Naturalizing Logic
A Case Study of the *ad hominem* and Implicit Bias

Madeleine Ransom

Abstract

The fallacies, as traditionally conceived, are wrong ways of reasoning that nevertheless appear attractive to us. Recently, however, Woods [27] has argued that they don't merit such a title, and that what we take to be fallacies are instead largely virtuous forms of reasoning. This reformation of the fallacies forms part of Woods' larger project to naturalize logic. In this paper I will look to his analysis of the argumentum ad hominem as a case study for the prospects of this project. I will argue that the empirical literature on implicit bias presents a difficulty for the reformation of the ad hominem as cognitively virtuous. Cases where implicit bias influences our assessment of the truth or claim or argument are instances of ad hominem reasoning, and these qualify as fallacious on Woods' own definition.

Introduction

How good are we at reasoning? The fallacies, as traditionally conceived, are wrong ways of reasoning that nevertheless appear attractive to most reasoners when they have the chance. They continue to exert their pull even after one has been made aware of the error of one's ways. If we are prone to often committing logical fallacies, then this suggests a negative answer to our question. While we of course get many things right (presumably the 'easy' stuff), there are nevertheless many ways we often go wrong, despite our best efforts.

What I term here 'Woodsian optimism' stands against this rather pessimistic view of human reasoning. Our reasoning is for the most part cognitively virtuous, including patterns of reasoning that closely resemble the fallacies. This optimism arises from Woods' (2013) overall project to provide the framework for a naturalized logic that is both 'empirically sensitive' and 'epistemologically aware' (p.

2).[1] The first desideratum calls for input from the empirical sciences when formulating accounts of human reasoning. The second puts the central focus back onto reasoners and the contexts in which reasoning occurs. Woods' naturalized logic thus calls for scrutiny of the way reasoning takes place on the ground, and places descriptive fidelity before ease of modeling or normative considerations. Woods claims that the 'empirical record' points to the conclusion that most reasoning that takes place is 'third-way reasoning', where this is reasoning that is beholden neither to the assessment standards of deductive validity nor inductive strength (p. 10). Naturalized logic will therefore largely be preoccupied with providing an account of such third way reasoning and proposing appropriate standards of evaluation.

Woodsian optimism arises because, he claims, descriptive fidelity attests to the fact that we have abundant knowledge (p. 86). While we do commit many errors of reasoning, we are nevertheless "right enough about enough of the right things enough of the time to survive and prosper" (p. 88). Thus, the fallacies represent an important test case for Woodsian optimism. If we do commit such fallacies with high occasioned frequency then this puts pressure on an optimistic view of human reasoning.

Accounting for the fallacies also serves another central aim of Woods' overall project. He conceives of his work as setting the stage for a theory of errors of reasoning, answering Hamblin's question of why there is no theory of fallacy despite our having theories of correct reasoning and inference. He contends that the traditional (mis)treatment of the fallacies is central to understanding this lack of progress towards a unified theory. Woods rejects that the traditional list of fallacies, the so called 'gang of eighteen' immortalized in contemporary logic textbooks, are in fact fallacies. This is the *concept-list misalignment* thesis — the traditional list of fallacies are not in the extension of the concept 'fallacy' (p. 6). While he accepts that the traditional characterization of such fallacies qualify as errors of reasoning, actual human reasoners do not commit them with enough frequency to qualify them as fallacies. Instead, Woods argues for the *cognitive-virtue* thesis, which is that the methods of reasoning traditionally labeled as fallacies are instead better understood as largely cognitively virtuous ways of reasoning (p. 7). Woods argues for both claims throughout the book by way of in-depth analysis of each purported fallacy.

In this paper, I will focus on Woods' treatment of one of the gang of eighteen: the *ad hominem*, in the hope of using it as a test case for the prospects of his larger project.[2] In section one I explain why on Woods' account the ad hominem

[1] Hereafter all subsequent page references will be to [27] unless otherwise indicated.

[2] Hereafter I drop the italization of 'ad hominem'.

does not qualify as a fallacy. In section two I raise a preliminary worry for his project: while we may not succumb to the ad hominem in our actual belief-forming processes, our post-hoc justifications may nevertheless be vulnerable to the fallacy. In section three, I consider whether we do in fact succumb to the ad hominem in our actual belief-forming processes. I argue that evidence from implicit bias suggests that we do. Finally, in section four I draw some broader lessons from this case study for the prospects of a naturalized logic.

1 Woods on the Ad Hominem Attack

1.1 What is the argumentum ad hominem?

As commonly conceived of, ad hominem arguments are those in which an argument or claim put forward by an agent is rejected on the basis of a fact about the agent's character, background or circumstances.[3] When such information is irrelevant to the truth or falsity of the argument itself, then the argument is in error. Call this the *malignant* ad hominem. When it turns out that such information is relevant to the truth of the agent's claims, the argument is not fallacious. Call this the *benign* ad hominem. As I use the term below, the benign ad hominem also includes cases where information about character is not being used to assess the truth of arguments or claims at all, but bringing it up in argument instead serves some other end (to be specified further).

Ad hominem arguments can be broken down into further sub-categories:

Circumstantial ad hominems are those in which the agent's argument is rejected on the basis of the circumstances she is in. Specifically, these are circumstances that may bias the agent towards favoring the conclusion of the argument or the claim that she puts forward. For example, a theist that puts forth an argument for the existence of God has a bias towards the conclusion that God exists. However, that an argument originates from a biased source does not make the argument itself false, and so to dismiss the argument on this basis is considered an error of reasoning.

Tu quoque ad hominems are those in which the agent's argument is rejected on the basis of some pragmatic inconsistency she has committed. She has argued against something that she herself implicitly endorses, or vice versa. For example, suppose a smoker argues or claims that you ought not to smoke cigarettes. The error arises if you use such an inconsistency to judge the argument or claim false

[3] In fact, the history of the ad hominem is more complicated than I am letting on here, and it was not properly thought of as a fallacy until relatively recently (see [6]). Nevertheless my intention here is to illustrate the currently received view, which is the view Woods focuses on.

— the smoker's argument ought to be assessed on its own merits, regardless of the recreational habits of its originator.

Abusive ad hominems are those in which the reasoner moves from the premise that the agent is of unsound character to the conclusion that the argument or claim she has put forward is likely to be, or is, false. For example, if a reasoner dismisses a philandering politician's arguments for gun reform on the basis that she is a cheat, then this is said to be an error because the argument ought to be assessed independently of the politician's character.

1.2 Why the ad hominem is not a fallacy

In each case, Woods grants that if we were to reason in the indicated patterns, then this would count as an error of reasoning. However, he argues that this is not in fact what we are up to. In order for the ad hominem to qualify as a fallacy, it must fulfill two conditions. First, it must be the case that the ad hominem is of the malignant rather than the benign variety. Only the malignant version is an error of reasoning, and fallacies must be errors of reasoning. For the ad hominem to be malignant, the reasoner must move in some fairly strict way from the premise that an individual is of a certain character, circumstance or background to the conclusion that the individual's argument or claim is likely to be false. And it must also be the case that the information about the individual's character or background is irrelevant to assessing the truth or falsity of her claims. Second, it must be an error of reasoning that is committed with high occasioned frequency — when we have the opportunity to reason along these lines, we take it.

Woods' strategy for denying that the ad hominem is a fallacy involves denying that both of these conditions are met. Many ad hominem attacks are not properly understood as part of argument evaluation at all, and so are benign. Rather, they are rhetorical devices used to embarrass opponents and put them off their game. Pointing out a politician's marital indiscretions changes the subject, perhaps distracting her from the point she was trying to make, or perhaps causing her to continue in a debate with less certainty. This sort of tactic is what Woods labels 'slanging'. While it is used with some frequency in public discourse, it does not qualify as fallacious because there is no suggested link between the slang and the veracity of the person's argument or claims. It does not qualify as premise-conclusion reasoning.

Moreover, those ad hominem attacks that are part of argument evaluation are largely virtuous cognitive strategies. In the circumstantial ad hominem case, consider the example of Leila, who has strong religious convictions and is opposed to stem cell testing. In discussing stem cell testing with Sarah, Sarah points out that Leila's religious commitments forbid it. If Sarah were to reason from this to

the conclusion that Leila's argument against stem cell testing is false, then this would qualify as erroneous reasoning. However, Woods contends that this is not what is going on in such situations. Rather, Sarah's pointing out Leila's religious convictions may serve to challenge Leila's openness to sincerely considering the merits of any arguments to the contrary (p. 451-3).

Woods points out that it is rational for Sarah to suspect that Leila's argument is defective in some way, based on her religious convictions, because such biases often cause us to improperly discount evidence and arguments that go against our beliefs and assign inordinate weight to arguments and evidence that confirm our beliefs. So, based on what we know about confirmation bias, there is a good chance that Leila's argument or position is defective in the sense that she hasn't properly considered arguments or evidence contrary to her position. Ad hominem attacks in this context can also therefore serve to put one's interlocutor on notice — they had better bring their A-game when making an argument for their position because there is reasonable suspicion that the argument is defective in the sense outlined here.

Woods provides three potential benign motivations for someone who issues a tu quoque ad hominem. Consider the case of Harry who asserts to Sarah that people ought not smoke. If Sarah responds by pointing out that Harry himself is a smoker, is she guilty of committing a fallacy? Woods thinks there is a better analysis of her motivation for doing so (p.453-5). First, Sarah may be expressing a doubt about Harry's actual position. While Harry may say that people ought not smoke, perhaps he is speaking loosely, and his actual position is more nuanced — perhaps it is rather that people shouldn't smoke heavily. Sarah's ad hominem remark in this case is a method of prompting Harry to clarify his position.

Second, Sarah may be putting in doubt whether Harry's supporting reasons for his position are sufficient to motivate a rational person to adopt the claim. After all, Harry is presumably a rational person, and he has not heeded his own advice. So Sarah's ad hominem remark serves to prompt Harry to say more about his reasons.

Third, Sarah may be questioning Harry's sincerity as an informant. If he himself doesn't act in accordance with his professed belief, perhaps he does not really believe it. In this case, Sarah's ad hominem serves to prompt Harry to reassure her that he is being genuine.

The abusive ad hominem can be understood in many cases as not involving argument assessment at all, but rather as promoting a desire to punish the individual in question. To illustrate this consider political attack ads, which are paradigmatic cases of abusive ad hominem attacks, as they often portray the candidate in a negative light by dredging up past personal indiscretions or perceived

defects in character. Such attack ads have been extremely effective in swaying the popular vote, and can break a political career (see [25]). The general tendency has been to view the decision to withhold one's vote as at least tacit endorsement of a fallacious line of reasoning — that a given candidate's moral indiscretions cast doubt upon the veracity of her claims or her ability to govern responsibly, despite her proven track record. However, Woods instead provides an alternate diagnosis of how attack ads and the like affect our decisions. For Woods "the object of the [negative "character" campaign] is to strengthen the causal tie between personal antipathy and a decision to withhold one's vote" (p. 444). By smearing the candidate's character, the attack ads cause viewers to dislike the candidate. This dislike, in turn, causes viewers to desire to punish the candidate, and the most expedient way to do so is to withhold one's vote. The alternate hypothesis Woods proposes is thus that attacks on character serve to influence the affect one feels towards the target, and such negative affect can cause us to desire to punish said target. Importantly, while voters who engage in this practice are committing a punitive ad hominem, it is benign — no fallacy is committed because they do not reason from the premise that the candidate's character is flawed to the conclusion that her arguments or claims are likely false.

Woods also offers an error theory as to why such voter decisions are commonly thought to be commissions of the malignant ad hominem when they are really instances of a benign ad hominem. It is because when challenged to justify one's decision to withhold one's vote, voters may mischaracterize their decisions as probative rather than punitive. That is, they may endorse a malignant ad hominem line of reasoning to justify their decision, post hoc: "The misbehaving candidate is not fit for office because he is unreliable. If he'll betray his wife, who's to say that he wouldn't betray his country? He is a person of bad character in marital matters. So who is to say that he wouldn't be a person of bad character in national security matters?" (p. 444).

In summary, the ad hominem is not a fallacy because the sort of practice we do engage in does not involve moving from the premise pertaining to an agent's character, circumstance or background to the conclusion that their argument or claim is likely false. In the case of the abusive ad hominen, this is because it does not involve premise-conclusion reasoning at all, and in the case of the tu quoque and circumstantial ad hominems, the reasoner is better described as rationally doubting that the argument is as strong as it is made out to be. Ad homineming, in such circumstances, is actually virtuous insofar as it invites one's interlocutor to clarify their position, provide stronger arguments, potentially be more receptive to arguments to the contrary, and overall keeps the conversation going, so that progress on this issue can be made (p. 466).

2 A preliminary worry: belief formation versus justification

The error theory that Woods advances for the abusive ad hominem is that while we do not reason malignantly in forming our decision to withhold our vote from the despised candidate, we nevertheless tend to invoke such malignant ad hominem arguments in a post hoc justification process. If this is an attractive form of reasoning, and common enough, then it seems that ad hominem is a fallacy after all. No matter that it is not the way we actually form our decisions.

In order for Woods' argument to go through, it must be the case that such malignant ad hominem reasoning is committed with low occasioned frequency. But if this is how voters commonly justify such decisions made with punitive intentions, then by Woods' own lights it seems that while they are not committing an error of reasoning during voting, they are committing something that looks dangerously like a fallacy during the post hoc justification process. That voters may seldom be asked for such justification is of little help here — the error theory proposes that when voters are in fact asked, they are likely to concoct a story that invokes the malignant ad hominem. And this is uncomfortably similar to saying that the malignant ad hominem is an attractive or compelling line of reasoning to voters seeking to explain their actions to others or to rationally reconstruct their own thought processes. Attractive but false — the hallmark of a fallacy. However here we might note that Woods' error theory is at this point mere speculation. We do not know how often actual voters, when called upon to justify their voting decisions, put forth such reasoning. So the question of whether the ad hominem is a fallacy or not is beholden to the results of such empirical investigation.

Nevertheless, the discussion here highlights a tension in Woods' account. He wants to allow that we have abundant knowledge, and so adopts a reliabilist account whereby our cognitive faculties can produce knowledge without needing to be able to access internal reasons for belief (ch. 3). Woods advances proposition 3.5a, '*a causal response description of knowing*', whereby "a subject knows that α provided that α is true, he believes that α, his belief was produced by belief-forming devices in good working order and functioning herein the way they are meant to, operating on good information and in the absence of environmental distraction or interference" (p. 93). We can know something without being able to justify it. Given this account of knowledge, Woods is largely concerned with demonstrating that we do not fall prey to the reasoning patterns traditionally labeled fallacies when we form our beliefs. However, we are also often called upon in discourse to justify our beliefs. This activity equally deserves the title of reasoning — we spend a lot of time trying to explain ourselves to others. And

if we are susceptible to the fallacies here then this puts Woodsian optimism in danger. Perhaps we are only half virtuous, half vicious.

Putting this issue aside, our virtue can also be called into question in a way that is more concerning for Woods' project. Our belief-forming processes themselves may be susceptible to malignant, or at least non-benign, patterns of reasoning. To assess whether we use ad hominem reasoning when forming our beliefs, I will turn to a research program in psychology dedicated to examining how information about categories of individuals is brought to bear on our beliefs: implicit bias.

3 Implicit bias and the ad hominem

Implicit biases may be thought of as unconscious tendencies to associate a given social group with a certain way of behaving or certain character traits. As Holroyd [7, p. 275] describes the phenomenon, '[an] individual harbors an implicit bias against some stigmatized group (G), when she has automatic cognitive or affective associations between (her concept of) G and some negative property (P) or stereotypic trait (T), which are accessible and can be operative in influencing judgment and behavior without the conscious awareness of the agent'[4] (though see [8] for discussion of heterogeneity of implicit bias).

For example, CV studies involve providing different groups of subjects with identical resumes except for the name, which varies along social category lines. One group will receive a CV with a female name, and the other will receive the identical CV with a male name (the studies have also been done with African American vs. Caucasian sounding names, amongst other variants). Subjects are then tasked with rating the quality of job applicants and making suggestions of 'hireability'. Several such studies have found that the CVs of men are consistently rated more highly than those of women, despite their being otherwise identical. For example, in one study exploring bias in the sciences, physics, chemistry and biology professors were asked to assess the application materials of either a (fictional) male or female student applying for a lab manager position ([13], see also [22]). The male applicant was rated as more competent and 'hireable', and faculty recommended a higher starting salary, despite the application materials being identical.

Implicit biases are typically thought to operate below the level of awareness — the automatic associations that influence our belief-formation processes are not directly open to our scrutiny and control. They thus fall squarely in the do-

[4]Note that implicit biases can also be 'positive'. For example, Asians may be associated with exceptional math skills. While this bias may be harmful for the social group in some contexts, it can nevertheless work in the social group's favour in other contexts.

main of reasoning that Woods is concerned with. Their influence is on how we actually arrive at our beliefs, not on the post-hoc justifications we offer for such beliefs. Implicit biases are also candidates for ad hominem reasoning, because as I shall elaborate below, information about a person's social category can influence argument assessment. While not all instances of implicit bias qualify as ad hominem reasoning, and not all ad hominem reasoning will involve implicit bias, implicit bias is nevertheless an important source of ad hominem reasoning. Our most general analysis of the ad hominem attack is one where an argument or claim put forward by an agent is rejected on the basis of a fact about the agent's character, background or circumstances. It is not a stretch to add the apparent social category of the agent to this list.[5] If this is right, then several questions must be answered about implicit bias in order to assess whether it is benign or malignant ad hominem reasoning and whether it may qualify as a fallacy.

3.1 Are implicit biases involved in argument assessment?

First we need to know whether we bring to bear our implicit biases when assessing the veracity of people's claims or arguments. This would establish that information about the social category of the agent is involved in argument or claim assessment, and so would make the phenomenon a proper candidate for malignant ad hominem reasoning. While much of the research focuses on evaluative judgments such as assessments of a person's competence or abilities, there is some evidence that implicit bias plays a part in argument and claim assessment. Take a widely discussed study that looked at peer review in psychology journals, prior to the days of widespread blind review practices. Peters and Ceci [17] selected 12 papers from 12 top 'high impact' psychology journals with different (though in several cases overlapping) specializations that had been published in the last 18 to 32 months, where at least one author of the study was affiliated with a prestigious institution, such as Harvard, Stanford, and UC Berkeley. They then changed the author names (though not gender), and affiliations to fictitious institutions such as the 'Tri-Valley Center for Human Potential'. These altered papers were sent out to the same journals that had originally published them for review. Only 3 papers were detected as having been submitted and published under a different author. Of the remaining 9, 8 were rejected. The reviewers for at least 5 of these journals cited serious methodological errors as reasons for rejecting the studies. For example, "A serious problem is the range of difficulty of ... material within groups. No account of this range is given and no control of its possible effects is of-

[5] Cf. [9], who argues that implicit bias should be understood in terms of status quo bias, and so commits the *ad verecundiam* fallacy, or appeal to authority. Though I suspect that there will be cases of both sorts of fallacy, as implicit bias is not a wholly unitary phenomenon [8].

fered. Similarly the comparability of material across groups is unknown" (p.190). Or "It is not clear what the results of this study demonstrate, partly because the method and procedures are not described in adequate detail, but mainly because of several methological defects in the design of the study" (p. 190).

Peters and Ceci argue that the best explanation for this result is prestige bias. Perhaps the lack of authorial prestige in the altered papers caused reviewers to misjudge the paper as having serious methodological flaws when they did not in fact have them. Or, perhaps the prestige of the authors positively contributed to the original papers' acceptance, in that it caused reviewers to overlook what were in fact serious methodological flaws. Because the institutional affiliation of the altered papers is not prestigious, this allowed reviewers to catch what might have been overlooked. Either way, prestige influences the assessment of the quality of the arguments or claims made by the authors.

There is also research that implicit bias influences the credibility of witnesses during trials, which can in turn lead to their claims being dismissed as false.[6] For example, Nagle *et al.* [14] found that male witnesses in real trials were consistently rated as more trustworthy than female witnesses by independent raters who watched their testimony on videotape. While more research is needed to understand the link between argument assessment and implicit bias in such cases, there is nevertheless some reason to believe that implicit bias is involved in assessing the truth of claims and arguments.

3.2 Are implicit biases relevant to argument assessment?

The second question relevant to determining whether implicit bias counts as a malignant reasoning is whether our use of such categorical information in assessing the claims and arguments of others is warranted or not. If it turns out that such information is relevant to argument assessment, then such ad hominem reasoning would not be malignant but benign. The answer to this question is less clear, in part because there may be cases of implicit biases where the information is in fact relevant to belief formation.

In many instances, implicit bias is a cognitively virtuous strategy (see for example [1, 24]). Categorization, part of the basis for implicit bias, is a valuable feature of human cognition. Sorting the individuals we encounter into categories

[6]Implicit bias is thus closely associated with testimonial injustice [4], where the social identity of a speaker can cause listeners to judge the speaker's testimony less credible than they otherwise would have had the speaker possessed a more favorable social identity. However, Saul [20] takes it that implicit bias is different from this phenomenon; while testimonial injustice involves only misjudging someone's credibility as a source of knowledge, "the research on implicit bias shows us that we are actually being affected by biases about social groups *when we think we are evaluating evidence or methodology*" (p. 248).

helps cut down on informational complexity — instead of gathering information on each individual one encounters, generalizations relative to a category of individuals allows us to access this information for new individuals of the same category. This in turn allows us to associate whole categories of individuals with other concepts that can facilitate timely and advantageous action — categorizing an individual as a member of the category 'lion', say, and associating this category with danger, can allow us to prepare to flee or protect ourselves, even if a particular lion we encounter does not attack us. Or, if you are in a given society where women do not characteristically hold leadership positions, then believing that the woman sitting at the head of the boardroom table is not the boss (perhaps the boss's secretary?) may be perfectly reasonable, given the base rate of women leaders [28]. However, this kind of case does not involve argument assessment and so is outside of the scope of our interest.

In the CV case discussed earlier, the quality of a person's achievements may be downgraded if they are a woman, or a person with an ethnic-sounding name, but this doesn't count as an error of reasoning in the sense that interests us here — the candidate's claims in the CV are not judged to be false, and no argument is being evaluated. While this may count as an error of reasoning, and a systematic one at that, it is not the error of reasoning we are looking for, the ad hominem. For this, we require both the information to be irrelevant and for that information to be applied to the assessment of a person's argument or claim.

In the prestige bias case, one might attempt to deny that the information is irrelevant. If we take it that institutional affiliation is a good proxy for the quality of a study, then the information may be relevant, and so reasonable to use it as a heuristic in assessing the studies. Perhaps schools like Harvard, Stanford and the like really do put out better papers on average, because they attract better researchers. However, this is a doubtful line of argumentation, especially in today's difficult job market where less prestigious schools are able to hire well-qualified candidates with stellar research track records. Perhaps prestigious institutions are able to provide lighter teaching loads or better financial support for researchers that in turn ensure higher quality output. If so, the correlation between the quality of a paper submitted and the prestige of the institution is still likely to be very weak — there is good quality research being done at all sorts of institutions.[7] So, if prestige does indeed influence our assessment of the quality of a person's arguments or the veracity of their claims, and such influence is irrelevant to this assessment, then it seems that this is a cognitively vicious, rather than virtuous

[7] An additional problem here is how to measure paper quality. Suppose we decided to look at the number of citations as a proxy for quality. If it is the case that papers from prestigious institutions get cited more often because they are published in higher ranked journals (which is due in turn to prestige bias), then this would provide us with a circular measure of quality.

pattern of reasoning.

The case of witness credibility is complicated by the fact that a person's honesty is relevant to whether or not we ought to discount their claims as potential lies. There are many social cues we use to determine whether an individual is likely telling the truth, such as eye contact and body language. However, it is not generally legitimate to use social category membership as a cue because there is no correlation between social group membership and honesty. While there may be some special cases where social group is relevant because members of this category have a known history of lying to protect each other (such as with police officers), this is the exception rather than the rule. So social category is irrelevant to assessing the veracity of their claims.

However, even here we need a better understanding of the mechanism by which the witness's testimony is discounted. Does a lower credibility rating actually cause us to judge their testimony false, or is it taken as true but simply given less weight when considering all the evidence? This matters because only the first option represents an error of reasoning of the ad hominem variety. While the second may also be an error, it is of a different sort, along the lines perhaps of confirmation bias. Take for example the case of Dr. Christine Blasey-Ford's testimony against now Supreme Court Justice Brett Kavanaugh. She testified in great detail how he had sexually assaulted her. Afterwards, several Republican senators said that, while they found her to be a credible witness, they nevertheless supported Kavanaugh due to lack of corroborating evidence for her accusations [26].

In summary, there is some reason to think that at least in most cases of implicit bias involving argument or claim assessment, social category information is indeed irrelevant to such assessments. As such, it qualifies as malignant ad hominem reasoning.

3.3 How pervasive is implicit bias?

Finally, to qualify as a fallacy the malignant ad hominem would have to be committed with high occasioned frequency. If it is a way of reasoning that we seldom indulge in then Woods' strategy would also be applicable here — since we don't often suffer from implicit bias then it doesn't merit the label of fallacy. On this count, there is considerable empirical evidence that most everyone in western society likely suffers from implicit biases involving social categories. The implicit association test is one such measure of implicit bias (Greenwald et al. 1998). Subjects are tasked with sorting either pictures or words into categories, where these appear with either stereotype-consistent or stereotype-inconsistent pairings. For example, a photo of an African American or a European American may be paired

with the word 'good' or 'bad'. Subjects must then sort the photo into the relevant social group. Over 70% of Caucasian and approximately 40% of African American subjects have been found to be slower at categorizing stereotype-inconsistent photo-word pairings (e.g. African American paired with good) than stereotype consistent photo-word pairings. This speed difference is taken as evidence for the presence of implicit associations ([16, 15, 2, 3]; though see [19, 11] for criticism). Such results also suggest biases are not limited to dominant social groups, but extend to marginalized groups as well. For example, in the CV studies, women were just as likely as men to judge the CVs of men as superior ([22, 13]). Implicit bias seems to affect people regardless of gender, age, ethnicity or political orientation (for a review see [15]).

Jennifer Saul [20] takes us to be at risk of forming beliefs on the basis of illegitimate implicit bias "whenever we consider a claim, an argument, a suggestion, a question, etc. from a person whose apparent social group we're in a position to recognize" (p.251). Since we are often in such situations, there is reason to think that not only does implicit bias exert an influence on our beliefs with high occasioned frequency, but also that many of our actual beliefs will be biased.

In summary, on all three counts implicit bias looks like it fits the definition of a fallacy. Implicit bias does involve bringing to bear information about a person's social identity in assessing the arguments and claims they make, this information is at least in many instances irrelevant (though perhaps not always so, and so one might resist categorizing it as a fallacy on this basis), and the phenomenon is widespread.

4 Some lessons for a naturalized logic

The discussion in section two suggests that a naturalized logic of reasoning may need to be bifurcated. We ought to separately analyze our actual belief forming processes and the sort of justification that we engage in during argumentation and discussion, where these two activities are beholden to different standards and produce different sorts of errors with different occasioned frequencies. But we might wonder whether a further (though related) distinction is in order here, between automatic belief-formation and deliberate belief formation, where the latter engages our conscious reasoning processes. For even if Woods is right that most belief-formation does not involve internalized argument, or 'case-making' (p. 99), it is nevertheless sometimes the case that we do form our beliefs this way. And there is evidence that we do make systematically different mistakes when engaging in these different kinds of reasoning processes (e.g. [10]). So a naturalized logic should proceed with caution in generalizing the lessons or standards from one

kind of reasoning to cover the others — human reasoning is not unitary and so we should not expect a naturalized logic to be either.

The discussion of implicit bias suggests that a naturalized logic of our belief formation processes should be particularly sensitive to the distinction between a truth conducive process and one that is conducive to evolutionarily advantageous beliefs. While these two are not mutually exclusive — under certain conditions it is evolutionarily advantageous to have true beliefs [23] — they can nevertheless come apart. In the case of implicit bias, there is some worry that in many instances the two do come apart. For example, we may form beliefs about whether or not abortion is morally justified based on the social category of the person advancing the argument, rather than on the quality of their argument. Having the same beliefs as someone from a dominant social category, regardless of whether these beliefs are true, may be advantageous for several reasons. Sharing beliefs makes two people more alike, and so more likely to cooperate. There is more potential that one may come to join this dominant social group, and less chance for disagreement and argument, along with social repercussions.

Naturalized logic should therefore be highly sensitive to context — we may expect our belief-formation processes to be largely reliable when it comes to perceptual beliefs, but not when it comes to value-laden beliefs regarding social norms, for example (though, if this is the case then the prospects for Woodsian optimism don't look good). Distinguishing between different contexts can also allow us to get a grasp on errors of reasoning — our belief-forming processes will be distorted in systematic ways in certain contexts. However, implicit bias suggests that distinguishing context is no easy process. Given the pervasiveness of our reliance on testimony and the prevalence of the availability of information about social category, implicit bias may occur in all sorts of contexts, and identifying conditions under which one's beliefs are the product of bias may be practically impossible [20].

These factors complicate the task of answering what Woods considers the most important question for a naturalized logic: "What is right about reasoning that is neither truth-preserving nor probabilistically clinching?" (p. 67).

Ought we count the systematic distortions in our belief-formation processes as errors of reasoning, by the standards of third way reasoning? Or should we — paying attention to the motivations of actual reasoners on the ground, as Woods has asked us to do — instead view third way reasoning as having a different aim than truth, and so not count these as errors of reasoning at all?[8] More practically, if we cannot clearly specify the contexts in which distortions in our belief-formation processes are likely to occur, then it seems we cannot clearly

[8]On the subject of the aim(s) of reasoning, see also [12, 21].

specify the cases in which right and wrong reasoning occurs. So the task before a naturalized logic is a difficult one, though perhaps this is to be expected given the complexity and heterogeneity of human cognition. An account of reasoning that pays attention to human reasoners will invariably inherit this complexity.

5 Conclusion

In this paper I have argued that implicit bias poses a challenge both to the concept-list misalignment thesis and to the cognitive virtue thesis. Implicit bias may be understood as a fallacy insofar as it is a widespread phenomenon that involves bringing to bear irrelevant information about a person's social identity in assessing their arguments and claims.

However, it should be noted that the success of Woods' project does not depend on the correctness of his analysis of the ad hominem. As Woods notes, he would be content to show that at least in most cases the 'gang of eighteen' are not really fallacies, as this would 'crater' the received view of the fallacies, inflicting a 'mortal wound' on the view (p. 8). So if it turns out that we still commit the ad hominem fallacy, then perhaps this would not prevent his cratering the traditional view. There are, after all, seventeen other purported non-fallacies to catapult at the opposition.

Nevertheless, the case study here raises a challenge to Woods' project, and in particular to his commitment to our cognitive virtue as reasoners. It suggests a rather more pessimistic view of human reasoning whereby our belief-formation processes are systematically distorted in ways that, while advantageous, are not truth conducive. This complicates the task of explaining what right reasoning amounts to, in naturalized logic.[9]

References

[1] Antony, L. (2016). "Bias: Friend or foe? Reflections on saulish skepticism." in Brownstein & Saul (eds.) *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*. Oxford University Press.

[2] Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black Americans' implicit racial associations and their implications for intergroup judgment. *Social Cognition*, 21(1), 61-87.

[3] Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research*, 17(2), 143-169.

[9]Thanks to John Woods and the audience of the 2014 WCPA for much helpful discussion of an earlier draft of this paper.

[4] Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

[5] Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6), 1464.

[6] Hitchcock, D. (2007). Is there an argumentum ad hominem fallacy? In H. Hansen & R. Pinto (Eds.),*Reason Reclaimed* (pp. 187—199). Newport News: Vale Press.

[7] Holroyd, J. (2012). Responsibility for implicit bias. *Journal of Social Philosophy*, 43(3), 274-306.

[8] Holroyd, J., & Sweetman, J. (2016). "The heterogeneity of implicit bias" in Brownstein & Saul (eds.) *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*. Oxford University Press.

[9] Hundleby, C., (2016). "The Status Quo Fallacy: Implicit Bias and Fallacies of Argumentation", in Brownstein & Saul (eds.) *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*. Oxford University Press.

[10] Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

[11] Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of personality and social psychology*, 105(2), 171.

[12] Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *The Behavioral and Brain Sciences*, 34(2), 57—74; discussion 74—111. doi:10.1017/S0140525X10000968

[13] Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41), 16474-16479.

[14] Nagle, J. E., Brodsky, S. L., & Weeter, K. (2014). Gender, smiling, and witness credibility in actual trials. *Behavioral sciences & the law*, 32(2), 195-206.

[15] Nosek, B. A., Greenwald, A. G., and Banaji, M. R. (2007). "The Implicit Association Test at age 7: A methodological and conceptual review." In Bargh, J. A. (ed.), *Automatic Processes in Social Thinking and Behavior*. Philadelphia, PA: Psychology Press.

[16] Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, 6(1), 101.

[17] Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. Behavioral and Brain Sciences, 5(2), 187-255.

[18] Stephens, C. L. (2001). When is it selectively advantageous to have true beliefs? Sandwiching the better safe than sorry argument. *Philosophical Studies*, 105(2), 161-189.

[19] Tetlock, P. E., & Mitchell, G. (2009). Implicit bias and accountability systems: What must organizations do to prevent discrimination? *Research in organizational behav-*

ior, 29, 3-38.

[20] Saul, J. (2012). Scepticism and implicit bias. *Disputatio*, 5(37), 243-263.

[21] Sperber, D. (2006). An evolutionary perspective on testimony and argumentation. In R. Viale, D. Andler, & L. Hirschfeld (Eds.), *Biological and cultural bases of human inference* (pp. 179—189). Mahwah, NJ: LEA.

[22] Steinpreis, R. E., Anders, K. A., & Ritzke, D. (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex roles*, 41(7-8), 509-528.

[23] Stephens, C. L. (2001). When is it selectively advantageous to have true beliefs? Sandwiching the better safe than sorry argument. *Philosophical Studies*, 105(2), 161-189.

[24] Sullivan-Bissett, E. (2015). Implicit bias, confabulation, and epistemic innocence. *Consciousness and Cognition*, 33, 548-560.

[25] Walton, D. N. (2000). Case Study of the Use of a Circumstantial Ad Hominem in Political Argumentation. *Philosophy and Rhetoric*, 33(2), 101—115. doi:10.1353/par.2000.0015

[26] Werner, E. (September 27, 2018). "Some GOP senators concede Ford's credibility, but point to lack of corroboration" Washington Post. Accessed December 16, 2018. `https://www.washingtonpost.com/business/economy/some-gop-senators-concede-fords-credibility-but-point-to-lack-of-corrobora 2018/09/27/6d97c484-c287-11e8-b338-a3289f6cb742_story.html?noredirect=on&utm_term=.e8dda6ac1c5f`

[27] Woods, J. (2013). *Errors of Reasoning: Naturalizing the Logic of Inference*. London: College Publications.

[28] Valian, V. (1998). *Why so Slow? The Advancement of Women*. Cambridge, MA: MIT Press.