Episteme (2025), **22**, 707–730
doi:10.1017/epi.2025.17

episteme

ARTICLE

Identity-protective reasoning: an epistemic and political defense

Carolina Flores

University of California Santa Cruz, Santa Cruz, CA, USA
Email: caro.flores@ucsc.edu

(Received 17 February 2025; accepted 21 February 2025; first published online 26 March 2025)

Abstract

Identity-protective reasoning – motivated reasoning driven by defending a social identity – is often dismissed as a paradigm of epistemic vice and a key driver of democratic dysfunction. Against this view, I argue that identity-protective reasoning can play a positive epistemic role, both individually and collectively. Collectively, it facilitates an effective division of cognitive labor by enabling groups to test divergent beliefs, serving as an epistemic insurance policy against the possibility that the total evidence is misleading. Individually, it can correct for the distortions that arise from taking ideologically skewed evidence at face value. This is particularly significant for members of marginalized groups, who frequently encounter evidence that diminishes the value of their identities, beliefs, and practices. For them, identity-protective reasoning can counter dominant ideological ignorance and foster resistant standpoint development. While identity-protective reasoning is not without risks, its application from marginalized and counter-hegemonic positions carries epistemic benefits crucial in democracies threatened by elite capture. Against dominant views in contemporary political epistemology and psychology, identity-protective reasoning should be reconceived as a resource to be harnessed and not a problem to be eradicated.

Keywords: Identity-protective reasoning; motivated reasoning; group cognition; political epistemology; standpoint

1. Introduction

Here is a seeming truism: one's beliefs should be driven only by the evidence and never by social identities (# FactsNotFeelings). As far as getting at the truth is concerned, one ought not to interact with evidence with the aim of defending a social identity. In addition to being thought to be individually irrational, such identity-protective reasoning is frequently also considered collectively noxious, insofar as it splinters society into polarized factions each with their own non-negotiable take on reality.

In this paper, I will argue against this gloomy picture of identity-protective reasoning. My argument is two-pronged.

First, I will argue that identity-protective reasoning has distinctive benefits at the collective epistemic level, enabling the preservation and development of minority views

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

that go against the total balance of evidence at a time. This provides a valuable insurance policy against the possibility that the total balance of evidence is misleading.

Second, in contexts of oppression, the balance of evidence is often ideologically skewed against the interests of marginalized groups. In such contexts, identity-protective reasoning from marginalized positions can be conducive to veritistic aims. Applying the idea that epistemic norms in non-ideal contexts can differ from those in ideal contexts, I will argue that identity-protective reasoning is epistemically permissible in such circumstances. In situations of oppression, subjects are not only ethically, but also epistemically licensed to resist the dominant gaze.[1]

To be clear, I will not argue that identity-protective reasoning is never to blame for epistemic or societal ills. In fact, identity-protective reasoning often consists of little more than grasping for fabrications at the service of dominant ideological ignorance. In such cases, it is epistemically impermissible and politically toxic. At the same time, identity-protective reasoning can shield against systematically misleading evidence. In doing so, it can support the articulation of marginalized standpoints that pierce through the veil of ideology. As such, it has an under-noticed potential for epistemic resistance.

My arguments in this paper can be seen as part of a larger project of vindicating cognition driven by group attachments.[2] In this vein, some have argued that epistemic bubbles and outgroup distrust sometimes constitute reliable filters on disinformation (Coady 2024; Lackey 2021; Nguyen 2021; Westfall 2024). Others have provided models of how relying on pre-existing views in assessing information can protect us from being misled and help us efficiently use cognitive resources (Coady 2024; Dorst 2023; Westfall 2024). Finally, perhaps group membership is at least sometimes about shared values and interests that encroach on how we interact with evidence (Lepoutre 2020). However, none of these views challenge the claim that motivated reasoning driven by social identities is unavoidably pernicious. This paper aims to show that, in fact, identity-protective reasoning can bring distinctive epistemic benefits at both the individual and collective levels.

To bring out these positive epistemic roles, I will proceed as follows. In §2, I will briefly survey existing literature, describing both what identity-protective reasoning consists in and why it is considered problematic. In §3, I draw on discussions by Hallsson and Kappel (2020) and Lepoutre (2020) of the collective value of motivated reasoning and dogmatic group cognition to offer a line of defense of identity-protective reasoning based on its collective epistemic benefits. In §4, I turn to arguing that identity-protective reasoning can promote veritistic aims in conditions where evidence is systematically biased. Based on this point and on the idea that epistemic norms in non-ideal contexts can differ from those at play in ideal contexts, I will argue that identity-protective reasoning is epistemically permissible in such cases. In §5, I offer some more speculative remarks on the role of identity-protective reasoning in democracy. The mainstream view holds that polarization is the key problem in contemporary democracies. Given that identity-protective reasoning can entrench polarization, it looks like a villain to eliminate. In contrast, when we see elite capture as the main challenge democracies face (Bagg 2024), identity-protective reasoning emerges as a resource for epistemic resistance.

[1]Sara Ahmed writes: "The moment of queer pride is a refusal to be shamed by witnessing the other as being ashamed of you. The queer who is happily queer still encounters a world that is unhappy with queer love, but refuses to be made unhappy by that encounter" (Ahmed 2020, 116–17). In a sense, this article develops this insight about the value of resisting the dominant ideological perspective into a non-ideal epistemology of identity-protection.

[2]I provided a defense of this approach aimed at general audiences in Flores 2023.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

2. The psychology of identity-protective reasoning and the received view in epistemology

When we encounter evidence that challenges beliefs tied to an important part of our identity – such as our political affiliation, nationality, or gender – our desire to protect that identity often makes us reluctant to change our views. We engage in *identity-protective reasoning*.

This is a familiar and widely studied phenomenon. For instance, men for whom masculinity is closely related to meat-eating often respond defensively to evidence suggesting that meat-eating is not necessary for health (Piazza et al. 2015). Many progressive climate activists do not budge on the risks associated with nuclear energy, even in light of evidence that they might be over-estimating those risks. Republicans and Democrats in the USA maintain radically different risk assessments of permissive gun laws (among many other topics) (e.g., Kahan et al. 2011; Kahan 2012; Kahan 2016) even if given the same evidence. Generalizing:

> **Identity-protective reasoning:** A subject *S* engages in *identity-protective reasoning* with respect to *p* when *S* interacts with evidence bearing on *p* in ways influenced by motivation to defend a social identity that matters to *S*, where that social identity is connected to a specific take on *p*.[3]

This needs to be unpacked. First, a mandatory technical point: by *evidence a subject has* I mean any considerations of which the subject is aware that make a difference to what the subject is justified in believing (Kelly 2008; Conee and Feldman 2004). Evidence for *p* probabilifies *p* and evidence against *p* makes *p* less likely to be true, relative to the subject's overall evidence (Kelly 2016). This conception of evidence is controversial, in that it is non-factive and limits what counts as evidence for a subject to only what they are aware of.[4] I assume this conception of evidence as I think it best fits research on identity-protective reasoning, which typically focuses on the psychological effects of considerations of which the subject is aware, regardless of their truth value. I will sometimes talk of "genuine evidence" to refer to the subset of the agent's evidence that is factive.

Second, identity-protective reasoning typically supports *maintaining* doxastic attitudes (that is, belief, disbelief, and suspension) in the face of counter-evidence.[5] In particular, identity-protective reasoning leads to belief perseverance (Anderson et al. 1980; Anderson and Sechler 1986; Slusher and Anderson 1989) and belief polarization (Festinger et al. 1956; Lord et al. 1979; Liberman and Chaiken 1992; McHoskey 1995; Taber and Lodge 2006). These phenomena occur when people receive evidence that contradicts their beliefs and either maintain those beliefs or become even more convinced of them, respectively.

[3]This definition attempts to make unified sense of a range of work on identity-protective reasoning. The existing definitions in the literature are imprecise. For example, even focusing exclusively on Dan Kahan's prominent work on this topic, we encounter incompatible definitions that are not particularly plausible. In Kahan 2017, he claims that individuals are motivated to defend beliefs that are standard in their group, and in Kahan 2015 the motivation is to defend one's status within a group. The former implausibly includes common sense beliefs like "it is a good idea to take an umbrella if it is going to rain," which do not elicit identity-protective reasoning. Similarly, defending one's status within a group includes reasoning aimed at defending one's positive image to others, which is not specifically tied to social identity.

[4]For diverging conceptions of evidence, see for example McWilliams 2021; Simion 2024; Williamson 2002.

[5]For simplicity, I will standardly talk in terms of beliefs.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

Belief perseverance and polarization can come about through many mechanisms. Agents may fail to update due to inattention, reasoning mistakes, limitations of computational power, or reliance on System 1 reasoning. Differences in beliefs between subjects can result from different evidence available in their environment (O'Connor and Weatherall 2019; Pennycook et al. 2022). A subject may fail to update in light of evidence because they do not trust the source of the information (Begby 2024; Levy 2021; Nguyen 2020; Rini 2017). And resistance to evidence can be explained by pre-existing worldviews, values, and cognitive skills (Dorst 2023; Druckman and McGrath 2019; Lepoutre 2020). In particular, in some cases subjects judge that some piece of evidence is likely to be misleading based on their prior beliefs about the world, resulting in belief perseverance without any desire to hang on to a specific view. For example, if a friend tells me that they returned from the Moon yesterday, I may think that they are joking based on the implausibility of the claim, without having any specific interest in defending the claim that they did not return from the Moon.[6]

Identity-protective reasoning differs from these sources of belief perseverance and polarization in that it is a form of *motivated reasoning* (Kunda 1990). In motivated reasoning, subject's non-truth-related desires influence how they interact with evidence, in a way that is not reducible to differential patterns of trust, different background beliefs, plausibility assessments of the evidence received, or any of the factors listed above.[7] In identity-protective reasoning, the relevant motivation is the motivation to defend a social identity. Social identities encompass gender, racial, and other similar identities as well as partisan political identities (e.g., Democrat; anarcho-syndicalist), professional roles, and other group affiliations (e.g., being a Swiftie or a Real Madrid fan) or traits with social significance (e.g., being a runner or a dog person).

The desire to defend an identity can motivate agents to maintain a wide range of beliefs. Most obviously, an identity-protective motivation redounds in defending beliefs about the goodness of the identity and members of the corresponding group (cf. Social Identity Theory – Tajfel 1982; Turner and Oakes 1986; Turner et al. 1987). Additionally, subjects may be motivated to act as press secretaries and defend beliefs that matter for the group's status (Williams 2023a), or that signal loyalty to that group (Kahan 2012; Funkhouser 2022). Such beliefs need not be about the group, identity, or its members. Indeed, one might think that there are no *a priori* constraints on which beliefs can be connected to an identity through cultural mechanisms. Identity-protective reasoning is a mechanism that can be "switched on" to enable selective dogmatism on specific issues by tying them to identity.

This dogmatism is not implemented via lazily employing heuristics or simply refusing to accept evidence. Although emotional factors (attachment to an identity) trigger motivated reasoning processes, the process itself is cognitively sophisticated (Kahan 2012), involving two types of cognitive processes that alter the epistemic situation of the subject.[8]

[6]The potential for such explanations of evidence-resistance causes *the motivated reasoning observational equivalence problem* (Druckman and McGrath 2019) for experimental results attempting to show motivated reasoning. I will stay out of the debate on how to interpret the relevant experiments and restrict my focus to cases that fall under my definition of identity-protective reasoning. That there are cases that fall under its scope is overwhelmingly plausible, as long as one thinks that we are *not* disinterested believers in all cases. Williams (2023a) discusses this debate in more detail.

[7]This does not mean that the subject is *more* evidence-resistant than they would be if one of those mechanisms were in play. The results could be equivalent.

[8]See Flores forthcoming for more discussion of the role of motivation in the revision of beliefs. The processes I will describe do not exhaust identity-protective cognition, which includes what Ellis (manuscript)

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

The first process is one where subjects receive undesired counter-evidence and their motivation leads them to uncover additional evidence. The (implicit) aim is for the subject's total evidence to support their preferred view. Subjects may do this by scrutinizing the counter-evidence received more than they scrutinize favorable evidence (Lord et al. 1979). This enables them to find flaws with the counter-evidence (where such flaws constitute additional evidence) that they do not find with supporting evidence. As a result, their total evidence including these found flaws may end up supporting their preferred view. Alternatively, subjects might search their memory to recall relevant (believed) facts that support their preferred view. Again, they end up with a body of total evidence that allows for the maintenance of their preferred view.[9]

The second process is one in which subjects devise alternative hypotheses that both explain the evidence received and are compatible with their preferred view. For instance, a subject motivated to reject anthropogenic climate change might consider the following hypotheses as explanations for increases in annual average temperature: that this data is the result of cyclical climate oscillations that have nothing to do with human activity, and that the scientific evidence was produced by biased actors. In light of this enlarged hypothesis space, the evidence no longer offers such strong support for revising their view. After all, how strongly evidence supports a claim depends on the hypothesis space considered to account for that evidence (Kelly 2008).

By engaging in these processes, subjects arrive at an overall internal epistemic state that evidentially supports the view they in fact hold. For this reason, agents can be seen as updating rationally in light of the body of evidence and hypothesis space at which they arrive (Kelly 2008), even if they are not updating rationally based on the evidence they receive *taken at face value*.[10]

Nonetheless, identity-protective reasoning is generally considered epistemically impermissible. The reason is that the differential scrutiny and hypothesis generation described appear epistemically impermissible, as they are driven by the truth-irrelevant desire to defend a social identity.[11] Because "one's handling of the evidence [is not] solely responsive to truth-indicating concerns" (Avnur and Scott-Kakures 2015, 23), the responses to *shared* evidence are not those that a disinterested observer (even with the same background beliefs) would have. In fact, these responses conflict with standard evidentialist norms and sometimes directly contradict Bayesian standards (Mandelbaum 2019), in that subjects update in light of the shared evidence in the opposite direction from the ideal Bayesian prescription.

calls "micro-motivated cognition," where subjects' assessments of gradable properties are slightly biased by motivation. Such phenomena are not covered by my discussion in this paper.

[9]Subjects may also engage in inquiring activities that go "outside the head," such as looking up additional information online, seeking out like-minded interlocutors, reading books, etc. Although I do not have space to defend this, I think my arguments in this paper apply to identity-protective inquiry, not just identity-protective reasoning. In any case, in my view, the boundary between the two is blurred (Flores and Woodard 2023; Hughes 2023).

[10]McWilliams (2021) argues that we should not see the subject as updating rationally based on their evidence, because we should include in their evidence "motivated defeaters," defeaters that could be grasped by reflecting further on one's current evidence. This relies on a different conception of evidence than the internalist one with which I am working in this paper.

[11]Dorst (2023) and Westfall (2024) have both argued that differential scrutiny of evidence can be reasonable, insofar as it is a response to one's judgment that counter-evidence is implausible in light of one's background beliefs (Westfall 2024) or to the fact that one expects to have gains in accuracy by scrutinizing counter-evidence (Dorst 2023). But in identity-protective reasoning, differential scrutiny is due to the motivation to defend a social identity – not in response to plausibility judgments or to expected gains in accuracy.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

Because identity-protective reasoning involves sensitivity to non-truth-directed concerns, it is generally unreliable. As Avnur and Scott-Kakures (2015) put it about motivated reasoning in general, "believing according to one's desires is about as reliable as believing randomly . . . so, when desire is directionally influential, this (all else equal) reduces the reliability of the process towards chance" (Avnur and Scott-Kakures 2015, 22–3).[12] What helps us defend our identities is orthogonal to the truth. For this reason, identity-protective reasoning is liable to lead to false beliefs (cf. Kahan 2012; Kahan 2015; Williams 2021; Williams 2023a). Identity-protective reasoning, then, is taken to be epistemically impermissible because it involves sensitivity to non-truth-relevant concerns, which makes it an unreliable way to interact with evidence.[13]

The epistemic irrationality of identity-protective reasoning is taken to pose grave collective dangers. The idea is simple. Democracy requires rationality; identity-protective reasoning is irrational; therefore, identity-protective reasoning threatens democracy (Achen and Bartels 2017; Brennan 2016; Somin 2017). More specifically, identity-protective reasoning is taken to make policies hostage to identity and disconnected from evidence. And it is accused of undermining voters' ability to reliably identify who serves their interests, thereby compromising democratic accountability.

More pointedly, identity-protective reasoning is one mechanism behind group polarization. Indeed, if a population is segmented into social groups with strong identities, each reasoning in identity-protective ways, the result is polarization. Identity-protection thus leads group to divergent, non-negotiable views of reality, preventing them from agreeing on key points or acting together.

3. In favor of identity-protective reasoning, part 1: the division of epistemic labor and epistemic insurance

Two recent strands of work put pressure on the idea that identity-protective reasoning is epistemically noxious at the collective level. The first of these strands focuses on collective deliberation and problem-solving. The other centers on ideas in philosophy of science about the value of exploring diverse views. Drawing on these, I will argue that identity-protective reasoning can help us arrive at true beliefs collectively.

For all I will say in this section, identity-protective reasoning is collectively epistemically good, albeit individually irrational; a form of what some theorists have called "Mandevillian intelligence" (Smart 2018; Peters 2021). In §4, I will challenge the idea that identity-protective reasoning is always epistemically irrational at the individual level.

On to the first strand. Following Mercier and Sperber (2017)'s argument that "myside bias" is beneficial for collective deliberation, Hallsson and Kappel (2020) argue that motivated reasoning (including identity-protective reasoning) can facilitate a helpful division of epistemic labor. The core idea is that by having advocates for opposing views challenge each other, we can more effectively uncover the truth – the principle that underpins the structure of our legal system. The rationale for this procedure is that each

[12]It is common for theorists to acknowledge that there could be cases where motivated reasoning supports true beliefs, or even where it is reliable (Avnur and Scott-Kakures 2015; Carter and McKenna 2020; Ellis 2022). But this is generally left unexplored. §4 will explore that possibility, serving to isolate a range of cases where identity-protective reasoning promotes veritistic aims.

[13]This claim of epistemic irrationality is compatible with the claim that identity-protective reasoning is *expressively rational* (Kahan 2015), in that it carries benefits for subjects (self-esteem, inclusion, and status in a group) that vastly outweigh the costs (having a false belief about a fairly remote issue). Expressive rationality and epistemic irrationality are compatible.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

side finds the best reasons for their position, and truth wins out as the result of critical engagement with one another's reasons.

> Motivated reasoning can play this positive role in facilitating an effective division of epistemic labor because it can increase one's ability to find good reasons in favor of one's view, and to critique reasons against it, compared to more dispassionate reasoning. When both sides of an issue are represented in deliberation, this results in a broad range of reasons being considered for each side, and the selective retention of the best ones. (Hallsson and Kappel 2020, 2824)

As long as participants remain willing to change their minds if the counter-evidence becomes sufficiently strong, this division of labor collectively improves our chances of converging onto the truth.

In support of this view, Hallsson and Kappel (2020) point to two types of experiment on the benefits of deliberation among disagreeing agents. First, subjects do much better on the Wason Selection Task when they determine their final response after collectively deliberating with people who initially disagree on the correct answer, as long as they discuss their reasons for their answer (Trouche et al. 2014).[14] Indeed, groups did well in these settings even if *all* participants had initially selected an incorrect answer (in one experiment, 100% of such groups ended up with the correct answer, whereas only 9% of people do so individually; Moshman and Geil 1998). Second, in "hidden profile" experiments, where a group has to find the correct solution to a problem after deliberating, the group does much better when they start from a position where each member has a body of evidence that supports a different conclusion (Schulz-Hardt et al. 2006).

These experiments suggest that disagreement among group members and members having different bodies of evidence epistemically enhances collective deliberation. Now, identity-protective reasoning can enable agents to possess different bodies of evidence and consider different hypotheses (§2). If a community is made up of agents whose identities motivate them to defend different views, identity-protective reasoning can provide the advantages for collectively arriving at the truth that Hallsson and Kappel (2020) discuss.

Lepoutre (2020)'s defense of dogmatic group cognition generalizes the idea – found in Lakatos – that dogmatism can be fruitful in science. Lakatos's key insight is that the balance of existing evidence is not always a good indicator of truth. Sometimes, scientific theories that lack evidential support later prove superior to their competitors (Lakatos 1970). For example, the view that peptic ulcers are often caused by *H. pylori* bacteria is now well-established. But Barry Marshall and Robin Warren were ridiculed when they first proposed this theory, as the evidence at the time strongly supported the view that bacteria could not live in the acidic environment of the stomach (Thagard 2000).[15]

In light of the potential misleadingness of bodies of evidence, it is advisable – for arriving at the truth – that some agents in the community pursue hypotheses that are not supported by the total evidence. In doing so, agents should hold on to their hypotheses

[14]In the classical version of the Wason selection task, participants are shown a set of four cards placed on a table, each with a number on one side and a color on the other. For example: the visible faces of the cards might show 3, 8, blue, and red. Participants are asked which card(s) they must turn over to test that if a card shows an even number on one face, then its opposite face is blue. The correct response is that they must turn over the red card and the '8' card.

[15]See also Solomon 1992 for discussion of how belief perseverance "on the part of geologists during the geological revolution was, contrary to what might be expected, in fact conducive to scientific success" (Solomon 1992, 443).

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

much as motivated reasoners hold on to their beliefs, treating counter-evidence as an anomaly that indicates problems with research methods or with auxiliary assumptions, *not* with their preferred hypothesis.[16]

Bolstering this point, Zollman (2010) formally modeled situations where some agents pursue minority views against the total evidence, finding that "endowing individuals with dogmatic priors has a good effect when the overall behavior of the community is in focus" (Zollman 2010, 84). As long as agents share the evidence they gather, such an epistemic community is less likely to prematurely discard and more likely to converge onto superior theories than one where all agents believe according to the total evidence at each time.[17]

To the extent that identity-protective reasoning enables dogmatically pursuing theories that are not supported by the total evidence, these results suggest that it can play a positive epistemic role. This conclusion is the same as the one derived from Hallsson and Kappel (2020)'s argument. However, note that neither Hallsson and Kappel (2020) nor Lepoutre (2020) think dogmatism is unconditionally good. Some conditions must be in place for the collective benefits identified to arise.

First, the claims that subjects defend must be ones which evidence bears on, paradigmatically empirical beliefs.[18]

Second, agents cannot be fully dogmatic. They must be willing to change their mind if the counter-evidence to their view is strong enough. Fully dogmatic agents would forever be stuck with false views, instead of reaping the benefits of collective deliberation and the exchange of reasons (Gabriel and O'Connor 2024). Correspondingly, identity-protective reasoning is collectively helpful only if agents abandon their identity-connected beliefs once the counter-evidence is strong enough. Fortunately, there is little reason to think that paradigmatic identity-protection involves total dogmatism. Standardly, identity-protection does not involve wanting to believe that p come what may, with no interest in whether p is true. Instead of a complete disinterest in truth, identity-protection involves finding it more costly to falsely believe that p than to falsely believe not-p (Avnur and Scott-Kakures 2015).

Third, agents who disagree must share evidence and engage in collective deliberation. Otherwise, they will not reap the benefits of different agents collecting different evidence. In particular, for identity-protective reasoning to have positive collective epistemic effects, a society needs to include groups who disagree with one another, and these cannot be completely siloed off or deeply mistrust each other's testimony.

Fourth, the arguments surveyed assume that agents are not just grabbing wildly for any rationalization that suits their preferred views. Such behavior would be unlikely to help us arrive at the truth or at a better body of evidence. Indeed, Lakatos and Zollman both assume that the agents at hand are scientists who hold on to a theory dogmatically but handle the evidence in broadly reasonable ways. In particular, the hypotheses that

[16]Kitcher (1993)'s arguments for the view science progresses best if a good balance of 'orthodox' and 'maverick' strategies are pursued within the scientific community and offer further support to this point.

[17]In a more recent agent-based model, Gabriel and O'Connor (2024) echo this, finding that moderate confirmation bias (where agents are, in the authors' operationalization, more likely to reject information that does not accord with their prior beliefs) improves collective epistemic outcomes. Although this again supports the collective value of dogmatism, there are some differences between confirmation bias as modeled and identity-protective reasoning. First, in identity-protective reasoning, agents do not simply reject information. Second, identity-protective reasoning is specifically about identity-connected beliefs. Third, identity-protective reasoning goes beyond assessing the plausibility of evidence in light of one's prior beliefs (which, arguably, is what confirmation bias involves; Westfall 2024). For other agent-based models that echo this pro-dogmatism conclusion, see Frey and Šešelja 2018; Frey and Šešelja 2020; Xu et al. 2016.

[18]To be clear, moral beliefs are covered if we think that they are such that evidence bears on them.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

agents consider must be reasonable, and the evidence they retrieve in motivated ways must include factive evidence that supports the conclusions it purports to. Fortunately, identity-protective reasoning can be rigorous; as Susanna Siegel notes, in motivated reasoning, "desire could be mediated by epistemically well-founded processes" (Siegel 2017, 414).[19]

In summary, a community where agents engage in identity-protective reasoning on controversial issues will in some conditions be epistemically better off than one where all agents are impartial reasoners. The relevant conditions are the following: there are identity groups that disagree about factual claims; groups share information and take other groups' evidence into account; agents are willing to revise their beliefs if the counter-evidence is strong enough; and agents employ well-founded methods in scrutinizing evidence. In such conditions, identity-protective reasoning facilitates productive collective deliberation instead of wrecking it. Against the concerns about collective rationality and democracy that we saw in §2, identity-protective reasoning can sometimes help us collectively arrive at the accurate judgments that democracy requires.

One might object that the conditions listed above only rarely hold. If that is right, then real-world identity-protective reasoning remains for the most part toxic. In particular, isn't identity-protective reasoning typically accompanied by epistemic bubbles and echo chambers (Nguyen 2020)? If that is the case, then identity groups generally do not access outgroup evidence and, when they do, they reject that evidence because they do not trust the outgroup (Joshi 2024). Indeed, it appears that negative feelings of the sort that break down communication are commonplace when it comes to partisan identities in the US (Iyengar et al. 2019), and that some social groups are systematically marginalized and not heard by more privileged groups (Wu 2023). If all of this is correct, then the arguments given might describe a purely theoretical possibility, leaving unscathed concerns about the collective disvalue of identity-protective reasoning in our world.

Clearly, it is an empirical question of how often the conditions under which identity-protective reasoning supports collective epistemic goods are met. Perhaps worries about identity-protective reasoning in the case of partisanship in the US are well-founded. But this is only one case of identity-protective reasoning. Identity-protective reasoning can encompass any social identity that matters to agents (being a man, a dyke, a philosopher, etc.). Once we move outside of the context of contemporary partisanship in the US, the relevant conditions for collectively beneficial identity-protective reasoning are met for many identity divisions. In particular, many identity groups are not isolated from non-members and do not fully dismiss their testimony. As long as groups do not see outsiders as evil enemies and society includes some public forums, we should expect agents to share evidence and have some mutual trust.

Regardless of this empirical point, the arguments in this section shift the dialectical terrain. These arguments suggest that the collective problem that we should be concerned about is *not* that subjects engage in identity-protective reasoning *per se*. Instead, the problem is structural: subjects do not share evidence across group lines. The fix for this problem is to construct epistemic networks with healthier information flow and address pathologies of social trust. Identity attachments themselves can be used as positive epistemic resources in facilitating a healthy division of epistemic labor and insuring us against the risk of being stuck with views that are evidentially supported at a time but false.

[19]Similarly, Avnur and Scott-Kakures (2015) and Ellis (2022) both note that the steps in reasoning in a particular instance of motivated reasoning could be identical to those used in accuracy-guided reasoning.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

The collective rationality of identity-protective reasoning in good structural conditions gives individuals the freedom to indulge their cognitive tendencies without incurring consequences. In particular, Nguyen (2022) argues that *playfulness* (the extremely undogmatic willingness to explore views that are not one's own) is needed to protect us against being stuck with false views. Interestingly, the arguments in this section suggest that a collective structure with agents who are dogmatic in different directions might provide a suitable insurance against being stuck with false views. We can stay dogmatic and outsource playfulness to the collective level.

4. In favor of identity-protective reasoning, part 2: Epistemic resistance in the face of evidential distortion

One of the key ideas in the last section was that identity-protective reasoning can be collectively beneficial, since the total evidence at a given time can be misleading. In light of this possibility, it is good for some members of the community to dogmatically pursue views that depart from the balance of evidence.

From this argument we can draw a lesson: although identity-protective reasoning can enable the maintenance of *false* beliefs in the face of *genuine* counter-evidence (the focus in the literature on identity-protective reasoning), it can also enable the maintenance of *true* beliefs in the face of *misleading* counter-evidence.

The dialectic here mirrors debates about the permissibility of systematically filtering out evidence from some sources (often framed in terms of whether echo chambers/epistemic bubbles can be good). While the standard view is that such filtering is epistemically pernicious, some, most notably Lackey (2021), have argued that epistemic bubbles can be reliably truth-tracking, as long as they in fact exclude unreliable testifiers.[20]

More to the topic of this paper, Battaly (2018) argues that dogmatism (and closed-mindedness more generally) is a "burdened virtue" (Tessman 2005), a trait that can have a preponderance of epistemic benefits in epistemically hostile environments. When the agent is "surrounded by falsehoods, incompetent sources, and diversions, closed-mindedness about options that conflict with what she knows will minimize the production of bad epistemic effects *for her*" (Battaly 2018, 39). Specifically, it will enable the agent to hold on to the true beliefs that she has while protecting her both from false beliefs and from devoting epistemic resources to misguided projects.[21] Similarly, Westfall (2024) suggests that dogmatism can be a way to escape "the gravitational force" of bad ideology. By scrutinizing "things that are epistemically good to scrutinize" (Westfall 2024, 87), we improve our epistemic position.

The argument I will offer develops these insights to argue that identity-protective reasoning is a form of dogmatism which can systematically support true beliefs. Specifically, I will argue that, under conditions of group-based oppression, identity-protective reasoning based on marginalized identities is often a *well-calibrated* form of dogmatism, with a preponderance of good epistemic effects. This is in contrast with the kinds of cases centrally discussed in the literature on identity-protective reasoning, where identity-protective reasoning does not serve to correct for evidential distortion in the environment and, therefore, tends to retrench false beliefs.

[20]Coady (2024), Nguyen (2021), and Westfall (2024) make similar points.

[21]Does this also enable the subject to hold on to their *knowledge*? Battaly speaks in those terms. This view is controversial, as it is plausible that subjects lose justification when they engage dogmatically. See the large literature on the puzzle of Kripkean dogmatism (Kripke 2011) for more on this point.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

4.1. The social distortion of evidence and evidential oppression

In this subsection, I will argue that, instead of living in a system that supports their self-interested ignorance, marginalized groups are surrounded by misleading evidence against the goodness of their identity and in favor of a worldview that does not reflect their interests. In later sections, I will argue that being in such a situation licenses identity-protection.

Theorists often focus on how identity-protective reasoning is part of a broader system of ignorance and false belief (e.g., white ignorance; Mills 2007). Joshi (2022) argues that the social pressures that sustain identity-based groups of sufficient power often also support epistemic bubbles where the total evidence the group has is biased toward the group's views. Williams (2023b) proposes that our motivations generate *rationalization markets*, social structures that produce psychologically compelling rationalizations for our preferred views. To the extent that these conditions are at play, identity-protective reasoning will compound structural factors that push groups toward false beliefs (Joshi 2024). Thus, it will reduce the likelihood that agents arrive at the truth.

However, these concerns only apply to those who have the social power to enforce external distortions of the evidence – the powerful, not the marginalized. Instead of living in epistemic bubbles, marginalized people are often "outsiders within" (Collins 1986). They are enmeshed in the dominant culture and in possession of much of the same evidence as the dominantly situated, in part because they cannot afford to be ignorant of the dominant worldview (Wu 2023). Further, when a group is marginalized, it is less likely that there is a thriving marketplace of rationalizations that caters to group interests, much as it is less likely that there is a thriving market of *anything* catering to the group's interests.

Instead, marginalized groups often face *evidential oppression*, whereby "social distortion causes the available evidence to disproportionately reflect an oppressive ideology with respect to that group" (Saint-Croix 2025, 403). The *social distortion* of evidence, occurs when "social factors, such as ideology or institutional policy, influenc[ing] the prevalence of evidence in that environment in a way that impels agents to take up a particular doxastic attitude toward [a proposition]" (Saint-Croix 2025, 403).

The mechanisms that socially distort evidence relevant to marginalized groups are manifold. Marginalized groups are often under-represented in knowledge-production and dissemination. Hermeneutical injustice can make it harder to articulate relevant facts (Fricker 2007). The questions asked, the hypotheses considered, and the interpretations favored in inquiry about marginalized groups are often systematically distorted (Hays-Gilpin and Whitley 1998). And those with marginalized identities often face testimonial injustice, silencing, and smothering (Dotson 2011), struggling to contribute evidence to the collective pool. As a result, the available evidence on marginalized social groups or identities often disproportionately reflects false negative views about them and topics of concern to them.

For example, as Saint-Croix discusses, media coverage of crime in the US disproportionately spotlights Black and Hispanic perpetrators (Dixon and Linz 2000) and portrays them as more threatening and less excusable than white perpetrators (Chiricos and Eschholz 2002). Such an evidential environment arguably supports the ideological beliefs that Black and Hispanic men are dangerous. Consider also how, in the 19th century, the consensus among medical experts was that strenuous physical activity, such as long-distance running, was extremely dangerous to women's health (Gregg and Gregg 2017). The total evidence at the time presumably supported this false view. After all, experts agreed on that view and women were not allowed to participate in such

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

sporting events, reducing the counter-evidence available. The history of race science and the study of sex differences provide a plethora of similar examples.[22]

Let us take a step back. I have argued, following Saint-Croix (2025), that oppression is articulated in part via the distorted availability of evidence. In oppressive contexts, some evidence is systematically biased away from the truth. Instead, that evidence supports views that form a part of the ideological system that helps sustain the corresponding social order. To put it more bluntly, there is a bias in the evidence towards view that reflect the interests of the powerful. In particular, marginalized groups often encounter evidential oppression, with evidence systematically biased toward negative views of them.

A very important point emerges from this. Evidential oppression and the social distortion of evidence constitute social-structural analogues of motivated reasoning. As we have seen, individuals engage in motivated reasoning when they reason in ways that are biased toward whatever views fit their goals, arriving at bodies of evidence that support those views. Similarly, oppressive social structures incorporate mechanisms that bias the total evidence toward views that suit the system's maintenance.

With respect to marginalized groups specifically, these mechanisms operate to bias the total evidence toward negative views of those groups. Social distortion of evidence in general, and evidential oppression in particular, "break the presumed connection between truth and indications of truth [evidence] within their domain of influence" (Saint-Croix 2025, 406). The question now is: How should agents interact with evidence in such evidentially toxic contexts?

4.2. Identity-protective reasoning as corrective partiality

O'Connor and Weatherall (2019) show that under conditions where there is a preponderance of misleading evidence, Bayesian updating systematically leads subjects astray. For example, where industries attempt to hide the harmful health or environmental effects of their products and shape research in the relevant areas to produce evidence favorable to their interests, Bayesian updating leads agents to fail to believe that those damaging effects occur. This is not surprising. After all, in such situations, bad actors have put money into severing the connection between evidence and truth. Subjects thus need some alternative way of reasoning.

Ideally, agents would find a way to push in the exact opposite direction to that in which the evidence is systematically distorted, to the exact same degree. This would fully correct for the biases in the total evidence, assuming that subjects considered reasonable explanatory hypotheses and arrived at genuine evidence. Agents who reasoned in this way would be implementing a perfectly well-calibrated compensation mechanism. They would employ motivated evidence-resistance to compensate for the social-structural pressures that bias evidence away from the truth.

Sadly, I do not believe that there are perfectly well-calibrated compensation mechanisms. Having a motivation that is perfectly targeted at only truths that ideology

[22]Regarding these examples, an objector might question whether the total evidence available at a given time actually supports the false view indicated. For example, one might think that the evidence provided by sensationalist media supports false beliefs about the dangerousness of Black and Hispanic men, but that the overall evidence available in our society also includes reputable studies, as well as undercutting defeaters for the evidence provided by sensationalist media. In response, it is plausible that all the evidence at the disposal of ordinary agents supports false beliefs in these cases, even if some more sophisticated body of evidence does not. Indeed, one might think that identity-protective reasoning and inquiry may precisely be put at the service of arriving at such a sophisticated body of evidence, one that is not so easily accessible to ordinary agents.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

obscures and nothing else would amount to miraculous anti-ideological clairvoyance. Even so, from a veritistic perspective, someone motivated to defend a number of true beliefs that ideology obscures – even if they also end up defending some false beliefs – would epistemically benefit, as long as they maintain an appropriate balance of defending true beliefs relative to false ones.[23] Such motivated reasoning would constitute a well-calibrated (though not perfect) compensation mechanism.

In the rest of this subsection, I will argue that identity-protective reasoning around marginalized identities in conditions of evidential oppression – which I will label *identity-protective reasoning from below* – can function precisely as this sort of well-calibrated compensation mechanism.

To see this, consider a subject who has some marginalized social identity. Specifically, this social identity is dominantly seen as bad: as incompetent along some dimension, cold, unpleasant, or untrustworthy (Fiske et al. 2007). How should such a subject reason on topics connected to their identity?

The first thought a well-read epistemologist might have about defending identity-protection in such a context goes through moral encroachment (Basu 2019). According to moral encroachment views, agents are epistemically permitted to resist views that are offensive, diminishing, and part of ideological structures, where agents' acceptance of such views about themselves would compromise their self-esteem and agency. Specifically, the moral costs of accepting such views encroach on epistemic normativity, generating an epistemic permission to carefully scrutinize the evidence for these views, as identity-protective reasoning allows them to.[24]

Although I am friendly to encroachment arguments for the permissibility of some instances of identity-protective reasoning, I am more interested in this paper in giving an argument that focuses on veritistic goals, as the central concerns about this kind of reasoning center there. The rest of my argument will follow a veritistic line, without assuming encroachment.

Under evidential oppression, the total evidence supports negative views about marginalized groups (e.g., that Black men are dangerous, or that long-distance running is harmful to women's health), as we saw in the last section. If the subject is motivated to defend the goodness of their identity, then they will scrutinize evidence for these negative views (and other claims that this identity or group is bad) more than a neutral observer would, consider additional explanations for the evidence, and so on.

In a world where the total evidence suggests that negative false claims about the group or identity are true, such biased cognitive maneuvering is truth-conducive. At least, such identity-protective reasoning is truth-conducive where it meets minimal conditions on epistemic conduct: agents would change their mind if the counter-evidence is strong enough, and the methods they employ are well-founded.

Now, one might be concerned that identity-protective reasoning *also* supports a large number of false beliefs, even under evidential oppression. Maybe some or many negative claims about the group are true. Or perhaps the identity is connected with many views

[23]This balance needs to be weighed by significance: it is more important to get it right on some questions than others. For an excellent treatment of how to weigh significance epistemically, see Munton manuscript.

[24]This is a simplification of the rich and vast literature on encroachment; see Bolinger 2020 for an overview. Further, if one grants the possibility of cases where more than one response to the same evidence is permitted (so-called permissive cases), one may think that there is nothing wrong with opting for the response that best fits one's desires (as Ellis (2022) sketches). This is even more plausible if one thinks that identity motivations often subsume substantive ethical values, which are generally taken to be acceptable settling factors in permissive cases (Lepoutre 2020).

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

that are not in the scope of evidential oppression (i.e., views for which the total evidence is a guide to truth). I will consider these two possibilities in turn.

I acknowledge that sometimes negative claims about groups, identities, and their members are true. Indeed, this is a standard point in theories attuned to ideology and the potential of evidential oppression. Ideology interpellates us into the subject positions it designates (Althusser 1970/2006); looping effects causally shape agents into meeting ideological expectations (Hacking 1995); and stereotype threat leads subjects to live up to negative stereotypes (Steele and Aronson 1995).

That said, evidential oppression still plays a role in shaping the body of evidence supporting true negative views about marginalized groups. As a result, extremely negative views may appear true when only mildly negative views are true. Quantities or proportions of some negative trait might be overstated. And true negative claims might receive overwhelming evidential support and incorrectly appear unquestionable. As such, even with true negative claims, identity-protective reasoning from below can still play a positive role in compensating for ideological distortion.

In addition, one might think that the false beliefs supported will tend to be less important than the false beliefs avoided by identity-protective reasoning from below.[25] For example, perhaps identity-protective reasoning leads a subject belonging to a group stereotyped as dangerous to incorrectly think that a reported crime by a group member did not occur. But it will help this subject have more accurate beliefs about the dangerousness and dispositions to violence of their group, which matters more.

In line with this point, Munton (2019) argues that true generalizations implicitly represent an inaccurate modal profile corresponding to that generalization. Specifically, they represent that the generalization is robustly true across many conditions, where it is only true given oppression. Identity-protective reasoning can ensure that agents do not believe this false modal generalization, protecting them from a false theory of how the social world functions. In particular, identity-protective reasoning can protect agents from believing that negative generalizations about the group derive from a "group essence," perhaps by motivating agents to consider structural explanations for statistical generalizations about groups instead of biological essentialist ones.[26]

In sum, even when some negative claims about a group are true, identity-protective reasoning from below often still plays a positive role from a veritistic perspective. This is because evidential oppression also influences the body of evidence behind those true negative claims, so that identity-protective reasoning still plays a helpful corrective role relative to the biases in the evidence. Second, the true beliefs that identity-protective reasoning enables should arguably be weighed more heavily than more trivial false beliefs in assessing this mode of reasoning.

What if identity-protective reasoning from below also covers empirical matters that are not about the group's goodness (but, e.g., about the health consequences of a diet or the risks of permissive gun regulation)? Is there any reason to think that identity-protective reasoning can be truth-conducive in such a situation?

The answer depends on where those beliefs come from. Sometimes, beliefs end up connected with an identity as the result of a broadly reliable process of selecting which

[25]This appeal to importance is not an appeal to encroachment. I am not claiming that subjects ought to set different thresholds for different propositions. Instead, I rely only on the claim that assessing the veritistic value of a reasoning method should weigh the significance of the true and false claims that the reasoning method supports.

[26]Thanks to Kate Ritchie for discussion.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

beliefs to defend.[27] For instance, in at least some cases, groups arrive at packages of beliefs and these become tied to a group identity via consciousness raising: careful collective reflection based on a rich but neglected body of evidence from the experience of group members (Toole 2023).

We should expect that many of these beliefs are true but that the overall balance of evidence on them is biased. (Indeed, this is one reason why consciousness raising is needed to uncover them!) As such, identity-protective reasoning can still constitute veritistically helpful corrective partiality against the biases in the evidence. More generally, when group beliefs are the result of broadly reliable processes for collecting true beliefs subject to the social distortion of evidence and tying them to an identity, identity-protective reasoning still constitutes a well-calibrated compensation mechanism. This is so even though in such cases it expands beyond beliefs about a group's goodness to include views about the social world writ large.

Identity-protective reasoning can enhance subjects' epistemic position beyond helping them hold on to true beliefs and avoid false ones. As we saw in §2, in identity-protective reasoning, subjects use their capacity for careful reflection to change their epistemic position, modifying their total evidence and hypothesis space. In fact, subjects engage in such mental gymnastics *precisely* so that their preferred belief is well-integrated into their mental life, and does not give rise to the unpleasant feelings of dissonance that accompany tension between beliefs (Aronson 1992; Elliot and Devine 1994; Festinger et al. 1956). As long as subjects' reasoning employs well-founded methods, therefore, identity-protective reasoning both corrects for distortion in the evidence agents receive and enables them to develop an overall more explanatory coherent worldview.

One last important epistemic benefit of identity-protective reasoning from below lies in the downstream protection of subjects' epistemic agency. Some instances of identity-protective reasoning from below are ones in which the subject's sense of basic epistemic competence is put into question. Consider, for example, attacks on the claim that people can genuinely be gender fluid or appeals to biology to argue that trans people are just confused. Being motivated to defend one's trans identity and, therefore, scrutinize such arguments with an eye to rejecting them can be crucial for maintaining epistemic agency downstream.[28]

As Battaly (2018) notes, dogmatism on the part of marginalized groups might help agents "ward off the vice of intellectual servility" (Battaly 2018, 40), by preventing them from losing confidence in their cognitive skills or over-attributing negative traits to themselves. As the literature on gaslighting demonstrates, trusting one's cognitive capacities can be essential to the ability to form beliefs about the world, exercise one's critical capacities, and maintain a cohesive worldview (Abramson 2024; Kirk-Giannini 2022). Hence, identity-protective reasoning from below can provide important downstream epistemic benefits by allowing subjects to continue to exercise their critical capacities.

To summarize, identity-protective reasoning from below has many epistemic benefits. It helps preserve important true beliefs with comparatively low costs in false beliefs; aids subjects in arriving at a cohesive worldview that incorporates these true beliefs; and it protects downstream epistemic agency.

[27]This argumentative move is analogous to Lackey (2021)'s move in defending echo chambers, when she notes that "If you are surrounded by like-minded people through careful discrimination, this is obviously epistemically different than ending up in this situation through chance" (Lackey 2021, 214).

[28]Thanks to M.J. Crockett for discussion.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

Identity-driven dissonance – the unpleasant emotion experienced when one receives counter-evidence to one's cherished beliefs – might therefore be seen as an *outlaw emotion* that helps members of marginalized groups to glom onto the truth (Silva 2021). To take a page from feminist understandings of the epistemic role of emotion, it provides "the first indications that something is wrong with the way alleged facts have been constructed, with accepted understandings of how things are . . . , help[ing] us to realize that what are taken generally to be facts have been constructed in a way that obscures the reality of subordinated people" (Jaggar 1989, 168).

Identity-protective reasoning from below may confer a systematic epistemic advantage to the marginalized (Dror 2023; Harding 2013; Hartsock 1983; Toole 2023; Wylie 2003), one that has gone under-noticed in standpoint theory. Specifically, by providing motivation to scrutinize misleading evidence, a marginalized social identity allows subjects to maintain true beliefs, beliefs that others are misled into rejecting by the social distortion of evidence. This epistemic advantage, in turn, enables agents to collect more evidence against the consensus, contributing to the development of more adequate standpoints.

This is in contrast to what happens in contexts without evidential oppression – like those that members of privileged groups encounter with respect to privileged identities. In such contexts, identity-protective reasoning only serves to further reinforce biases in the evidence. It keeps subjects from calibrating their worldview to what the world is actually like, and it supports the vicious refusal to own up to limitations as opposed to allowing subjects to maintain the basic confidence needed for epistemic agency.

4.3. The epistemic standing of identity-protective reasoning

What does all this mean for the epistemic standing of identity-protective reasoning from below and the beliefs it supports?

I can see two ways of going on the question of whether the true beliefs defended by identity-protective reasoning from below retain justification. Following Srinivasan, one might argue that situations of evidential oppression show that "what intuitively matters most is whether the subject's truth-tracking capacities are distorted by ideological forces, or whether the subject is endowed with capacities that allow her to pierce through ideological distortion" (Srinivasan 2020, 408). According to this externalist view of justification, as long as identity-protective reasoning pierces through ideological distortion, beliefs so defended retain justification. It is irrelevant that differential scrutiny is driven by non-truth-related considerations.

At the same time, the fact that non-truth-related considerations are involved makes it somewhat intuitive to think that subjects' beliefs do not retain justification. Carter and McKenna (2020) argue that, even when motivated reasoning leads to true beliefs, the resulting beliefs are not justified because they are not properly based on reasons. Specifically, agents do not believe on the basis of the reasons that their belief is based on because they are good reasons, but because they want to defend an identity. (So: I might believe that philosophers are very smart based on various facts about philosophers, but I only believe on the basis of these facts because I am motivated to defend my identity as a philosopher.) On this conception of justification, true beliefs supported by identity-protective reasoning are not justified.

But, as Saint-Croix emphasizes, "justification is valuable because it is truth-directed," and the social distortion of evidence "prevents normal justification-conferring practices of gathering and responding to evidence from being truth-directed" (Saint-Croix 2025, 407). In other words, even if these beliefs are not justified, justification in these contexts is not an epistemic good that we should prioritize because it is disconnected from

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

epistemic success. This makes the question of whether these beliefs are justified less important than the question of the epistemic standing of identity-protective reasoning from below as a method, which I now turn to.

At a minimum, identity-protective reasoning from below is *epistemically innocent*: even if irrational, it has significant *epistemic* benefits (listed above) that could not be easily attained otherwise (Bortolotti 2020). To see this, let us consider some alternative ways of reasoning.

First, as we have seen, taking the evidence at face value and updating accordingly would leave agents stuck with false beliefs (and risk compromising their epistemic agency by acquiring unduly negative beliefs about themselves). Second, accepting the evidence while sticking to one's beliefs would incur the cost of incoherence and fragmentation. Third, ignoring or avoiding the evidence (perhaps by seeking out a friendly epistemic bubble) would not enable subjects to engage in generative scrutiny of counter-evidence that can enable more accurate views. Lastly, scrutinizing counter-evidence driven by one's prior beliefs and corresponding plausibility assessments (Westfall 2024), without additional motivation to hold on to those beliefs, would not secure the same veritistic benefits. If the available evidence is sufficiently lopsided, then the unmotivated subject's sense of plausibility is likely to converge with dominant views, so that relying on it will not provide protection from evidential distortion. For these reasons, the benefits of identity-protective reasoning cannot be easily acquired otherwise.

The claim that identity-protective reasoning from below is epistemically innocent still grants that it is a violation of epistemic norms, albeit an epistemically advantageous one. More controversially, I think that identity-protective reasoning from below is in fact epistemically permissible. It is not a violation of the epistemic norms in play in contexts of evidential oppression.

My argument here relies on the idea that the epistemic norms at play in a non-ideal context can be quite different from those at play in ideal contexts. This view finds support in what economists call "the Theory of the Second Best" (Lipsey and Lancaster 1956). According to this theory, when a set of conditions is necessary for attaining a particular state of affairs and one of those conditions is not met, it may no longer be desirable to attain the remaining conditions. For example, if someone wants caffeine, loves cappuccinos but cannot stand the bitterness of espresso, and does not have milk at home, then it is no longer desirable for them to pull a shot of espresso. Grabbing a caffeinated soda will be preferable, although under ideal conditions (with milk at home) they would have pulled the espresso shot. Less trivially, if workers are exploited, going on strike and inflicting costs on service-recipients might be the best option – though, if workers were compensated fairly (the ideal condition) it would be better if they provided the service they are paid to perform.

Drawing on the theory of the second best, DiPaolo (2019) has convincingly argued that the epistemic norms and standards at play in non-ideal contexts may be quite different from those that govern behavior in ideal contexts. We should countenance epistemic norms of compensation, which "help us achieve the least irrational mental state consistent with living in unfavorable epistemic conditions" (DiPaolo 2019, 2053). DiPaolo focuses on compensating for individual fallibility, arguing that we can be required to be incoherent to compensate for our imperfections. My claim is that we may also be epistemically permitted to deviate from the norms of ideal epistemic rationality (such as updating according to the evidence taken at face value) to compensate for shortcomings in our evidential environment.

Accordingly, we can grant that, in an environment where the evidence is a good guide to the truth, agents epistemically ought to adjust their beliefs to the evidence taken at face

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

value. However, under evidential oppression, adjusting one's beliefs to the evidence taken at face value is no longer epistemically desirable. It is better to scrutinize evidence for dominant beliefs in a biased manner.[29] I have argued that identity-protective reasoning from below enables agents to do so in an effective and reasonably reliable way. For this reason, it is permitted under conditions of evidential oppression.[30] As Rawls put it, "two wrongs can make a right in the sense that the best available arrangement may contain a balance of imperfections, an adjustment of compensating injustices" (Rawls 1993, 247). Given the "wrong" of evidential oppression, identity-protective reasoning from below is licensed as an adequate adjustment.

One might object that an agent could do even better if they carefully reflected on the total evidence in an unbiased way, came to the true conclusion that the evidence on topics *X*, *Y*, and *Z* is socially distorted, and scrutinized the evidence in proportion to the degree of distortion that they justifiedly believe to exist. However, reasoning in this way demands a high degree of detachment, cognitive sophistication, as well as time and resources. It might be too much to demand from ordinary agents. In addition, it is unclear whether unbiased reflection on a body of distorted evidence would in every case enable agents to successfully conclude that the evidence is distorted. Social distortion can cover its own tracks well enough that the total evidence is still distorted.[31]

A second worry one might have about sanctioning identity-protective reasoning from below is that this position is ripe for abuse. After all, agents might mistakenly believe that they are in a situation of evidential oppression.

This is a serious risk. Conspiracy theories and other mechanisms of evidential pre-emption (Begby 2021) work precisely by persuading agents that they are in a situation where the evidence against their views is distorted and should not be trusted. Such agents might falsely believe that they are engaging in permissible identity-protective reasoning from below. More generally, given that we can easily be wrong about whether we are under social distortion of evidence, the permission to engage in identity-protective reasoning defended here cannot be used by agents to guide their own epistemic conduct.

In response, it is true that the view I defend can be misapplied. But this does not show that the view is wrong or offers *no* guidance. It only shows that subjects are fallible in determining whether the background conditions for the applicability of the permission are in place – which is arguably in general true anyway (Srinivasan 2015). I take it to be a strength of my position that it yields a rationalizing explanation of the behavior of those who are suspicious of the evidence. Specifically, if those suspicions were correct, they would indeed be epistemically permitted to resist the evidence. In contrast, rejecting my

[29]Note that, if sufficiently averse to black coffee, the mere *risk* of not having milk at home would license buying a coke to take home. Similarly, one might think that the mere risk that one is under evidential oppression is enough to make it permissible to scrutinize evidence for dominant beliefs in a biased manner. This argument requires the permissibility of disvaluing falsely believing dominant claims much more than one disvalues falsely believing their negation, which I will not assume.

[30]This is a *permission*, not an obligation, as there is presumably more than one way to compensate for shortcomings in our environment. In general, the theory of the second-best countenances multiple permitted second-best options.

[31]Employing Sosa (2021)'s helpful taxonomy, we might think that identity-protective reasoning can support animal knowledge but not reflective knowledge or knowledge full well, which requires that the subject exercise their reliable competences guided by the second-order knowledge that these competences are reliable in the context. In this case, subjects would only achieve knowledge full well if they were dogmatic on identity-connected claims guided by the second-order knowledge that such dogmatism is epistemically reliable. I am friendly to the view that subjects would achieve more epistemic points in the latter case than in the former, while still holding that identity-protective reasoning that supports animal knowledge is epistemically permitted.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

view imputes two mistakes to such subjects: first they incorrectly doubt the evidence, then they interact with that evidence in a way that would be norm-violating even if those doubts were accurate.

Note also that this same objection can be leveraged against any account that proposes that whether a particular epistemic structure or procedure is good or bad depends on external features. For example, the objection about difficulties first-personally applying a norm also covers the view that echo chambers are epistemically good when they protect true views from misleading evidence and bad when they do not (Coady 2024; Lackey 2021; Nguyen 2021; Westfall 2024).

In my view, it is overwhelmingly plausible that our epistemic assessment is in fact sensitive to external features in this way. For example, whether you should trust someone depends on their actual reliability. In much the same way, genuine identity-protective reasoning from below is epistemically permissible, but identity-protective reasoning from dominant ideological positions is not. That is because in such a context, given that the surrounding evidence is already biased toward dominant interests, identity-protective reasoning is a much worse option from a veritistic perspective than reasoning according to standard evidentialist norms.

5. Identity-protection, democracy, and elite capture

I want to end by connecting my discussion back to the role of identity-protective reasoning in democratic politics.

In §2, we saw that many theorists have worried that identity-protective reasoning is harmful to democracy because it leads to polarization (Achen and Bartels 2017; Brennan 2016; Somin 2017). But against troves of popular discourse, one might think that the focus on polarization as the source of contemporary democratic pathology is a red herring (Nguyen 2021).

Instead, we should focus on elite capture (Bagg 2024; Táwò 2022): the hijacking of projects and resources by elites (comparatively well-resourced and powerful groups in the context), who steer them towards their narrow interests and aims. I will argue that if the main threat to democracy is elite capture, then the fact that identity-protective reasoning can promote polarization is not necessarily bad. In fact, I will draw on my argument in §4 to argue that identity-protective reasoning can be a resource to protect against elite capture. In other words, if we shift our underlying picture of democracy, our assessment of the political role of identity-protective reasoning also changes.

To start, if elite capture is the major risk, what protecting democracy requires are not interventions that depolarize us into placid conformity with the dominant consensus, as attempts to reduce identity-protective reasoning aim to do. Instead, the focus should be on building countervailing power, so as to "ensure that whichever interests are hegemonic in a given society face stiff opposition from well-resourced and well-organized counter-hegemonic groups" (Bagg 2024, 131).

Political projects that oppose the dominant consensus are not necessarily a risk to democracy, as the moral panic around polarization has it. Instead, genuinely counter-hegemonic groups grounded in the interests of non-elites are required for a well-functioning democracy.

Similarly, under conditions of elite capture, the apparently reasonable middle ground on many social and political issues will often be shaped by elite interests. In slogan form: When evidence is captured, (evidence-)resistance is warranted. We should welcome and encourage *epistemic countervailing power*: groups that develop views against the distorted total evidence, driven by motivations to defend non-elite interests.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

Countervailing groups will need to compensate for the social distortion of evidence in developing alternative views of the social world. By connecting group identities to the rejection of views that reflect elite interests, such groups can engage in epistemically productive evidence-resistance, exploring and developing alternative positions that can move us beyond the elite-driven consensus. In other words, identity-protective reasoning based on countervailing identities has the potential to help pierce through the veil of ideology.

From this elite-capture-centric perspective, §4 presents just one case study of this dynamic of capture and countervailing power. Specifically, the relevant elites there are socially dominant groups along the relevant axis of oppression (gender, race, and so on), and countervailing power is represented by marginalized groups subject to evidential oppression. What I am suggesting now is that the lesson that identity-protective reasoning under social distortion can be advantageous can be generalized to countervailing groups, as long as they are savvy about which views they connect to the group identity.

In sum, when we center elite capture, identity-protective reasoning emerges as a resource that can be harnessed in the construction of countervailing power. This is in stark contrast to what the polarization narrative suggests, which is that identity-protective reasoning is a toxic source of the kind of deep disagreement that, in such a narrative, corrodes democracy.

At the same time, the elite capture angle suggests a different danger of identity-protective reasoning to guard against, one that is not about polarization. This is the risk that identities themselves can be captured (Táwò 2022), stymieing the development of countervailing power and directing dogmatism toward elite interests. Elites can come to control not only the evidence available but also how social identities are construed. If elites come to control the views that are connected to different social identities and the salience that identities acquire, they can channel dogmatism to views that are favorable to their interests.

This does not imply a rejection of the potential value of identity-protective reasoning. It simply means that this positive value is fragile and that identity-protective reasoning can be captured – much like any other process of reasoning or resource. Indeed, as we have seen, ideally rational cognition is also ripe for elite capture via the capture of bodies of evidence.

As Nguyen (2023) puts it, limited beings in hostile epistemic environments are "locked in an unending arms race" (Nguyen 2023, 28) of reasoning methods. The lesson is not that we should eschew identity-protection. Instead, we are well-served by playfully shifting between identity (and other) lenses (Camp and Flores 2024; Nguyen 2022) to avoid being stuck in an elite-captured mode of reasoning. More importantly, we should collectively cultivate a diversity of forms of genuine countervailing power and corresponding identities, providing an insurance policy against the elite capture of any one of these.

6. Conclusion

I have offered two lines of argument for the claim that identity-protective reasoning can be epistemically positive, challenging the received view that it is epistemically vicious and politically dangerous. My arguments imply that we should reject blanket recommendations to reduce identity-protective reasoning, whether by cultivating humility (Carter and McKenna 2020) reducing the salience of social identities in contexts of reasoning or deliberation (Talisse 2019; Klein 2020), or attempting to avoid

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

the connection of empirical beliefs with identities (Kahan 2012; Kahan 2016; Kahan 2017).

Instead, we should see identity-protective reasoning as a resource in contexts of pluralistic disagreement with open communication or where the evidence on identity-connected topics is systematically distorted. In the former context, it can provide an insurance policy against the possibility that the total evidence is misleading; in the latter, it has the power to keep us tethered to reality in the face of systematic distortion of evidence by dominant interests. For this reason, it is an important cognitive tool in the perpetual struggle to build countervailing power in the face of elite capture.

Acknowledgments. Thanks to the audience at the 2024 *Episteme* workshop and to Thi Nguyen, Kate Ritchie, and Ernie Sosa for discussion. Special thanks to Jennifer Lackey and Elise Woodard for comments on past drafts.

References

Abramson K. (2024). *On Gaslighting*. Princeton: Princeton University Press.

Achen C. and Bartels L. (2017). *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton: Princeton University.

Ahmed S. (2020). *The Promise of Happiness*. Durham, NC: Duke University Press.

Althusser L. (1970/2006). 'Ideology and Ideological State Apparatuses (Notes Towards an Investigation).' *The Anthropology of the State: A Reader* **9**(1), 86–98.

Anderson C., Lepper M. and Ross L. (1980). 'Perseverance of Social Theories: The Role of Explanation in the Persistence of Discredited Information.' *Journal of Personality and Social Psychology* **39**(6), 1037.

Anderson C.A. and Sechler E.S. (1986). 'Effects of Explanation and Counterexplanation on the Development and Use of Social Theories.' *Journal of Personality and Social Psychology* **50**(1), 24.

Aronson E. (1992). 'The Return of the Repressed: Dissonance Theory makes a Comeback.' *Psychological Inquiry* **3**(4), 303–311.

Avnur Y. and Scott-Kakures D. (2015). 'How Irrelevant influences Bias Belief.' *Philosophical Perspectives* **29**, 7–39.

Bagg S.E. (2024). *The Dispersion of Power: A Critical Realist Theory of Democracy*. Oxford: Oxford University Press.

Basu R. (2019). 'The Wrongs of Racist Beliefs.' *Philosophical Studies* **176**(9), 2497–2515.

Battaly H. (2018). 'Can Closed-Mindedness be an Intellectual Virtue?' *Royal Institute of Philosophy Supplement* **84**, 23–45.

Begby E. (2021). 'Evidential Preemption.' *Philosophy and Phenomenological Research* **102**(3), 515–530.

Begby E. (2024). 'From Belief Polarization to Echo Chambers: A Rationalizing Account.' *Episteme* **21**(2), 519–539.

Bolinger R.J. (2020). 'Varieties of Moral Encroachment.' *Philosophical Perspectives* **34**(1), 5–26.

Bortolotti L. (2020). *The Epistemic Innocence of Irrational Beliefs*. Oxford: Oxford University Press.

Brennan J. (2016). *Against Democracy*. Princeton: Princeton University Press.

Camp E. and Flores C. (2024). 'Playing with Labels: Identity Terms as Tools for Building Agency.' *The Philosophical Quarterly* **74**(4), 1103–1136.

Carter J.A. and McKenna R. (2020). 'Skepticism Motivated: On the Skeptical Import of Motivated Reasoning.' *Canadian Journal of Philosophy* **50**(6), 702–718.

Chiricos T. and Eschholz S. (2002). 'The Racial and Ethnic Typification of Crime and the Criminal Typification of Race and Ethnicity in Local Television News.' *Journal of Research in Crime and Delinquency* **39**(4), 400–420.

Coady D. (2024). 'Stop Talking About Echo Chambers and Filter Bubbles.' *Educational Theory* **74**(1), 92–107.

Collins P.H. (1986). 'Learning from the Outsider Within: The Sociological Significance of Black Feminist Thought.' *Social Problems* **33**(6), s14–s32.

Conee E. and Feldman R. (2004). *Evidentialism: Essays in Epistemology*. Oxford, England: Oxford University Press.

Elliot A.J. and Devine P.G. (1994). 'On the Motivational Nature of Cognitive Dissonance: Dissonance as Psychological Discomfort.' *Journal of Personality and Social Psychology* **67**(3), 384–394.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

DiPaolo J. (2019). 'Second Best Epistemology: Fallibility and Normativity.' *Philosophical Studies* **176**, 2043–2066.
Dixon T.L. and Linz D. (2000). 'Overrepresentation and Underrepresentation of African Americans and Latinos as Lawbreakers on Television News.' *Journal of Communication* **50**(2), 131–154.
Dorst K. (2023). 'Rational Polarization.' *Philosophical Review* **132**(3), 355–458.
Dotson K. (2011). 'Tracking Epistemic Violence, Tracking Practices of Silencing.' *Hypatia* **26**(2), 236–257.
Dror L. (2023). 'Is There an Epistemic Advantage to Being Oppressed?.' *Noûs* **57**(3), 618–640.
Druckman J.N. and McGrath M.C. (2019). 'The Evidence for Motivated Reasoning in Climate Change Preference Formation.' *Nature Climate Change* **9**(2), 111–119.
Ellis J. (2022). 'Motivated Reasoning and the Ethics of Belief.' *Philosophy Compass* **17**(6), e12828.
Ellis J. (manuscript). *Casting stones: The ethics and epistemology of motivated reasoning*.
Festinger L., Riecken H.W. and Schachter S. (1956). *When Prophecy Fails*. Minneapolis, MN: University of Minnesota Press.
Fiske S.T., Cuddy A.J.C. and Glick P. (2007). 'Universal Dimensions of Social Cognition: Warmth and Competence.' *Trends in Cognitive Sciences* **11**(2), 77–83.
Flores C. (2023). Groupthink is for mindless pawns, but group thinking helps us thrive. *Big Think*. https://bigthink.com/the-well/groupthink-vs-group-thinking/.
Flores C. (forthcoming). 'Resistant Beliefs, Responsive Believers.' *Journal of Philosophy*.
Flores C. and Woodard E. (2023). 'Epistemic Norms on Evidence-Gathering.' *Philosophical Studies* **180**(9), 2547–2571.
Frey D. and Šešelja D. (2018). 'What is the Epistemic Function of Highly Idealized Agent-Based Models of Scientific Inquiry?.' *Philosophy of the Social Sciences* **48**(4), 407–433.
Frey D. and Šešelja D. (2020). 'Robustness and Idealizations in Agent-Based Models of Scientific Interaction.' *The British Journal for the Philosophy of Science* **71**(4), 1411–1437.
Fricker M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
Funkhouser E. (2022). 'A Tribal Mind: Beliefs that Signal Group Identity or Commitment.' *Mind & Language* **37**(3), 444–464.
Gabriel N. and O'Connor C. (2024). 'Can Confirmation Bias Improve Group Learning?.' *Philosophy of Science* **91**(2), 329–350.
Gregg A.E. and Gregg H.V. (2017). 'Women in Sport: Historical Perspectives.' *Clinics in Sports Medicine* **36**(4), 603–610.
Hacking I. (1995). 'The Looping Effects of Human Kinds.' In *Causal Cognition: A Multidisciplinary Debate*, pp. 351–394. Oxford: Clarendon Press/Oxford University Press.
Hallsson G.B. and Kappel K. (2020). 'Disagreement and the Division of Epistemic Labor.' *Synthese* **197**, 2823–2847.
Harding S. (2013). 'Rethinking Standpoint Epistemology: What is "Strong Objectivity"?.' In L. Alcoff and E. Potter (eds), *Feminist Epistemologies*, pp. 49–82. London: Routledge.
Hartsock N.C.M. (1983). *Money, Sex, and Power: Toward a Feminist Historical Materialism*. Boston: Northeastern University Press.
Hays-Gilpin K. and Whitley D.S. (1998). *Reader in Gender Archaeology*. London: Psychology Press.
Hughes N. (2023). 'Epistemic Feedback Loops (or: How Not to Get Evidence).' *Philosophy and Phenomenological Research* **106**(2), 368–393.
Iyengar S., Lelkes Y., Levendusky M., Malhotra N. and Westwood S.J. (2019). 'The Origins and Consequences of Affective Polarization in the United States.' *Annual review of political science* **22**(1), 129–146.
Jaggar A.M. (1989). 'Love and Knowledge: Emotion in Feminist Epistemology.' *Inquiry: An Interdisciplinary Journal of Philosophy* **32**(2), 151–176.
Joshi H. (2022). 'The Epistemic Significance of Social Pressure.' *Canadian Journal of Philosophy* **52**(4), 396–410.
Joshi H. (2024). 'Socially Motivated Belief and Its Epistemic Discontents.' In *Philosophic Exchange*. London: Taylor & Francis.
Kahan D.M. (2012). 'Ideology, Motivated Reasoning, and Cognitive Reflection: An Experimental Study.' *Judgment and Decision Making* **8**, 407–24.
Kahan D.M. (2015). 'The Expressive Rationality of Inaccurate Perceptions.' *Behavioral & Brain Sciences* **40**, 26–28.
Kahan D.M. (2016). 'The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning is and How to Measure it.' In *Emerging Trends in the Social and Behavioral Sciences*. New York: John Wiley & Sons, Inc.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

Kahan D.M. (2017). Misconceptions, misinformation, and the logic of identity-protective cognition. Working Paper 164, Cultural Cognition Project, Yale Law School.
Kahan D.M., Jenkins-Smith H. and Braman D. (2011). 'Cultural Cognition of Scientific Consensus.' *Journal of Risk Research* **14**(2), 147–174.
Kelly T. (2008). 'Disagreement, Dogmatism, and Belief Polarization.' *Journal of Philosophy* **105**(10), 611–633.
Kelly T. (2016). 'Evidence.' In E.N. Zalta (ed), *The Stanford Encyclopedia of Philosophy*. Stanford, CA: Metaphysics Research Lab, Stanford University.
Kirk-Giannini C.D. (2022). 'Dilemmatic Gaslighting.' *Philosophical Studies* **180**(3), 745–772.
Kitcher P. (1993). *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.
Klein E. (2020). *Why We're Polarized*. New York City: Simon & Schuster, Inc.
Kripke S.A. (2011). *Philosophical Troubles: Collected Papers, Volume 1*. Oxford: Oxford University Press.
Kunda Z. (1990). 'The Case for Motivated Reasoning.' *Psychological Bulletin* **108**(3), 480.
Lackey J. (2021). 'Echo Chambers, Fake News, and Social Epistemology.' In S. Bernecker, A.K. Flowerree and T. Grundmann (eds), *The Epistemology of Fake News*. Oxford: Oxford University Press.
Lakatos I. (1970). 'Falsification and the Methodology of Scientific Research Programmes.' In I. Lakatos and A. Musgrave (eds), *Criticism and the Growth of Knowledge*, pp. 91–196. Cambridge: Cambridge University Press.
Lepoutre M. (2020). 'Democratic Group Cognition.' *Philosophy and Public Affairs* **48**(1), 40–78.
Levy N. (2021). *Bad Beliefs: Why They Happen to Good People*. Oxford: Oxford University Press.
Liberman A. and Chaiken S. (1992). 'Defensive Processing of Personally Relevant Health Messages.' *Personality and Social Psychology Bulletin* **18**(6), 669–679.
Lipsey R.G. and Lancaster K. (1956). 'The General Theory of Second Best.' *The Review of Economic Studies* **24**(1), 11–32.
Taber C.S. and Lodge M. (2006). 'Motivated Skepticism in the Evaluation of Political Beliefs.' *American Journal of Political Science* **50**, 755–769.
Lord C., Ross L. and Lepper M. (1979). 'Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence.' *Journal of Personality and Social Psychology* **37**, 2098–2109.
Mandelbaum E. (2019). 'Troubles with Bayesianism: An Introduction to the Psychological Immune System.' *Mind & Language* **34**(2), 141–157.
McHoskey J.W. (1995). 'Case Closed? On the John F. Kennedy Assassination: Biased Assimilation of Evidence and Attitude Polarization.' *Basic & Applied Social Psychology* **17**(3), 395–409.
McWilliams E.C. (2021). 'Evidentialism and Belief Polarization.' *Synthese* **198**(8), 7165–7196.
Mercier H. and Sperber D. (2017). *The Enigma of Reason*. Cambridge: Harvard University Press.
Mills C. (2007). 'White Ignorance.' In S. Sullivan, and N. Tuana (eds), *Race and Epistemologies of Ignorance*, pp. 26–31. Albany, NY: University of New York Press.
Moshman D. and Geil M. (1998). 'Collaborative Reasoning: Evidence for Collective Rationality.' *Thinking & Reasoning* **4**(3), 231–248.
Munton J. (2019). 'Beyond Accuracy: Epistemic Flaws with Statistical Generalizations.' *Philosophical Issues* **29**(1), 228–240.
Munton J. (manuscript). *Priority and Prejudice: The Epistemic Evaluation of Attention and Salience*. Oxford: Oxford University Press.
Nguyen T.C. (2020). 'Echo Chambers and Epistemic Bubbles.' *Episteme* **17**(2), 141–161.
Nguyen T.C. (2021). 'Was it Polarization or Propaganda?.' *Journal of Philosophical Research* **46**, 173–191.
Nguyen T.C. (2022). 'Playfulness Versus Epistemic Traps.' In *Social Virtue Epistemology*, pp. 269–290. London: Routledge.
Nguyen T.C. (2023). 'Hostile Epistemology.' *Social Philosophy Today* **39**, 9–32.
O'Connor C. and Weatherall J.O. (2019). *The Misinformation Age*. London: Yale University Press.
Pennycook G., McPhetres J., Bago B. and Rand D.G. (2022). 'Beliefs About Covid-19 in Canada, the United Kingdom, and the United States: A Novel Test of Political Polarization and Motivated Reasoning.' *Personality and Social Psychology Bulletin* **48**(5), 750–765.
Peters U. (2021). 'Illegitimate Values, Confirmation Bias, and Mandevillian Cognition in Science.' *The British Journal for the Philosophy of Science* **72**, 1061–1081.
Piazza J., Ruby M.B., Loughnan S., Luong M., Kulik J., Watkins H.M. and Seigerman M. (2015). 'Rationalizing Meat Consumption. The 4 Ns.' *Appetite* **91**, 114–128.
Rawls J. (1993). 'The Law of Peoples.' *Critical inquiry* **20**(1), 36–68.

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press

Rini R. (2017). 'Fake News and Partisan Epistemology.' *Kennedy Institute of Ethics Journal* **27**(S2), 43–64.
Saint-Croix C. (2025). 'Evidence in a Non-Ideal World: How Social Distortion Creates Skeptical Potholes.' In H.C. Hänel, and J.M. Müller (eds), *The Routledge Handbook of Non-Ideal Theory*, pp. 401–414. London: Routledge.
Schulz-Hardt S., Brodbeck F.C., Mojzisch A., Kerschreiter R. and Frey D. (2006). 'Group Decision Making in Hidden Profile Situations: Dissent as a Facilitator for Decision Quality.' *Journal of Personality and Social Psychology* **91**(6), 1080.
Siegel S. (2017). 'How is Wishful Seeing Like Wishful Thinking?.' *Philosophy and Phenomenological Research* **95**(2), 408–435.
Silva L. (2021). 'The Epistemic Role of Outlaw Emotions.' *Ergo* **8**, 23.
Simion M. (2024). *Resistance to Evidence*. Cambridge: Cambridge University Press.
Slusher M.P. and Anderson C.A. (1989). 'Belief Perseverance and Self-Defeating Behavior.' In *Self-defeating Behaviors*, pp. 11–40. Boston, MA: Springer.
Smart P.R. (2018). 'Mandevillian Intelligence.' *Synthese* **195**, 4169–4200.
Solomon M. (1992). 'Scientific Rationality and Human Reasoning.' *Philosophy of Science* **59**(3), 439–455.
Somin I. (2017). *Democracy and Political Ignorance*. Redwood City, CA: Stanford University Press.
Sosa E. (2021). *Epistemic Explanations: A Theory of Telic Normativity, and What It Explains*. Oxford: Oxford University Press.
Srinivasan A. (2015). 'Normativity without Cartesian Privilege.' *Philosophical Issues* **25**, 273–299.
Srinivasan A. (2020). 'Radical Externalism.' *Philosophical Review* **129**(3), 395–431.
Steele C.M. and Aronson J. (1995). 'Stereotype Threat and the Intellectual Test Performance of African Americans.' *Journal of Personality and Social Psychology* **69**(5), 797.
Táwò O.O. (2022). *Elite Capture: How the Powerful Took Over Identity Politics (and Everything Else)*. Chicago: Haymarket Books.
Tajfel H. (1982). 'Social Psychology of Intergroup Relations.' *Annual Review of Psychology* **33**(1), 1–39.
Talisse R.B. (2019). *Overdoing Democracy: Why We Must Put Politics in its Place*. Oxford: Oxford University Press.
Tessman L. (2005). *Burdened Virtues: Virtue Ethics for Liberatory Struggles*. Oxford: Oxford University Press.
Thagard P. (2000). *How Scientists Explain Disease*. Princeton: Princeton University Press.
Toole B. (2023). 'Standpoint Epistemology and Epistemic Peerhood: A Defense of Epistemic Privilege.' *Journal of the American Philosophical Association* **10**(3), 409–26
Trouche E., Sander E. and Mercier H. (2014). 'Arguments, More than Confidence, Explain the Good Performance of Reasoning Groups.' *Journal of Experimental Psychology: General* **143**(5), 1958.
Turner J.C. and Oakes P.J. (1986). 'The Significance of the Social Identity Concept for Social Psychology with Reference to Individualism, Interactionism and Social Influence.' *British Journal of Social Psychology* **25**(3), 237–252.
Turner J.C., Hogg A.M., Oakes P.J., Reicher S.D. and Wetherell M.S. (1987). *Rediscovering the Social Group: A Self-Categorization Theory*. Oxford: Basil Blackwell.
Westfall M. (2024). 'Polarization is Epistemically Innocuous.' *Synthese* **204**(3), 87.
Williams D. (2021). 'Socially Adaptive Belief.' *Mind & Language* **36**(3), 333–354.
Williams D. (2023a). 'The Case for Partisan Motivated Reasoning.' *Synthese* **202**(3), 89.
Williams D. (2023b). 'The Marketplace of Rationalizations.' *Economics & Philosophy* **39**(1), 99–123.
Williamson T. (2002). *Knowledge and its Limits*. OxfordUSA: Oxford University Press.
Wu J. (2023). 'Epistemic Advantage on the Margin: A Network Standpoint Epistemology.' *Philosophy and Phenomenological Research* **106**(3), 755–777.
Wylie A. (2003). 'Why Standpoint Matters.' In R. Figueroa and S.G. Harding (eds), *Science and Other Cultures: Issues in Philosophies of Science and Technology*, pp. 26–48. London: Routledge.
Xu B., Liu R. and He Z. (2016). 'Individual Irrationality, Network Structure, and Collective Intelligence: An Agent-based Simulation Approach.' *Complexity* **21**(S1), 44–54.
Zollman K.J.S. (2010). 'The Epistemic Benefit of Transient Diversity.' *Erkenntnis* **72**(1), 17–35.

Cite this article: Flores C. (2025). "Identity-protective reasoning: an epistemic and political defense." *Episteme* **22**, 707–730. https://doi.org/10.1017/epi.2025.17

https://doi.org/10.1017/epi.2025.17 Published online by Cambridge University Press