propositional or the pure associative models. Accordingly, some theorists propose novel accounts that aim at explaining the heterogenous character of implicit bias. Johnson (2020), for example, argues for a functional characterization of implicit bias that leaves open which kinds of mental states bridge the gap between the inputs and outputs in implicit bias. Other models aim to account for the heterogeneity of implicit bias by incorporating multiple processes. For example, Sullivan-Bissett (2019) defends an account of implicit bias as unconscious imaginings that can involve both associations *and* propositions. This model can account for the heterogeneity of implicit bias. However, the explanation is based on the thesis that unconscious imaginings exist which some theorist deny (e.g. Kind 2001). Moreover, it assumes that there is one single state – unconscious imaginings – that covers multiple processes. That means the heterogeneity of bias is explained by a diversity of processes, which then are subsumed under one single state. One might rather prefer a model that has the same explanatory power but invokes only one single state with a particular, *unified,* nature.

So, can the heterogeneity as well as the key features of implicit bias be explained without subsuming two distinct kinds of processes under one single state? I think so. In the following, I introduce a novel view that aims at finding a middle ground between the associative and the propositional models. This middle ground is found by allowing that implicit bias has a propositional structure, though the propositional content is represented in a special, phenomenal, way. The proposed view explains the biases´ heterogeneity by incorporating different aspects in a single state, rather than postulating two different kinds of mental states that both realize implicit bias.

3. A theory of *qualiefs*

I propose that implicit bias is best analyzed as a belief-like state that involves a special usage of phenomenal concepts. I dub these states 'qualiefs'.[14] The notion 'qualief' is a term of art that refers to a specific kind of mental state that constitutively involves *qualitative* (or phenomenal) experiences and shares some of the distinctive features of proper

[14]The term 'qualief' is inspired by Gendler´s notion of *aliefs* because there are some parallels between these models (see Section 3.3.).

beliefs. In this Section 3, I develop the general theory of qualiefs. In Section 4, I analyze the specific kind of qualiefs that account for implicit bias.

One might wonder how the suggestion that implicit bias involves *phenomenal* experiences fits with the claim that we are unaware of the bias. To see that there is no tension, we have to distinguish between two kinds of qualiefs: *singular qualiefs,* which are introspectively accessible, and *generic qualiefs,* which – due to their particular internal structure – have a content that is hard to access introspectively. I suggest that implicit bias is best analyzed as a generic qualief. Since the qualief model can account for the main explanatory desiderata of a theory of implicit bias – its implicitness, automaticity, insensitivity to evidence and inferential role – I provide an abductive argument for the proposed view.

3.1. Phenomenal concepts

The proposed account draws upon an insight from the debate about the metaphysical nature of phenomenal states, namely that we can conceptualize these states in a phenomenal way, via 'phenomenal concepts'. In the following, I outline the connection between phenomenal concepts and 'qualiefs'.

Traditionally, the notion of *phenomenal concepts* is used by physicalists to explain away anti-physicalist arguments. What is known as the *phenomenal concept strategy* is based on the following line of argument. First, physicalists point at special, phenomenal, concepts that directly pick out phenomenal states in terms of their phenomenal character. Next, they hold that this way of conceptualizing phenomenal states gives rise to the *hard problem of consciousness* and the related anti-physicalist intuition of distinctness of phenomenal states and physical states. Finally, they claim that phenomenal and physical concepts pick out the same physical referent, e.g. a neurophysiological state. By pointing to the particularities of phenomenal concepts (for example, their 'conceptual isolation' (Carruthers and Veillet 2007), i.e. the fact that phenomenal concepts lack any a priori connections with physical or functional concepts) an explanation of anti-physicalist intuitions is provided, without being committed to ontological anti-physicalist conclusions. Despite the diversity of views about the nature of phenomenal concepts,[15] most philosophers agree on

[15]Some philosophers think that phenomenal concepts are inner demonstratives (Levin 2007), whereas others (Balog 2012; Block 2007; Chalmers 2007; Loar 1997; Papineau 2007) hold that they use, quote, or are partly constituted by phenomenal states. For an analysis of which account can best

the basic idea that there is a special, first-person way to think about phenomenal states which involves tokens of these very states. This minimal agreement suffices for present purposes.

In what follows, I assume that there are phenomenal concepts, and I propose a broader application for this notion. The debate about phenomenal concepts aims to explain our intuitions about *experiences* and, hence, centers on phenomenal concepts referring to experiences. I suggest that phenomenal concepts understood as specific concepts that constitutively involve experiences – besides picking out experiences – can also be deployed to refer to *external objects* in terms of how they appear phenomenally to the subject.[16] This outward-directed usage of phenomenal concepts is a much-neglected phenomenon to which I want to draw the attention. The key-idea is the following: Suppose that you are thinking about *phenomenal states*, such as the experience of seeing a red car, in terms of what this experience is like. It is just a small step to think in these phenomenal terms about the *external object* as well, by taking the *car* as red. Thus, the phenomenal way of thinking is not restricted to the realm of experiences, but can be extended to the realm of external objects.

Let me clarify that there need not be a shift from referring to experiences first to referring to the external objects. It might turn out that thinking in terms of phenomenal concepts about the external world is antecedent to thinking in this way about experiences, which might require an additional reflective process. The crucial point is that we can refer to internal states *as well as to external objects* by deploying phenomenal concepts.

So, to a first approximation, we can make the following distinction:

1) We can use phenomenal concepts to refer to *internal states*.
2) We can use phenomenal concepts to refer to *external objects*.

A common feature of both usages is that cognitive states involving phenomenal concepts differ from standard beliefs insofar as they are

explain the conceptual isolation and the cognitive role of phenomenal concepts, see Chalmers 2007; Fürst 2014.

[16]To the best of my knowledge, this broader application has not been discussed in the literature on phenomenal concepts yet. Lehrer (2019) uses his notion of 'exemplarization' to explain how we can think in phenomenal terms about both experiences and external objects. On his account, an experience can be used to refer to the class it is an instance of *and* to external objects. Thus, an exemplarized state exhibits a 'Janus-faced' character by being at the same time inwardly and outwardly directed. The proposed account of qualiefs is inspired by Lehrer´s metaphor of the 'Janus-faced' character and draws upon his account of exemplarization.

difficult to influence via cognitive means. The reason for this is that the content is presented in a phenomenal way, namely in terms of what an experience or an external object *seems like* to the subject.[17] Phenomenal experiences are paradigm cases of states that are not reason-responsive, and their usage carries this feature over to the relevant cognitive states. Let me illustrate this with an example. Knowing that the lines of the Müller-Lyer illusion are of equal length does not change your phenomenal experience of them as different in lengths. On a widely held view, encapsulation explains the cognitive impenetrability and persistence of the phenomenal experience (Fodor 1983; Pylyshyn 1999). Accordingly, thinking about the Müller-Lyer illusion in terms of phenomenal concepts – i.e. in terms of concepts that involve an experience – represents the lines as of different lengths. Measuring the lines will lead to the *belief* that the lines are equal, but it does not influence the mental state which involves the experience. This insensitivity to evidence of mental states that use phenomenal concepts will be crucial when it comes to analyzing implicit bias.

3.2. Qualiefs

On the orthodox view, phenomenal concepts are attributed to internal referents, namely to phenomenal states. This view, and the resulting 'phenomenal concept strategy', provides significant insights in our understanding of phenomenal states. Here, however, I choose not to focus on this orthodox usage of phenomenal concepts.[18] Rather, I propose to extend the application of phenomenal concepts to external referents as well, a move which has not yet been discussed in the literature. In what follows, I focus on the usage of phenomenal concepts to refer to *external objects*. If a phenomenal concept is used to think about external objects, the phenomenal aspect fuses with the representation of the external object. The resulting state has the *propositional* content that an object is *F*, but this content is presented in a particular way, namely in a *phenomenal way*.[19]

Many philosophers think that a proposition can be entertained under different modes of presentation. To entertain a proposition under a

[17] I use the term 'seeming' in the *phenomenal sense* (rather than in the doxastic sense of the term). Moreover, an object´s *seeming* in a particular way does not implicate that this object is *not* that way.
[18] For an analysis of phenomenal concepts referring to experiences, see Fürst 2014; Fürst forthcoming b.
[19] The qualief account is a weaker thesis than the 'cognitive phenomenology thesis'. According to the cognitive phenomenology thesis *every* conscious thought (essentially) exhibits a phenomenal character (Fürst forthcoming a). Here I am only concerned with the phenomenal character of qualiefs and I remain neutral about the phenomenology of standard beliefs.

phenomenal mode of presentation is special insofar as it involves the instantiation of a phenomenal experience. In this respect, it is similar to entertaining a proposition under what Stanley and Williamson (2001) call a 'practical mode of presentation'. Just like thinking under a practical mode of presentation, thinking under a phenomenal mode of presentation is in some respects analogous to the first-person mode of presentation. Moreover, the phenomenal mode of presentation is rich and vivid. Notably, what is represented in this particular way is still a *propositional* content. Accordingly, the resulting mental state turns out to be a hybrid of phenomenal and external-representational features. I dub these mental states *qualiefs.*

A *qualief* is a mental state that constitutively involves phenomenal concepts and uses them to attribute properties to external referents. That means, a qualief is a hybrid of phenomenal and external-representational features that is not sensitive to evidence. Examples are easy to find: you can qualieve *that this car is red*.[20] One might think that the content of a qualief can be responsive to evidence; e.g. by learning that the car is white and illuminated by red light we come to qualieve that the car is white. This is not the case. Instead of changing the qualief due to the evidence, we might rather switch from the *qualief* that the car is red to the *belief* that the car is white. The qualief is partly constituted by an experience and since this experience remains unaffected by the new evidence, the qualief does not change either.

One reason for why one might (falsely) believe that the content of a qualief could be evidence-responsive lies in the notorious difficulty to publicly express a mental state that involves phenomenal concepts. As Chalmers (2003) notes, this difficulty applies in particular to pure phenomenal concepts that involve an occurrent experience.[21] Since qualiefs are partly constituted by phenomenal concepts that involve an experience, verbally expressing the phenomenal aspect t of a qualief is a difficult task. For this reason, the phrase 'qualieving that p' is hereinafter used as an auxiliary mean to express a mental state that constitutively involves a phenomenal experience to present its content.

[20]One might wonder under which conditions we are prone to think in terms of phenomenal concepts about external objects. Presumably, we tend to have qualiefs when having occurrent experiences, e.g., when directly interacting with the target object.

[21]Chalmers´ notion of 'phenomenal beliefs' is similar to qualiefs insofar as a 'phenomenal belief is partly constituted by an underlying phenomenal quality.' (2003, 235) However, phenomenal beliefs attribute phenomenal properties under phenomenal concepts to *mental states,* whereas qualiefs attribute properties to *external objects.* Moreover, qualiefs are not beliefs.

3.3. Beliefs, aliefs, and qualiefs

Qualiefs are not beliefs: although both mental states might have the same representational content, the qualief presents its content in a different, namely in the phenomenal, way. Beliefs might be formed on the basis of experiences as well, but only in qualiefs the experience becomes a constitutive part of the mental state. As a result, qualiefs are *evidence-insensitive*. One might note that also some beliefs are insensitive to evidence, as the literature on irrational beliefs shows (e.g. Bortolotti 2009). However, the evidence-insensitivity of qualiefs is pervasive and of a particular kind. It is the phenomenal experience which is part of the qualief that makes it resilient against evidence.

Despite this key characteristic that differentiates qualiefs from beliefs, qualiefs are still *belief-like* mental states. What makes them similar to beliefs is that they share parts of the representational-functional profile of beliefs. In particular, what is shared with beliefs is that qualiefs take the world to be in a certain way; they have propositional contents.

Aptly characterizing full-fledged beliefs is a complicated issue and elaborating on this question would carry us too far off course. Nevertheless, let me make some clarificatory remarks about the relation between beliefs and qualiefs. Some philosophers think that to qualify as a belief, a state must be governed by truth and criteria such as evidence-responsiveness and alignment with other beliefs need to be met (Gendler 2008a; Levy 2015). Helton (2020) argues for the 'revisability view of beliefs' which has it that if a subject cannot revise a mental state in response to evidence, then this state is not a belief. On these demanding criteria, qualiefs would not qualify as belief-like. However, my claim is weaker. I hold that qualiefs share the less sophisticated criterion of beliefs of taking the world to be in a specific way, and that by doing this they influence our behavior and reasoning.[22] In that sense, if you qualieve that p, you stand in a belief-like relation to the content that p.[23]

[22]Velleman (2000) argues that other states, such as imaginings or supposings, take the world to be in a specific way and thereby motivate actions as well. (Thanks to an anonymous referee for pointing this out to me). That imaginings share this feature with qualiefs explains why they can be integrated in the qualief account. I elaborate on this in Section 4.1.

[23]Given that both beliefs and qualiefs have propositional contents and can figure in inferences, one might ask: under which conditions are qualiefs involved in inferences? This is a tricky question that needs empirical investigation. My hypothesis is that explicit reasoning tends to involve beliefs, whereas implicit reasoning, which is triggered by occurrent experiences, is likely to involve qualiefs. Moreover, conflict-cases in which behavior and judgments are not in alignment with explicit beliefs point towards qualiefs figuring in the inference. (If the qualief is in alignment with the explicit belief, it is hard to find out which state figures in the inference).

Next, let me clarify the account further by delineating important differences to Gendler (2008a, 2008b) prominent notion of 'aliefs'. That I dub my account 'qualiefs', which sounds similar to 'aliefs', is no coincidence. Some parallels between these two accounts are worth noting. The most significant one is the tight connection between representational and phenomenal aspects.

Gendler emphasizes that an alief is a simple, unified state (rather than a cluster of causally related, but distinct, familiar states such as perceptions, emotions and behavioral reactions).[24] Motivating this 'unity strand' (Egan 2011, 67) is one of the main challenges faced by the alief account (Egan 2011; Currie and Ichino 2012). The qualief account shares the idea with Gendler´s account that representational and phenomenal features are intimately connected (though the behavioral aspect is seen as separate and causally related). It is an advantage of the qualief account that it can additionally provide an *explanation* of this tight connection: since the content of a qualief is phenomenally presented, these two aspects are fused and build a hybrid state. This hybrid displays a much tighter connection than a causal relation or co-activation of two separate representational and phenomenal states.[25]

There are significant differences between aliefs and qualiefs as well. First, an alief is held to be an innate kind of state, shared with non-human animals and conceptually antecedent to other cognitive mechanisms. In contrast, a qualief is a cognitively sophisticated mental state with a propositional content and one which, presumably, non-human animals have not yet developed.

Second, Gendler contrasts full-fledged beliefs, which reflect what one takes to be true, with aliefs, which are tied to how things merely seem. In contrast, the proposed model reconciles these two aspects. A qualief uses an experience of an object seeming in a specific way to represent a propositional content of what one takes to be true.

[24]Brownstein and Madva hold that affective and cognitive components of implicit bias are indissociable and that implicit biases 'consist in 'clusters' of semantic-affective associations.' (2018, 611) This view gets a lot right about implicit bias in that it does justice also to its phenomenal aspects. However, as a one-type model that involves clusters of phenomenal and cognitive components, it faces the same challenge as the alief-model – to motivate and explain the unity of the components.

[25]Some philosophers (Currie and Ichino 2012; Holroyd 2016) suggest analyzing implicit bias as co-activated or causally related representational and affective contents, plus behavioral responses. This shall explain why counterconditioning is a promising strategy to change implicit bias – it breaks the causal relation between the representational and the affective content. Notably, the qualief model also offers an explanation of the fruitfulness of counterconditioning: counterconditioning changes the experiences and thereby it changes the qualief.

Third, since no part of the representational-affective-behavioral triad need to be propositional, aliefs cannot be changed by inferential means and do not figure in inferential reasoning. In contrast, a qualief has a propositional content, though represented in a phenomenal way. As a result of the special, phenomenal, mode of presentation, qualiefs cannot be easily changed via reasoning. However, given their propositional structure, they can figure as inputs to inferential reasoning. I dub this characteristic – that qualiefs are insensitive to evidence but can serve as propositional input for other attitudes – an 'asymmetric inferential profile'. This feature is explored in more detail in Section 5.

Let us pause for a moment to consider where we stand. We started focusing on the usage of phenomenal concepts to attribute properties to external referents. I introduced 'qualiefs' as the mental states that draw upon an experience of what an external object seems like to attribute a property to this very object (or to its class). Thus, qualiefs are hybrid mental states that fuse phenomenal and representational features. Given the usage of an experience, the way of attributing the relevant property is richer and more vivid than in beliefs, and it is difficult to influence via evidence.[26]

4. Qualiefs involved in implicit bias

Until now, I have developed a general account of the special mental states that I dubbed 'qualiefs'. Next, I will focus on the structure of those qualiefs that underly implicit bias, starting with the *kinds of experiences* involved in those qualiefs before also discussing their specific *content*.

4.1. The experiences

Various kinds of experiences can be an essential part of qualiefs. With regard to the qualiefs involved in implicit bias, two kinds of experiences are particularly important: First, *actual* perceptual experiences of an individual seeming a specific way (e.g. when we directly interact with

[26]One might ask: what are the advantages of a model that suggests one single state with a hybrid character over a model that posits a doxastic state plus an associated phenomenal content? The problem with the latter view is that, if there is a doxastic state that is independent of (but related to) a phenomenal state, in principle one might change while the other remains the same. In contrast, on the qualief account the phenomenal aspect and the representational are fused and, hence, a change in the phenomenal aspect implies a change of the qualief. Moreover, as I will argue in Section 5, this tight connection – fusion of the representational and phenomenal aspects – explains the biases´ insensitivity to evidence best.

members of the target group) and, second, *imaginative* or recreative experiences (e.g. when we make judgements about members of the target group without them being actually present). For example, in the evaluation of a job applicant (Dovidio and Gaertner 2000), the subject´s noticing that the applicant´s name is Jamal might prompt an imaginative experience of a Black man seeming a particular way.[27] What is germane to qualiefs underlying implicit bias is that the experience, which is an essential part of the qualief, is shaped by *stereotype representations*. For example, in the case of an implicit racist bias the stereotype representations in our cultural environment might influence the experience of a particular Black man as seeming dangerous (Eberhardt et al. 2004).[28]

There is wide agreement that implicit bias is caused by the stereotypes that we find in our social environment. Stereotypes are typically characterized as explicit beliefs (Brownstein and Madva 2018, 612).[29] However, prior to forming the relevant stereotype-beliefs, we are already exposed to representations of these stereotypes such as images in movies, advertisements, but also real-world settings (for example, one might meet only male pilots or only female midwifes). Presumably, the awareness of these representations suffices to influence our experiences. Thus, stereotype representations that we encounter everyday already shape our

[27]In the literature, we find accounts that focus on the role of imaginative experiences in implicit bias. For example, Welpinghus (2020) argues for an imagination model: 'When you sit at your desk with CVs in front of you and choose whom to invite for a job interview, you will imagine the qualified candidates in the job to be given.' (Welpinghus 2020, 1621) On her view, implicit bias is the result of stereotypes that influence the process of imagination. This view aims at explaining implicit bias without positing unconscious mental states, but rather by analyzing it as a disposition that involves imagination. I agree that imaginative experiences, influenced by stereotypes, can play a key role in implicit bias. However, I share the widely held view that implicit bias is realized by mental states. Hence, I suggest to integrate imaginative experiences as part of a qualief. Moreover, Nanay refers to involuntary mental imagery, 'understood as early perceptual processing that is not directly triggered by sensory input ' (2021, 331 , 4) to account for implicit bias. On Nanay´s view, implicit bias does not have a propositional structure and it does not feature in inferences (2021). Accordingly, the findings about the inferential power of implicit bias cannot be accounted for by his view. In contrast, the qualief account explains these findings. Hence, Nanay´s view about mental imagery could benefit from being combined with the qualief account. In particular, holding that mental imagery can be used as a mode of presentation of a qualief, would result in an explanatorily more powerful view. Analyzing the possible combinations of these models with the qualief account is an interesting task, but for the lack of space I have to leave it to another paper.

[28]Here I assume that our experience of an individual seeming in a particular way is not restricted to low-level properties but can include high-level properties (such as being dangerous) as well (see, e.g., Bayne 2009; Siegel 2010; Toribio 2018). Does this mean that defenders of the view that the admissible contents of perception are restricted to low-level properties (e.g., Brogaard 2013; Tye 2018) cannot adopt the qualief account? It does not. The qualief account can be modified to be compatible with this view as well. One could flesh out the experience of an individual seeming F as an overall phenomenal experience consisting of a low-level sensory phenomenal element and high-level conceptual element deployed in judgement.

[29]For an alternative, intriguing, account of stereotypes as involving generic beliefs alongside other non-propositional contents, see Bosse 2022.

experiences, even if we do not have the explicit stereotype belief. We might actually disavow the stereotype belief and yet the exposure to stereotype representations exercises its influence on our experiences. As a result, members of the target groups seem to us to be a specific way.

At this point, one might wonder whether the influence of stereotype representations on our experiences counts as an instance of *cognitive penetration*. On the orthodox picture of cognitive penetrability, *doxastic* states influence experiences in a way that the experiences differ in phenomenal character and content even when external stimuli (and focal attention and fixation points) are held constant. In contrast, on the view developed here, the influence is not due to stereotype *beliefs* (which might be disavowed), but rather due to stereotypical representations such as images. This view – that the awareness of these representations shapes our experiences – is compatible with the thesis that the mind is modular. The stereotypical images could operate within the same module as the resulting experience (e.g. if we understand these modules as 'compiled transducers' (Fodor 1983, 41)). So even if our experiences are encapsulated relative to doxastic states, as the evidence-insensitivity suggests, they can still be open to influences within other modules. Importantly, the usage of such experiences when thinking about members of particular social groups often is epistemically problematic, for the experiences are rather reflecting the stereotypes than representing the individual accurately.

4.2. The generic content

In the previous section, I analyzed the experiences involved in the qualiefs underlying implicit bias. Next, I will focus on the content of those qualiefs.

Qualiefs can take various forms, e.g. qualiefs that attribute properties to *single* objects ('this car is red'), qualiefs that attribute properties to a *class* of objects (e.g. quantified generalizations such as 'all zebras are striped' or 'most dogs have tails' as well as generic generalization such as 'rattlesnakes are dangerous'), qualiefs that attribute properties to actions ('this jump is dangerous') etc.[30] This list is not meant to be exhaustive. For our present purposes, two kinds of qualiefs are particularly important:

1) *Singular* qualiefs
2) *Generic* qualiefs

[30]Thanks to an anonymous referee for pressing me on this point.

A *singular* qualief uses the experience of a particular external object to attribute a property to that very object. An example is the qualief *that this car is red*, where the mode of presentation of the content involves a red-experience. This builds on the assumption that an experience of an object *seeming* F supports the content that the object *is* F (Huemer 2001).

In contrast to singular qualiefs, a generic qualief implicitly operates on the experience of a *particular* to generate a *generic* content, e.g. the content that *rattlesnakes are dangerous*. Let me clarify that not all qualiefs which result from a generalizing process are generic qualiefs. However, as I will argue, generic qualiefs are the most plausible candidates to explain implicit bias. Therefore, in what follows, I will focus on generic qualiefs.

Generics have the form 'Fs are G', where G is supposedly, e.g. a normal (Nickel 2008), stereotypical (Declerk 1986), characteristic, or striking (Leslie 2008) property of the target group. Generics are often expressed by bare plurals (e.g. 'rattlesnakes are dangerous'), but they can also take the form of indefinite singulars ('A rattlesnake is dangerous') and of definite singulars ('The rattlesnake is dangerous'). (Analyzing definite and indefinite singular generics turns out to be a complicated task since they are often infelicitous where bare plurals are not. In what follows, I will focus on bare plurals.)

Most of the literature dedicated to generics deals with the question of how to analyze generics semantically. The standard model of generics posits a covert dyadic operator, *Gen,* which functions as an adverb of quantification (Lewis 1975).[31] Most theorists agree that generics are not reducible to, and are more basic than, quantifiers, but they disagree how *Gen* is best analyzed. Leslie (2008, 2012), for example, holds that *Gen* is semantically primitive and offers a disquotational semantics for *Gen* – a view which is criticized for not being able to account for the context-sensitivity of generics (Sterken 2015).[32] Sterken (2015, 2016) argues that *Gen* is an *indexical* over quantifiers. Nickel (2017) thinks that generics quantify over normal members of the kind, while Cohen (2004) understands generics in terms of comparative probabilities. Many other sophisticated theories of *Gen* have been developed (for an overview of the literature on generics, see Nickel forthcoming.) For present purposes, I will not focus on analyzing the existence and meaning of a covert

[31] Not all theorists posit *Gen* to explain generics. For example, Liebesman (2011) holds that generics are kind predications and Nguyen (2020) argues against a semantically effective operator *Gen* by proposing a pragmatic account of generic generalizations.

[32] On Sterken´s contextualist view of generics (2016), 'the truth-conditional variability of generics is not due to the complexity of some unified phenomenon of genericity, but rather to semantic context-sensitivity' (Plunkett, Sterken, and Sundell 2023, 50).

operator *Gen*. Rather, I will be concerned with the psychological mechanism that brings the generic generalization about.

Some theorists (e.g. Cimpian and Erickson 2012; Gelman 2010; Leslie 2012) think that the basis of generics can be found in a primitive psychological mode of generalizing that is prior to the acquisition of quantifiers. This hypothesis is supported by studies that show that young children understand generics more easily than quantifiers (Gelman et al. 2008) and that infants at the age of 30 months are already capable of forming generic generalizations (Leslie and Lerner 2016). Moreover, generic statements are more easily recalled in memory than overt quantificational statements in English (Gülgöz and Gelman 2015). Along with these empirical studies, Leslie (2012) argues that the absence in most languages of a word that articulates the 'Gen' operator also speaks in favor of the hypothesis that generics express basic generalizations: Presumably, a default way of generalizing might not require a word to signal a generic statement, whereas a deviation from the default mode might require an explicit instruction; e.g. by the word 'all' for processing universal statements. In line with these considerations, Leslie develops the 'generics-as-default hypothesis' (2012, 40), which has it that there is a fundamental, default mode of generalizing that picks up on characteristic or striking properties and links them to a kind. Following Leslie, I assume that the process that leads to generic generalizations is a basic, default cognitive mechanism. If so, then it is plausible that experiences of members of social groups can trigger this primitive generalizing mechanism. Accordingly, I suggest that there are generic qualiefs which are the result of a basic generalization mechanism.[33]

Next, my hypothesis is that such generic qualiefs explain implicit bias. The hypothesis that bias is closely linked to generics is not new (e.g. Wodak, Leslie, and Rhodes 2015; Hammond and Cimpian 2017; Leslie

[33]Not all theorists agree with Leslie. Sterken, for example, grants that there might be a primitive cognitive mechanism of generalization (Sterken 2015, 2494) but provides counterexamples against the thesis that *generics* express these cognitively primitive generalizations. If Sterken is right, the proposed account could be modified in the following way: one might hold that the qualiefs that realize implicit bias are not generic qualiefs but qualiefs that are the result of quantified generalizations. What is important for present purposes is the following: first, qualiefs that attribute properties to a class are the result of a primitive cognitive mechanism of generalization and they are easily triggered by experiences. Second, qualiefs that attribute properties to a class involve a phenomenal experience: this explains their evidence-insensitivity. Third, the content of these qualiefs is general and about a *kind*, whereas the mode of presentation is phenomenal and singular: this explains why the content is not easily introspectively accessible. These three claims about qualiefs realizing implicit bias are compatible with both the view that the relevant qualiefs have a generic content as well as that they have a quantified general content that is about a kind. (However, holding that implicit bias involves quantified generalization would have to restrict the account to quantified generalizations, which are about kinds. In contrast, the claim that the target content is about a kind fits naturally with *generic* contents).

2017). This view is often combined with the hypothesis that generics tend to *essentialize* the target group in way that overtly quantified generalizations do not (Haslam, Rothschild, and Ernst 2022; Gelman 2003; Rhodes, Leslie, and Tworek 2012).[34] For example, Haslanger (2011) argues that generics might be interpreted as falsely attributing an essential property to a social target group rather than a socially constructed one. If generics express our thinking of natural and social kinds as sharing a fundamental nature (which does not imply they necessarily share a *biological* nature), this suggests that generic contents are persistent. If a generics content is a content of a *qualief*, the phenomenal mode of presentation of the content reinforces this persistence, resulting in an insensitivity to counterevidence. This fits nicely with the findings about implicit bias and makes generic qualiefs a promising candidate to explain implicit bias.[35]

To recap: I suggest that there are singular and general qualiefs, among which generic qualiefs are a subset. In accordance with theorists like Leslie (2012, 2017), I think that generic generalizations are the result of a primitive cognitive mechanism and tend to essentialize the property attributed to the target group. Thus, generic *qualiefs* are also the result of a primitive cognitive mechanism and tend to essentialize the property attributed to the target group.

Next, let me further clarify the particularities of qualiefs underlying implicit bias. Generic qualiefs have two components – the experience (which is used to think about the target group) and the generic content. In implicit bias, both aspects display important particularities that differentiate them from other qualiefs: first, the experience involved is shaped by stereotype representations and, second, the content attributes essentialized properties to the target group. Notably, there is a further particularity which is crucial: the mode of presentation of the

[34] An important debate concerns the question whether generics that essentialize social groups should be rejected or can still be useful in some contexts. Langton, Haslanger, and Anderson hold that racial generic generalizations present 'social artifacts as racial essences' (2012, 765) and, thus, should be rejected as false and replaced by overt quantified statements. In contrast, Saul (2017) points to the usefulness of some social generics in our effort to establish social justice, while Ritchie (2019) has argued that some racial or gender generic generalizations more accurately describe structural oppression than overtly quantified sentences. While important, I have to leave these discussions aside, since I am mainly interested in the particular cognitive mechanism of generalizing that leads to generics.

[35] A clarification is in order here: In her analysis of generics and prejudice, Leslie (2017) focuses on a subclass of generics, namely those generics that generalize about a harmful or dangerous property. I do not think that the contents of generic qualiefs are restricted to involving negative properties. I rather think that generic qualiefs can also involve features that are considered as normal (e.g., 'women are nurturing'). On my view, what is key to generic qualiefs involved in implicit bias is not the kind of property about which we generalize but rather that it is seen as an *essential* property of the relevant social kind.

generic qualiefs involved in implicit bias differs from their contents in an important way. The mode of presentation involves an experience of a particular, whereas the generic content of the qualief is about a *kind*. Let me say more about this.

It is a widely held view that (at least some) generics do not generalize about individual members of a kind, but about the kind in general.[36] This can be fleshed out by distinguishing 'direct kind predications', which generalize over the target group in general (such as 'dinosaurs are extinct'), from 'characterizing generics' (Krifka 1987), which express generalizations about individual members of the kind (such as 'tigers are striped'). Some theorists provide powerful arguments for the view that characterizing generics in fact just are direct kind predicating generics (e.g. Liebesman 2011; Teichman 2019; Liebesman and Magidor 2017). Others defend the weaker view that characterizing generics – though not reducible to direct kind predications – still are about kinds in an important sense. Considerations that support this view include that characterizing generics differ significantly from quantified statements: e.g. they allow exceptions and are cognitively primitive (Leslie 2008; Nickel 2017).[37] For present purposes, it suffices to stick to the minimum assumption that the generics involved in implicit bias are about kinds. The idea that the generic content involved in implicit bias is about kinds fits well with the insight that it is the result of a primitive cognitive mechanism and that it attributes essentialized properties.[38]

4.3. Consequences of the overall structure of generic qualiefs

If our considerations so far are correct, then it becomes clear why the content of implicit bias is not easily introspectively accessible. Recall that a singular qualief uses an experience of an object *seeming F* to attribute to it the property of *being F*. In contrast, in a generic qualief there is a shift between the phenomenal mode of presentation and the content. The former involves an experience of a *particular* whereas the latter attributes a property to a *kind*.

The phenomenal mode of presentation draws the subject´s attention to the occurrent experience of the particular. Thus, the subject is

[36]For an illuminating discussion of this view and its implications for a metaphysics of kinds, see Liebesman and Sterken (2021).

[37]The view that generics are about kinds is often combined with the view that the target properties ascribed to the kind are understood as *normal* for the kind (Nickel 2017; Pelletier and Asher 2017) rather than *common* to the kind.

[38]Thanks to an anonymous referee for pressing me on this point.

primarily aware of the phenomenal experience of a particular. Importantly, this experience differs significantly from the generic content: the content is *not* phenomenal and it attributes a property to the target *kind*. As a result of these differences in both the phenomenality aspect and the content aspect, the generic content is eclipsed by the phenomenal mode of presentation. That means, it is the particular internal structure of generic qualiefs that makes it hard to introspectively access their contents. This characteristic of generic qualiefs is crucial for the present purposes and it explains how a phenomenal experience can be part of a mental state that is commonly held to be implicit. Accordingly, to analyze the introspective accessibility of a generic qualief, we have to look at its mode of presentation and at its content separately: First, the mode of presentation of a generic qualief is open to introspection, for it uses a phenomenal experience. Second, the generic content is not open to introspection in the same way, for it is not phenomenal and it is about a kind.

5. Explanatory power of qualiefs

I argued that implicit bias is best explained as a special kind of *generic qualief* – namely, as generic qualiefs that use experiences that have been shaped by stereotypes to attribute essentialized properties to a kind. Now I turn to the key features of implicit bias and demonstrate that the qualief model can account for them.

5.1. Key features of implicit bias

Implicit bias is subject to extensive empirical investigations that will provide us with new insights regarding its nature. The current findings suggest that implicit bias is a heterogenous phenomenon (Holroyd and Sweetman 2016, Johnson 2020 ; Del Pinal and Spaulding 2018) and some of its typical characteristics may not apply to all instances of implicit bias. However, any theory about implicit bias has to start somewhere. Therefore, in formulating the desiderata for a theory about implicit bias, I rely on a standard characterization of implicit bias that covers at least a broad range of cases. The main explanatory desiderata for a theory about implicit bias are its implicitness, automaticity and uncontrollability, insensitivity to evidence and its inferential role. Especially the latter two features taken together, which I label the biases´ '*asymmetric inferential profile*', are difficult to explain

on the extant models.[39] The qualief model is explanatorily powerful in these respects.

(A) *Implicitness*

First, recall the key feature that the bias is *implicit*. A common characterization of the implicitness has it that the bias is *not (easily) accessible via introspection*.[40] As Kelly & Roedder put it, neither 'introspection nor honest self-report are reliable guides to the presence of such mental states' (2008, 532). That implicit bias is not easily accessible introspectively even if a subject asks herself whether she harbors the relevant bias, differentiates the phenomenon from ordinary unconscious beliefs, such as one´s standing belief that 2 + 2 = 4 of which one can easily become aware as soon as one considers the target proposition. To shed light on this feature, it is helpful to discern three aspects of implicit bias that might be inaccessible via introspection: the *content* of the bias, its *source,* and its *impact* on our behavior and judgments (see Gawronski, Hofmann, and Wilbur 2006).

On the qualief account, we lack introspective access to the *source* and to the *impact* of implicit bias. However, we can become aware of these aspects via other methods (such as inferring from our behavior that we harbor a bias or learning from literature about its origins). With regard to the content, *one part* is open to introspection – namely its phenomenal mode of presentation –, but the generic content itself is not easily introspectively accessible. (This is a specific feature of *generic* qualiefs that differentiates them from singular qualiefs. If I qualieve that a *particular* man is dangerous, I can become easily aware of my attributing this property to the man due to him seeming so to me.)

Let me illustrate this feature of generic qualiefs by using Schwitzgebel (2010) example of Juliet, an aversive racist: 'When she gazes out on class

[39]This list of desiderata is not supposed to be exhaustive. (For an elaborated list of more desiderata, see Holroyd 2016.) Further explanatory desiderata are, e.g., to account for the motivational power of implicit bias and to show which methods to mitigate implicit bias appear promising and why. With regard to the first point, the qualief model offers an explanation of the motivational power, since a qualief involves a phenomenal experience which is a paradigmatic state that has motivational power. With regard to the second question, the qualief account provides a model of how to change implicit bias that then can be tested and supported empirically. Since the model explains the biases´ insensitivity to evidence, it thereby suggests methods other than rational argument for mitigating bias. In particular, methods such as counterstereotype exposure (Dasgupta and Greenwald 2001) and counterconditioning (Olson and Fazio 2006; Hu et el. 2017) appear promising, since they change the very experience involved in a qualief. For the lack of space, I have to leave a deeper analysis of these two further desiderata to another paper.

[40]Alternatively, bias can be held to be implicit because it does not figure in conscious reasoning (Mandelbaum 2016). For an analysis of the different interpretations of implicitness, see de Houwer 2014; Holroyd 2016.

the first day of each term, she can't help but think that some students look brighter than others – and *to her*, the black students never *look* bright.' (2010), 532 (my emphasis).[41] This description fits well with the qualief account of implicit bias. Juliet has an experience of a particular Black student not looking bright. This experience has been shaped by stereotype representations. Juliet is aware that this particular Black student does not look bright to her. (She may also form a corresponding *belief.)* Next, she might use this experience in a qualief. If so, she is aware of the phenomenal mode of presentation of her qualief.

However, Juliet is unaware that her experience of *this individual seeming F* kicks off the basic cognitive mechanism of generalization which results in the attribution of the relevant, essentialized, property to a social *kind*. Juliet is aware that this particular Black student does not look bright to her, but she is unaware of having a mental state with the generic content that *Black people are not bright*. In fact, she will deny this since she explicitly endorses anti-racist beliefs.

Notably, the tension between Juliet´s explicit anti-racist beliefs and her implicit racist bias reinforces the difficulty of introspectively accessing the generic content. This becomes clear when we differentiate between the *awareness of the mode of presentation* of implicit bias and the *endorsement of its content*. Plausibly, endorsement requires reflection. Since in qualieving Juliet is only aware that an individual seems F, she can reflect only on this aspect. She might consider whether to take this *individual* as being F and then endorse *this* content. But this need not conflict with explicit anti-racist beliefs. (After all, this particular Black student might not be bright, even though, in general, Black and White people are of equal intelligence). Moreover, Juliet is unaware of the generic content that the target kind is F. Hence, she is not in a position to reflect on and endorse this generic content. The only contents that would be open to her reflection and endorsement are a) the content of the singular qualief that an *individual* is F and b) the content of her explicit beliefs. As noted, these contents are not in tension with each other.

To recap: the phenomenal mode of presentation of implicit bias uses an experience of a particular, whereas the content is generic and about a (social) kind. Since the mode of presentation involves a phenomenal experience, it is *this* aspect which is salient and, since this aspect differs significantly from the content, it thereby occludes the content. Moreover,

[41] Schwitzgebel uses this example to illustrate his view of *in-between beliefs*, where beliefs are understood as dispositions. Since Juliet is disposed to endorse anti-racist views as well as to racist behavior and judgments, she in-between believes in racial equality.

there is a further factor that is accessible: one´s explicit beliefs. The mode of presentation of the bias and the explicit beliefs both differ significantly from the generic content of the bias and, therefore, prevent the subject from introspectively detecting the generic content. It is the particular structure of a generic qualief, combined with the awareness of explicit anti-discriminatory beliefs, which results in the unawareness of the bias.

Finally, let me note again that some theorists question the complete unawareness of implicit bias. For example, Hahn and colleagues (2014) found that people can predict their scores on prejudice IATs with a high degree of accuracy. Different explanations of this accuracy are available. On one view, this points towards the introspective accessibility of implicit bias. The explanation then is that people under normal circumstances are unwilling to report their bias, but believing that one´s bias will be uncovered leads to its admission. On an alternative explanation, one does not have directly access to one´s implicit bias but rather becomes aware of it by focusing on one´s affective reactions (Gawronski and Bodenhausen 2014). This explanation is supported by a recent study by Hahn and Gawronski (2019) that shows how merely attending to one's spontaneous affective reactions toward minority groups helps to acknowledge one´s bias.

The qualief account primarily aims at explaining those cases in which one is introspectively unaware of one´s implicit bias. Moreover, it is also compatible with the explanation that people predict their IAT scores via the awareness of affective reactions. The resulting picture is the following: the generic content of the implicit bias is not introspectively accessible and in this sense the bias is implicit. However, the phenomenal mode of presentation of the bias – the experience of an individual – is directly accessible. Plausibly, affective reactions are closely tied to this phenomenal mode of presentation and, hence, are accessible too. If one is asked to reflect about having a particular bias or not, one might focus on the affective reactions and doing so, one might find out about the the bias indirectly.

Thus, in contrast to other models that point at a doxastic state (or an imagining) that has a downstream phenomenal content to explain those cases of introspective accessibility, the qualief model can explain both the accessability cases *and* the cases that suggest an unawareness of the bias. The picture is the following: Given the particular structure of a generic qualief, one tends to be unaware of the generic content of one´s implicit bias. This explains those findings that suggest unawareness of implicit bias. However, if asked to reflect on whether one harbors a bias, the possibility of the relevant bias becomes salient. This process of reflecting on a salient possibility might draw the attention to the phenomenal

aspect and, as a result, one might find out about one´s bias. Thus, the nature of a generic qualief – that it presents a propositional content in a phenomenal way – offers a framework on which we can explain the unawareness of implicit bias as well as findings that suggest that under certain circumstances one might become aware of one´s bias.

(B) Automaticity and uncontrollability

A further characteristic of implicit bias is that it is automatic in origin, but also persistent and not under our direct control. We can influence our implicit bias in some way, as the literature on methods to mitigate bias shows (Madva 2016b; Byrd 2019). However, these methods influence the bias only indirectly and the effect is often only short-term (Lai et al. 2016). Any account of implicit bias should explain these features, and the qualief model meets this requirement.

Recall that a phenomenal experience is a constitutive part of every qualief. Phenomenal experiences are the paradigmatic kind of states that are automatically induced via incoming stimuli and that are not under our direct control. Accordingly, if we think in terms of an experience about external referents, we cannot deliberately choose to have or to change the very experience which is part of the qualief. Furthermore, as long as we are exposed to stereotype representations, we cannot deliberately choose whether those representations exercise their influence on our experiences. The only factor that might be under our control is blocking the causal impact of the qualief on our behavior and judgments, once we have learned about that impact via methods other than introspection.

(C) The asymmetric inferential profile of implicit bias

Finally, the qualief model sheds light on the *asymmetric inferential profile* of implicit bias. Let me explain.

Recent experimental findings point towards two aspects of implicit bias that seem hard to reconcile on the extant models. On the one hand, implicit bias seems insensitive to logical form (Madva 2016a; Gawronski et al. 2008) and insensitive to evidence and reasoning in most cases (Gregg, Seibt, and Banaji 2006).[42] On the other hand, the

[42]For an alternative explanation of these findings and the view that implicit bias can be changed via argument, see Mandelbaum 2016. Moreover, Kurdi and Banaji (2019) found out that verbal information sometimes shifts implicit bias. Del Pinal and Spaulding (2018) defend a concept-centrality account of implicit bias that can explain why in some cases implicit bias is more evidence-insensitive than in

content of implicit bias can serve as propositional input to further mental states (Gawronski, Hofmann, and Wilbur 2006; Mandelbaum 2016). Propositional models cannot easily explain the former findings, associative models fail to account for the latter. The qualief model can explain both.

First, the qualief model of implicit bias accounts for the insensitivity to logical form. For example, Gawronski et al. (2008, 376) show that thinking, 'it is not true that old people are bad drivers,' reinforces rather than undermines a negative implicit attitude toward elderly drivers. The proposed account explains this insensitivity to logical form: the sentence evokes an (imaginative) experience of an elderly person driving badly which then is used in the qualief.[43] Moreover, the qualief account explains the biases´ insensitivity to evidence. A phenomenal experience – which is paradigmatically evidence-insensitive – is used as the mode of presentation of a qualief. Since the experience is a constitutive part of a qualief, it carries its evidence-insensitivity over to the overall mental state. Notably, this explanation of the evidence insensitivity differs significantly from the ones we find in the literature. For example, Levy (2015) analyzes implicit bias as 'patchy endorsements' that are not reason-responsive because they are not well-integrated within the inferential network. Similarly, Egan (2011) holds that insensitivity to evidence is due to these doxastic states´ being fragmented. In contrast, on the qualief account, the reason for the evidence insensitivity is found in the internal structure of the mental state itself (i.e. that it is partly constituted by an experience) rather than in its relation to the overall inferential network.

Second, since generic qualiefs have a propositional content, they can serve as input to further mental states. For instance, if the implicitly racist teacher qualieves that Black people are stupid, she can infer from this qualief that she had better not ask a Black student to become her teaching assistant. Notably, she will not be aware of that inference since the generic content, which serves as input, is not introspectively accessible to her.

In short: the qualief model can explain experimental data concerning the biases´ insensitivity to logical form and to evidence as well as

others. On their view, implicit biases are encoded in dependency networks that are part of our representations of social categories. Depending on the concept-centrality, implicit bias may take substantially different logical forms and may exhibit different degrees of stability and, hence, evidence insensitivity. Given these findings, I confine myself to the weaker claim that in *many* cases implicit bias is not responsive to evidence.

[43]Nanay (2021) gives a similar explanation by noting that negation operation has no influence of mental imagery.

findings that point towards the fact that implicit bias can figure in inferences. Data that, taken together, pose a difficulty for both the propositional and the associative models.

6. Conclusion

I have proposed a novel account of implicit bias that accommodates both its phenomenal and its propositional aspect. In the literature, these two aspects have been accounted for mainly separately – either by the associative or by the propositional models. The qualief model reconciles these aspects by pointing towards a propositional content, though represented in a phenomenal way. Therefore, this model accounts for the heterogenous character of implicit bias. In contrast to other views that aim at explaining the heterogenous character by invoking one single state that covers multiple processes or states, the qualief account does not face the challenge to motivate such unity. Rather, it provides an *explanation* of the tight connection between the phenomenal and the propositional aspect of implicit bias.

I have argued that implicit bias is a specific instance of a generic qualief and that this model is explanatorily powerful. First, qualiefs are partly constituted by phenomenal experiences – this explains why implicit bias is automatic, not under our direct control, and hard to regulate via cognitive means. Second, as a result of the generic generalization, there is a shift between the phenomenal mode of presentation of a qualief, which involves an experience of a *particular*, and its content, which is a *generic* and attributes essentialized properties to a kind. This accounts for the unawareness of the biases´ generic content. Finally, the qualief model does justice to the asymmetric inferential profile of implicit bias by showing how the bias can be insensitive to logical form and evidence, but at the same time it can serve as propositional input to further mental states.

Now that we have traced the tenacity of implicit bias to its roots – which lie in experiences that have been shaped by stereotypes and that are used to (mis)attribute essentialized properties to social groups – we have a fixed point from which to explore further important questions, such as: which new strategies can be developed to mitigate implicit bias?

Acknowledgements

I am grateful to Marian David, Terry Horgan, Keith Lehrer, and Guido Melchior for insightful discussions of earlier drafts of this paper. For helpful feedback on the

paper, I am indebted to Rebecca Davis and Wes Siscoe. Many thanks to the audience of the Department Colloquium at the University of Arizona – the paper benefitted greatly from their comments. I presented earlier versions of the paper at the workshop 'Dissonance and Implicit Bias', University of Graz, the 'Epistemology Conference: Epistemic Virtues and Epistemic Skills', Bled; the 'Metaphysics Conference', Inter University Center Dubrovnik and the 9th SEFA-Conference, University of Valencia. I am grateful to the participants and the audiences for their valuable comments. Finally, I want to express my gratitude to all anonymous referees involved. This research has been supported by the Austrian Science Fund -Project P33710.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research has been supported by the Austrian Science Fund Project [grant number P33710].

ORCID


Martina Fürst http://orcid.org/0000-0001-6337-231X

References

Balog, Katalin. 2012. "In Defense of the Phenomenal Concept Strategy." *Philosophy and Phenomenological Research* 84 (1): 1–23. doi:10.1111/j.1933-1592.2011.00541.x.

Bayne, Tim.2009. "Perception and the Reach of Phenomenal Content." *The Philosophical Quarterly* 59 (236): 385–404. doi:10.1111/j.1467-9213.2009.631.x.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013. doi:10.1257/0002828042002561.

Block, Ned. 2007. "Max Black's Objection to Mind-Body-Identity." In *Phenomenal Concepts and Phenomenal Knowledge*, edited by T. Alter and S. Walter, 249–306. Oxford: Oxford University Press.

Bortolotti, Lisa. 2009. *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.

Bosse, Anne. 2022. "Stereotyping and Generics." *Inquiry*.

Briñol, Pablo, Richard E. Petty, and Michael J. McCaslin. 2009. "Changing Attitudes on Implicit Versus Explicit Measures." In *Attitudes: Insights from the New Implicit Measures*, edited by R. Petty, R. Fazio and P. Brinol, 285–326. New York: Psychology Press.

Brogaard, Berit. 2013. "Do we Perceive Natural Kind Properties?" *Philosophical Studies* 162 (1): 35–42. doi:10.1007/s11098-012-9985-5.

Brownstein, Michael, and Alex Madva. 2018. "Stereotypes, Prejudice, and the Taxonomy of the Implicit Social Mind." *Noûs* 52 (3): 611–644.

Brownstein, Michael, and JenniferSaul, eds. 2016. *Implicit Bias & Philosophy*. New York: Oxford University Press.

Byrd, Nick. 2019. "What we Can (and Can't) Infer About Implicit Bias from Debiasing Experiments." *Synthese* (2): 1–29.

Carruthers, Peter, and Benedicte Veillet. 2007. "The Phenomenal Concept Strategy." *Journal of Consciousness Studies* 14 (9–10): 212–236.

Chalmers, David.2003. "The Content and Epistemology of Phenomenal Belief." In *Consciousness: New Philosophical Perspectives*, edited by Q. Smith, and A. Jokic. Oxford: Oxford University Press.

Chalmers, David. 2007. "Phenomenal Concepts and the Explanatory Gap." In T. Alter & S. Walter (eds.), 167–154. OUP.

Cimpian, A., and L. C. Erickson. 2012. "Remembering Kinds: New Evidence That Categories are Privileged in Children's Thinking." *Cognitive Psychology* 64 (3): 161–185. doi:10.1016/j.cogpsych.2011.11.002.

Cohen, A.2004. "Generics and Mental Representation." *Linguistics and Philosophy* 27 (5): 529–556. doi:10.1023/B:LING.0000033851.25870.3e.

Currie, Gregory, and Anna Ichino. 2012. "Aliefs Don't Exist, But Some of Their Relatives Do." *Analysis* 72: 788–798. doi:10.1093/analys/ans088.

Dasgupta, Nilanjana, and Anthony Greenwald. 2001. "On the Malleability of Automatic Attitudes: Combating Automatic Prejudice with Images of Admired and Disliked Individuals." *Journal of Personality and Social Psychology* 81 (5): 800–814. doi:10.1037/0022-3514.81.5.800.

Declerk, R. 1986. "The Manifold Interpretations of Generic Sentences." *Lingua. International Review of General Linguistics. Revue internationale De Linguistique Generale* 68: 149–188. doi:10.1016/0024-3841(86)90002-1.

De Houwer, Jan.2014. "A Propositional Model of Implicit Evaluation." *Social and Personality Psychology Compass* 8 (7): 342–353. doi:10.1111/spc3.12111.

Del Pinal, Guillermo, and Shannon Spaulding. 2018. "Conceptual Centrality and Implicit Bias." *Mind & Language* 33 (1): 95–111. doi:10.1111/mila.12166.

Dovidio, John F., and Samuel L Gaertner. 2000. "Aversive Racism and Selection Decisions: 1989 and 1999." *Psychological Science* 11: 319–323.

Eberhardt, Jennifer L., Phillip A. Goff, Valerie J. Purdie, and Paul G. Davies. 2004. "Seeing Black: Race, Crime, and Visual Processing." *Journal of Personality & Social Psychology* 87 (6): 876–893. doi:10.1037/0022-3514.87.6.876.

Egan, Andy.2011. "Comments on Gendler's The Epistemic Costs of Implicit Bias." *Philosophical Studies* 156 (1): 65–79. doi:10.1007/s11098-011-9803-5.

Evans, J., and K. Frankish. 2009. *In Two Minds: Dual Processes and Beyond*. Oxford: Oxford University Press.

Evans, J., and David E. Over. 2004. *If*. Oxford: Oxford University Press.

Evans, J., and K. Stanovich. 2013. "Dual-Process Theories of Higher Cognition: Advancing the Debate." *Perspectives on Psychological Science* 8 (3): 223–241. doi:10.1177/1745691612460685.

Fodor, Jerry A.1983. *Modularity of the Mind*. Cambridge, MA: MIT Press.
Fürst, Martina.2014. "A Dualist Account of Phenomenal Concepts." In *Contemporary Dualism: A Defense*, edited by Andrea Lavazza, and Howard Robinson, 112–136. London: Routledge.
Fürst, Martina. forthcoming a. "Phenomenal Holism and Cognitive Phenomenology." Erkenntnis.
Fürst, Martina.forthcoming b. "Closing the conceptual gap in epistemic injustice". Philosophical Quarterly.
Gawronski, Bertram. 2019. "Six Lessons for a Cogent Science of Implicit Bias and its Criticism." *Perspectives on Psychological Science* 14: 574–595. doi:10.1177/1745691619826015.
Gawronski, Bertram, and Galen V Bodenhausen. 2006. "Associative and Propositional Processes in Evaluation: An Integrative Review of Implicit and Explicit Attitude Change." *Psychological Bulletin* 132 (5): 692–731. doi:10.1037/0033-2909.132.5.692.
Gawronski, Bertram, and Galen V.Bodenhausen. 2014. "The Associative—Propositional Evaluation Model: Operating Principles and Operating Conditions of Evaluation." In *Sherman, Gawronski, & Trope, Dual-Process Theories of the Social Mind*, 188–203. Guilford Press.
Gawronski, Bertram, Roland Deutsch, Sawsan Mbirkou, Beate Seibt, and Fritz Strack. 2008. "When "Just Say No" is not Enough: Affirmation Versus Negation Training and the Reduction of Automatic Stereotype Activation." *Journal of Experimental Social Psychology* 44: 370–377. doi:10.1016/j.jesp.2006.12.004.
Gawronski, Bertram, Wilhelm Hofmann, and Christopher Wilbur. 2006. "Are "Implicit" Attitudes Unconscious?" *Consciousness and Cognition* 15: 485–499. doi:10.1016/j.concog.2005.11.007.
Gelman, S. A. 2003. *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford Cognitive Development.
Gelman, S. A. 2010. "Generics as a Window Onto Young Children's Concepts." In *Kinds, Things, and Stuff: The Cognitive Side of Generics and Mass Terms*, edited by F. J. Pelletier, 100–112. New York: Oxford University Press.
Gelman, S. A., P. J. Goetz, B. S. Sarnecka, and J. Flukes. 2008. "Generic Language in Parent-Child Conversations." *Language Learning and Development* 4: 1–31. doi:10.1080/15475440701542625.
Gendler, Tamar Szabó. 2008a. "Alief and Belief." *Journal of Philosophy* 105: 634–663. doi:10.5840/jphil20081051025.
Gendler, Tamar Szabó. 2008b. "Alief in Action (and Reaction)." *Mind & Language* 23 (5): 552–585. doi:10.1111/j.1468-0017.2008.00352.x.
Gregg, A. P., B. Seibt, and M. R. Banaji. 2006. "*Easier Done Than Undone: Asymmetry in the Malleability of Implicit Preferences*." *Journal of Personality and Social Psychology* 90 (1): 1–20. doi:10.1037/0022-3514.90.1.1.
Gülgöz, S., and S. A. Gelman. 2015. "Children's Recall of Generic and Specific Labels Regarding Animals and People." *Cognitive Development* 33: 84–98. doi:10.1016/j.cogdev.2014.05.002.
Hahn, A., and Bertrand Gawronski. 2019. "Facing One's Implicit Biases: From Awareness to Acknowledgment." *Journal of Personality and Social Psychology* 116 (5): 769–794. doi:10.1037/pspi0000155.

Hahn, A., Charles M. Judd, Holen K. Hirsh, and Irene, V. Blair. 2014. "Awareness of Implicit Attitudes." *Journal of Experimental Psychology: General* 143 (3): 1369–1392. doi:10.1037/a0035028.

Hammond, Matthew D., and A. Cimpian. 2017. "Investigating the Cognitive Structure of Stereotypes: Generic Beliefs About Groups Predict Social Judgments Better Than Statistical Beliefs." *Journal of Experimental Psychology: General* 146 (5): 607–614. doi:10.1037/xge0000297.

Haslam, Nick, Louis Rothschild, and Donald Ernst. 2022. "Are Essentialist Beliefs Associated with Prejudice?" *British Journal of Social Psychology* 41: 87–100. doi:10.1348/014466602165072.

Haslanger, Sally. 2011. "Ideology, Generics, and Common Ground." In *Feminist Metaphysics: Explorations in the Ontology of Sex, Gender, and the Self*, edited by C. Witt, 179–209. ct: Springer.

Helton, Grace. 2020. "If You Can't Change What You Believe, You Don't Believe It". *Noûs* 54 (3): 501–526

Holroyd, Jules. 2016. "What Do We Want from a Model of Implicit Cognition?" *Proceedings of the Aristotelian Society* 116 (2): 153–179. doi:10.1093/arisoc/aow005.

Holroyd, Jules, Robin Scaife, and Tom Stafford. 2017. "What is Implicit Bias?" *Philosophy Compass*, doi:10.1111/phc3.12437.

Holroyd, Jules, and JosephSweetman. 2016. "The Heterogeneity of Implicit Bias." In *Implicit Bias and Philosophy*. Vol. 1: Metaphysics and Epistemology, edited by Michael Brownstein and Jennifer Saul, 80–103. New York: Oxford University Press.

Hu, Xiaoqing, Bertram Gawronski, and Robert Balas. 2017. "Propositional Versus Dual-Process Accounts of Evaluative Conditioning: II. The Effectiveness of Counter-Conditioning and Counter-Instructions in Changing Implicit and Explicit Evaluations." *Social Psychological and Personality Science* 8 (8): 858–866. doi:10.1177/1948550617691094.

Huemer, Michael. 2001. *Skepticism and the Veil of Perception*. Rowman & Littlefield.

Johnson, Gabbrielle M. forthcoming. "The Structure of Bias." *Mind; A Quarterly Review of Psychology and Philosophy*.

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kelly, Daniel, and E. Roedder. 2008. "Racial Cognition and the Ethics of Implicit Bias." *Philosophy Compass* 3 (3): 522–540. doi:10.1111/j.1747-9991.2008.00138.x.

Kind, Amy. 2001. "Putting the Image Back in Imagination." *Philosophy and Phenomenological Research* 62: 85–109. doi:10.1111/j.1933-1592.2001.tb00042.x.

Krifka, M. in collaboration with Claudia Gerstner 1987. *An Outline of Genericity in collaboration with Claudia Gerstner, SNS-Bericht 87–23*. University of Tübingen.

Kurdi, B., and M. R. Banaji. 2019. "Attitude Change via Repeated Evaluative Pairings Versus Evaluative Statements: Shared and Unique Features." *Journal of Personality and Social Psychology* 116 (5): 681–703. doi:10.1037/pspa0000151.

Lai, Calvin K., Allison L. Skinner, Erin Cooley, Sohad Murrar, Markus Brauer, Thierry Devos, Jimmy Calanchini, et al. 2016. "Reducing Implicit Racial Preferences: Ii. Intervention Effectiveness Across Time." *Journal of Experimental Psychology: General* 145 (8): 1001–1016. doi:10.1037/xge0000179.

Langton, R., S. Haslanger, and L. Anderson. 2012. "Language and Race." In *The Routledge Companion to Philosophy of Language*, edited by G. Russell, and D. Graff Fara, 753–767. New York: Routledge.
Lehrer, Keith. 2019. *Exemplars of Truth*. Oxford: Oxford University Press.
Leslie, Sarah-Jane. 2008. "Generics: Cognition and Acquisition." *The Philosophical Review* 117 (1): 1–47. doi:10.1215/00318108-2007-023.
Leslie, Sarah-Jane. 2012. "Generics Articulate Default Generalizations." *Recherches linguistiques de Vincennes* 41: 25–44.
Leslie, Sarah-Jane. 2017. "The Original Sin of Cognition." *Journal of Philosophy* 114 (8): 393–421. doi:10.5840/jphil2017114828.
Leslie, Sarah-Jane, and Adam Lerner. 2016. "Generic Generalizations". The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2016/entries/generics/.
Levin, Janet. 2007. "What is a Phenomenal Concept?" In *Phenomenal Concepts and Phenomenal Knowledge*, edited by Torin Alter and Sven Walter, 87–111. Oxford: Oxford University Press.
Levy, Neil.2015. "Neither Fish nor Fowl: Implicit Attitudes as Patchy Endorsements." *Noûs* 49 (4): 800–823.
Lewis, David. 1975. "Adverbs of Quantification." In *Formal Semantics of Natural Language*, edited by E. L. Keenan, 3–15. Cambridge: Cambridge University Press.
Liebesman, D.2011. "Simple Generics." *Nous (detroit, Mich)* 45 (3): 409–442.
Liebesman, D., and O. Magidor. 2017. "Copredication and Property Inheritance." *Philosophical Issues* 27 (1): 131–166. doi:10.1111/phis.12104.
Liebesman, David, and Rachel K.Sterken. 2021. "Generics and the Metaphysics of Kinds." *Philosophy Compass* (7): 1–14.
Loar, Brian, 1997. "Phenomenal States." In *The Nature of Consciousness*, edited by Ned Block. et al. MIT Press.
Madva, Alex. 2016b. "Virtue, Social Knowledge, and Implicit Bias", 191–215. In: Brownstein & Saul.
Madva, Alex. 2016. "Why Implicit Attitudes are (Probably) not Beliefs." *Synthese* 193: 2659–2684. doi:10.1007/s11229-015-0874-2.
Mandelbaum, Eric. 2013. "Against Alief." *Philosophical Studies* 165 (1): 197–211.
Mandelbaum, Eric. 2016. "Attitude, Inference, Association: On the Propositional Structure of Implicit Bias." *Noûs* 50 (3): 629–658.
Mitchell, Chris J., Jan De Houwer, and Peter F. Lovibond. 2009. "The Propositional Nature of Human Associative Learning." *Behavioral and Brain Sciences* 32 (02): 183–198. doi:10.1017/S0140525X09000855.
Nanay, Bence. 2021. "Implicit Bias as Mental Imagery." *Journal of the American Philosophical Association* 7 (3): 329–347.
Nguyen, A. 2020. "The Radical Account of Bare Plural Generics." *Philosophical Studies* 177: 1303–1331. doi:10.1007/s11098-019-01254-8.
Nickel, Bernhard. 2008. "Generics and the Ways of Normality." *Linguistics and Philosophy* 31 (6): 629–648.
Nickel, B. 2016. *Between Logic and the World: An Integrated Theory of Generics*. Oxford: Oxford University Press.

Nickel, B. 2017. "Generics." In *The Blackwell Companion to the Philosophy of Language, 2nd ed.*, edited by B. Hale, A. Miller, and C. Wright. Blackwell.

Nier, J. 2001. "How Dissociated are Implicit and Explicit Racial Attitudes? A Bogus Pipeline Approach." *Group Processes & Intergroup Relations* 8 (1): 39–52. doi:10.1177/1368430205048615.

Olson, Michael, and Russell Fazio. 2006. "Reducing Automatically-Activated Racial Prejudice through Implicit Evaluative Conditioning." *Personality and Social Psychology Bulletin* 32 (4): 421–433. doi:10.1177/0146167205284004.

Papineau, David. 2007. "Phenomenal and Perceptual Concepts." In *Phenomenal Concepts and Phenomenal Knowledge*, edited by T. Alter and S. Walter, 111–145. Oxford: Oxford University Press.

Payne, Keith. 2006. "Weapon Bias: Split-second Decisions and Unintended Stereotyping." *Current Directions in Psychological Science* 15 (6): 287–291. doi:10.1111/j.1467-8721.2006.00454.x.

Pelletier, F. J., and N. Asher. 2017. "Generics and Defaults." In *Handbook of Logic and Language*, edited by J. van Benthem, and A. ter Meulen, 1125–1177. Elsevier.

Plunkett, D., R. K. Sterken, and T. Sundell. 2023. "Generics and Metalinguistic Negotiation." *Synthese* 201: 50. doi:10.1007/s11229-022-03862-0.

Pylyshyn, Zenon W. 1999. "Is Vision Continuous with Cognition?: The Case for Cognitive Impenetrability of Visual Perception." *Behavioral and Brain Sciences* 22: 341–423. doi:10.1017/S0140525X99002022.

Rhodes, M., S.-J. Leslie, and C. M. Tworek. 2012. "Cultural Transmission of Social Essentialism." *Proceedings of the National Academy of Sciences* 109 (34): 13526–13531. doi:10.1073/pnas.1208951109.

Ritchie, Katherine. 2019. "Should we use Racial and Gender Generics?" *Thought: A Journal of Philosophy* 8 (1): 33–41. doi:10.1002/tht3.402.

Rooth, D. O. 2010. "Automatic Associations and Discrimination in Hiring: Real World Evidence." *Labour Economics* 17 (3): 523–534. doi:10.1016/j.labeco.2009.04.005.

Saul, Jennifer. 2013. "Implicit Bias, Stereotype Threat, and Women in Philosophy." In *Women in Philosophy*, edited by K. Hutchinson, and F. Jenkins, 39–60. Oxford: Oxford University Press.

Saul, Jennifer. 2017. "Are Generics Especially Pernicious?" *Inquiry*, 1–18. doi:10.1080/0020174X.2017.1285995.

Schwitzgebel, Eric. 2010. "Acting contrary to our professed beliefs or the gulf between occur- rent judgment and dispositional belief." *Pacific Philosophical Quarterly* 91 (4): 531–553. doi:10.1111/j.1468-0114.2010.01381.x.

Siegel, Susanna. 2010. *The Contents of Visual Experiences*. Oxford: Oxford University Press.

Sloman, Steven A. 1996. "The Empirical Case for Two Systems of Reasoning." *Psychological Bulletin* 119: 3–22. doi:10.1037/0033-2909.119.1.3.

Stanley, Jason, and Timothy Williamson. 2001. "Knowing How." *Journal of Philosophy* 98 (8): 411–444.

Sterken, R. 2015. "Leslie on Generics." *Philosophical Studies* 172 (9): 2493–2512. doi:10.1007/s11098-014-0429-2.

Sterken, R. 2016. "Generics, Covert Structure and Logical Form." *Mind and Languag* 31 (5): 503–529. doi:10.1111/mila.12118.

Sullivan-Bissett, Ema. 2019. "Biased by our Unconscious Imaginings." *Mind and Language* 34 (5): 627–647. doi:10.1111/mila.12225.

Teichman, M. 2019. "The Sophisticated Kind Theory." *Inquiry*, 1–47. doi:10.1080/0020174X.2016.1267407.

Toribio, Josefa. 2018. "Visual Experience: Rich but Impenetrable." *Synthese* 195 (8): 3389–3406. doi:10.1007/s11229-015-0889-8.

Tye, M. 2000. *Consciousness, Color and Content*. Cambridge, MA: MIT Press.

Van Dessel, P., Y. Ye, and J. De Houwer. 2018. "Changing Deep-Rooted Implicit Evaluation in the Blink of an Eye: Negative Verbal Information Shifts Automatic Liking of Gandhi." *Social Psychological and Personality Science* 1948550617752064.

Velleman, David. 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Welpinghus, Anna. 2020. "The Imagination Model of Implicit Bias." *Philosophical Studies* 177: 1611–1633. doi:10.1007/s11098-019-01277-1.

Wodak, D., S.-J. Leslie, and M. Rhodes. 2015. "What a Loaded Generalization: Generics and Social Cognition." *Philosophy Compass* 10 (9): 625–635. doi:10.1111/phc3.12250.