

# The Fallacy Fallacy

## From the Owl of Minerva to the Lark of Arete

Andrew Aberdein

the date of receipt and acceptance should be inserted later

**Abstract** The fallacy fallacy is either the misdiagnosis of fallacy or the supposition that the conclusion of a fallacy must be a falsehood. This paper explores the relevance of these and related errors of reasoning for the appraisal of arguments, especially within virtue theories of argumentation. In particular, the fallacy fallacy exemplifies the Owl of Minerva problem, whereby tools devised to understand a norm make possible new ways of violating the norm. Fallacies are such tools and so are vices. Hence a similar problem arises with argumentative vices. Fortunately, both instances of the problem have a common remedy.

**Keywords** Scott Aikin; fallacist's fallacy; fallacy fallacy; David Hackett Fischer; David Godden; Owl of Minerva problem; vices of argument; virtues of argument

### 1 Introduction

Several authors have proposed a “fallacy fallacy”, but not all their proposals take the same form. Hence, for William Lycan, “the ‘fallacy fallacy,’ [is] that of imputing fallaciousness to a view with which one disagrees but without doing anything to show that the view rests on any error of reasoning” (Lycan 1996, 69). Similar definitions include: “Improperly connecting a fallacy with an argument, so that the argument is errantly presumed to be debunked” (Miller and Miller 2015, 172) and “the fallacy of thinking that something is a fallacy when it isn't” (Hedden 2019, 43). Such flagging up of fallaciousness on insufficient grounds is a type-one error with respect to fallacy detection. The teaching of fallacy theory is chiefly concerned with the type-two error, that of letting insidious fallacies pass for good argument. Thus the fallacy fallacy represents a salutary warning against overcorrection. However, other authors have defined the fallacy fallacy differently, as the unwarranted inference that the conclusion

---

A. Aberdein  
School of Arts & Communication, Florida Institute of Technology, 150 West University Blvd, Melbourne,  
FL 32901, USA  
E-mail: aberdein@fit.edu

of a fallacious argument must be false. This includes Christian Cotton, in what seems to be the first stand-alone treatment of the fallacy fallacy: it “occurs when one reasons that because the argument for some conclusion is fallacious, the conclusion of that argument is false” (Cotton 2019, 125) and Scott Aikin: “One commits the fallacy fallacy when one identifies a fallacy in an interlocutor’s argument and thereby takes this as a reason to reject their conclusion” (Aikin 2020, 17).

The fallacy fallacy thus exhibits an “individuation problem” akin to that John Woods diagnoses in *ad hominem* arguments (Woods 2007, 124). He schematizes the general *ad hominem* strategy as follows:

1. Sarah makes her *ad hominem* retort.
2. She *concludes* from this that the adequacy of her opponent’s case is called into doubt.
3. She *concludes* from *this* that there is reason to think that her interlocutor’s position is false.

Woods remarks of his three part scheme that “Some people are of the view that an *argumentum ad hominem* is constituted by all three components, the retort of (1) and the inferences of (2) and (3). So understood . . . an *ad hominem* begins with (1) and ends with (3). Others are of the view that the *ad hominem* has a slighter constitution, one that begins with (1) and ends with (2)” (Woods 2007, 124). We should also notice that Woods is scrupulous, as some of the sources cited above are not, to distinguish the determination of fallaciousness from the identification of a specific pattern of argument. As he remarks, for most fallacies *X*, “it is possible to specify arguments that instantiate ‘*X*’ which nevertheless are not fallacious. In other words, being an argument of the *X*-sort underdetermines whether it is a fallacy” (Woods 2007, 131). In this spirit, I shall henceforth refer to fallacy schemes when it is not determined whether a given case is fallacious. The resemblance of Woods’s scheme to the fallacy scheme is immediate. Indeed, we could adopt it as a characterization of the fallacy scheme, with one amendment: keeping (2) and (3) but substituting (1′) for (1):

- 1′. Sarah makes her fallacy diagnosis.
2. She *concludes* from this that the adequacy of her opponent’s case is called into doubt.
3. She *concludes* from *this* that there is reason to think that her interlocutor’s position is false.

We may see that the first sort of fallacy fallacy discussed above runs from (1′) to (2) and the second from (1′) to (3). In order for the former to be a fallacy, Sarah’s diagnosis must be in error; indeed, if the diagnosis is correct, the step from (1′) to (2) is the normal process of fallacy recognition. This makes the fallacy scheme central to the application of fallacy theory. This constraint does not apply to instances of the latter type; the step from (2) to (3) is unwarranted independently of the status of (1′).

At least since the work of John Pollock, epistemologists have distinguished two of the ways of attacking an inference as undercutting, which challenges the connection between the premisses and the conclusion, and rebutting, which challenges the conclusion directly (Pollock 1986). As I have observed elsewhere, (2) is an undercutter and (3) is a rebutter, so the three-component *ad hominem* is an (undercutting

and) rebutting variety and the slighter, two-component ad hominem is a (solely) undercutting variety (Aberdein 2014, 83). Let us call them  $AH_R$  and  $AH_U$ , respectively. Likewise, the two-component fallacy scheme is an undercutting attack whereas the three-component fallacy scheme is an (undercutting and) rebutting attack. Call them  $FS_U$  and  $FS_R$ . An inference may be undercut but not rebutted, if its conclusion holds for other reasons, so not every  $FS_U$  is an  $FS_R$ . Strictly speaking, inferences may also be rebutted but not undercut, by challenging the conclusion without giving any grounds for disputing the argument's cogency. However, insofar as an  $FS_R$  derives the falsehood of its conclusion from an alleged inferential weakness, every  $FS_R$  must be an  $FS_U$ .

The close relationship between the fallacy scheme and the ad hominem is perhaps unsurprising: the fallacy fallacy is to non-agent-based argument appraisal what the ad hominem is to agent-based argument appraisal. If fallacy detection is to serve any useful purpose, there must be cases of  $FS_U$  that are not fallacious; similarly, if an agent-based (and, a fortiori, a virtue-theoretic) account of argument appraisal is to do any work, there must be cases of  $AH_U$  that are not fallacious. Their stronger and less defensible cousins,  $FS_R$  and  $AH_R$ , are immaterial to the defence of either sort of appraisal, since it is no part of the appraisal of an inference to determine that its conclusion is false.

In the next section I will discuss some anticipations of the fallacy fallacy that provide yet another possible reading. §3 addresses the Owl of Minerva problem, a broader phenomenon which the fallacy fallacy and its relatives exemplify. In §4, we will see that this problem also afflicts the practice of charging others with vices, including vices of argument. Lastly, §5 discusses a possible remedy.

## 2 Anticipations

The fallacy fallacy is anticipated in Schopenhauer's *Art of Controversy* (in which errors of reasoning are ironically proposed as advice to the unscrupulous arguer): "Should your opponent be in the right, but, luckily for your contention, choose a faulty proof, you can easily manage to refute it, and then claim that you have thus refuted his whole position. This is a trick which ought to be one of the first" (Schopenhauer 1896, 45).<sup>1</sup> Schopenhauer's "trick" is explicitly an  $FS_R$ . He also links it to ad hominem argument: "it is, at bottom, an expedient by which an *argumentum ad hominem* is put forward as an *argumentum ad rem*" (Schopenhauer 1896, 45). For Schopenhauer, an ad hominem shows that some proposition "is inconsistent with other statements or admissions of our opponent" whereas an ad rem shows "that the proposition is not in accordance with the nature of things" (Schopenhauer 1896, 13). Hence Schopenhauer does not draw the connection to the arguer's character typical of more recent accounts of the ad hominem.

The character of the arguer receives direct attention in a more recent anticipation of the fallacy fallacy from another classic work of fallacy theory, David Hackett Fischer's *Historians' Fallacies*. He designates as "the fallacist's fallacy" any of the following "profoundly wrong" propositions:

<sup>1</sup> For further discussion of *Art of Controversy*, see (Davies 2015; Lemanski 2021).

1. An argument which is structurally fallacious in some respect is therefore structurally false<sup>2</sup> in all respects.
2. An argument which is structurally false in some respect, or even in every respect, is therefore substantively false in its conclusion.
3. The appearance of a fallacy in an argument is an external sign of its author's depravity.
4. Sound thinking is merely thinking which is not fallacious.
5. Fallacies exist independent of particular purposes and assumptions (Fischer 1970, 305).

Prop. 1 and 2 closely resemble  $FS_U$  and  $FS_R$ , respectively—the resemblance is not exact, since their antecedents exclude wholly blameless inferences. The other propositions also represent salutary warnings. Prop. 4 cautions that the theory of argument is not exhausted by fallacy theory and Prop. 5 reminds us that classifications of fallacies are only possible within broader theoretical contexts. (Fischer's example is the dependance of Aristotle's fallacy of accident on the distinction in his metaphysics between essential and accidental properties.)

Prop. 3 holds much greatest interest. Notably, we may disaggregate it into two components, each of which rewards separate study:

- 3a. An arguer's fallacies are evidence of the arguer's persistent dispositions.
- 3b. The persistent dispositions responsible for fallacies are moral vices.

Prop. 3a could be taken as an instance of what social psychologists call the "fundamental attribution error" (Ross 1977). This "error of ignoring situational factors and overconfidently assuming that distinctive behaviour or patterns of behaviour are due to an agent's distinctive character traits" is one of the most significant challenges to virtue theories (Harman 1999, 315). The situationist critique has thus been the subject of considerable discussion. Elsewhere I have suggested how virtue theories of argument may respond (Aberdein 2016c): one major weakness of the critique is that the fundamental attribution error relies on a behaviourist conception of character traits. Virtue theorists from Aristotle onwards typically make clear that virtues are not just rote reactions to specific stimuli, they are situation-responsive (Webber 2006, 206). Our present concern is with vices, but similar considerations apply: psychologically realistic vices are also situation-responsive and thereby predict most, but not all, behaviour (Cassam 2016, 173). So, if Prop. 3a is interpreted as attributing fallacies to behaviourist character traits, it is indeed an instance of the fundamental attribution error, but is not a thesis any virtue theorist need defend. On the other hand, if Prop. 3a is interpreted as taking fallacies as evidence, however weak, of situation-responsive vices, then it is consistent with a virtue theory but evades the situationist critique. By contrast, Prop. 3b is straightforwardly false. It links failures of argument to failures of moral character. This is a tempting but fundamentally mistaken misreading of the virtue programme of argument analysis: as I have stressed elsewhere, only argumentative vices should be seen as relevant to the appraisal of arguments (Aberdein 2010, 171). Thus the virtue programme does not claim to ground the analysis of an argument in the arguer's *moral* virtues or vices. Indeed, this is an error which the virtue programme can help to expose.

<sup>2</sup> Fischer does not define "structurally false". I will treat it as a synonym for "structurally fallacious".

This error can be brought into clearer focus by looking at some of the close logical relatives of Prop. 3. The contrapositive of Prop. 3 states that “Evidence of an author’s moral righteousness is evidence that their arguments are free from fallacy”. This treatment of the moral status of an arguer as support for the cogency of their arguments or the truth of their conclusions is what Jamie Whyte has termed “morality fever”, under which heading he includes the ideas “that you can’t argue with the morally sincere, that caring licenses irrationality, . . . [and] that sincerity may substitute for reason” (Whyte 2005, 156). The converse of Prop. 3 states that “Evidence of an author’s depravity is evidence that their arguments are fallacious”. This bears some resemblance to Lawrence Powers’ version of the *ad hominem*: “the opponent is regarded as evil, deceitful, and diabolically clever. [Therefore,] we had better not listen to the opponent’s arguments; we listen to such arguments at our peril” (Powers 1998). None of these theses follow from a virtue theoretic approach to the appraisal of arguments.

To see why this is so, consider the dilemma that José Gascón presents as confronting any virtuous evaluation of arguments:

Either (a) one takes argumentative virtues as the basis from which the quality of arguments derives, or (b) one admits that cogency is not to be defined in terms of qualities of the arguer. Option (a) clashes with a widespread and well-established intuition that arguments should be evaluated on their own merits, regardless of who puts them forward. On the other hand, option (b) leads to a gap in virtue argumentation theory regarding argument quality (Gascón 2018, 163).

Gascón takes option (b), whereas I maintain that option (a) is viable. But how can (a) be reconciled with Gascón’s “widespread and well-established intuition”? As Gascón observes, “a virtuous arguer can put forward a bad argument, and a vicious arguer can put forward a cogent argument” (Gascón 2018, 168). A *perfectly* virtuous arguer would, *ex hypothesis*, only ever advance good arguments. However, it is unclear whether there are any perfectly virtuous arguers, or how we should recognise them if they exist. Instead, there are *typically* virtuous arguers, who can certainly put forward bad arguments from time to time. However, they are not arguing virtuously in so doing. Likewise, a vicious arguer can put forward a good argument on occasion, but only by arguing as a virtuous arguer would argue. Thus the practical measure of argument quality available to the virtue theorist is not whether the arguers are actually virtuous—perhaps an impossible question to answer—but whether they are arguing as virtuous arguers would argue. What this standard actually comprises may not be so very different from what more conventional accounts of good argument propose (at least, it should not be any laxer): good arguing will still be expected to start from true premisses and reason to conclusions that follow from them with certainty or high likelihood. But this is good arguing precisely because it is how a virtuous arguer is overwhelmingly likely to argue. Thus “evaluate arguments on the basis of who puts them forward”, with its disturbing echoes of Prop. 3 and its relatives, is not a proposal that a virtue theorist of argument should embrace and nor is it a serious rival to the familiar standard, “evaluate arguments on their own merits”. Rather, when properly understood, these are two differently incomplete descriptions of the same strategy:

evaluate arguments on their own merits as manifest in the actions of the arguers who put them forward (and are otherwise engaged in them).

This cluster of errors can shed some light on political phenomena that have lately attracted concern and theoretical discussion, including *issue alignment*, the tendency of individuals to coincide in their views even across not obviously linked issues, such as abortion and immigration (e.g. Joshi 2020), and *affective polarization*, the degeneration of political disagreement into acrimony and vitriol (e.g. Iyengar and Westwood 2015). Suppose that you and I disagree about some issue, say  $P$ , and that I believe, rightly or wrongly, that I have found one or more fallacies in your arguments. By Prop. 3, I could reason from the fallaciousness of your arguments to the badness of your character, justifying my treating you with hostility, increasing the severity of my disagreement and leading to polarization. If I then learn of your opinions on some other issue, say  $Q$ , on which I have no strong feelings, and perhaps even share your opinions, I could employ the converse of Prop. 3 to infer that your arguments must be fallacious, because of your terrible character. Lastly, by  $FS_R$ , I may conclude that you must be wrong about  $Q$  and that I should disagree with you about this as well as  $P$ , leading to issue alignment.<sup>3</sup> Notice that a virtue theory of argument does not endorse any of these steps: if I find fallacies in your arguments, Prop. 3a licenses me to revise my opinion of your *argumentative* character downwards, but does not justify any reappraisal of your moral character. Even if I conclude that your argumentative character is very bad, that does not in itself justify my treating you with hostility. And while I should treat your arguments about  $Q$  with greater scepticism if I take your argumentative character to be poor, I should still take those arguments seriously and could still end up accepting your conclusions (perhaps for other reasons).

### 3 The Owl of Minerva

Scott Aikin and Robert Talisse have recently posed the Owl of Minerva Problem, named by allusion to the Hegelian dictum that the Owl of Minerva flies only at dusk. As they gloss this aphorism,

Minerva's Owl, a symbol of wisdom, is active only at the end of the day, when all the deeds are done. One reason we believe this is the case is that we humans are changed by our insights, sometimes for the better and sometimes for the worse. We and our theories about ourselves have a looping pattern—we have concepts of ourselves, and they at once function as both descriptive and prescriptive of our behavior (Aikin and Talisse 2020, 477 f.).

Such loops can be virtuous, reinforcing newly acquired positive patterns of behaviour until they become entrenched. But they can also be vicious: the new conceptual framework can open up new possibilities for error. On Aikin's account, the fallacy fallacy originates in just such a vicious loop: "the fallacy fallacy is possible only if we have the *concept of fallacy*. Teaching fallacy theory makes a whole new kind of fallacy possible" (Aikin 2020, 17). Aikin, like many who have taught this material,

<sup>3</sup> This narrative is not intended to be an exhaustive account of the sources of polarization or issue alignment. I am grateful to an anonymous referee for stressing this point.

is concerned lest his students become what Edward Damer dubs “fallacy mongers”: “Some people, with a little knowledge of fallacious reasoning, develop a kind of obsession with identifying fallacies in the utterances of others. They sniff suspiciously at every argument and point of debate. Such pouncing on others often creates alienation” (Damer 2009, 60). The fallacy monger is an inveterate perpetrator of fallacy fallacies.

David Godden has observed that the Owl of Minerva Problem, “the worsening of our discursive circumstances, following on our attempts to make explicit the norms of our practices”, gives rise to two importantly distinct faults: it may be accidental or deliberate (Godden 2022, 45). This mirrors a traditional distinction in fallacy theory between paralogisms, innocent mistakes in reasoning, and sophisms, designed to deceive (cf. Aberdein 2016b, 415). In each case, it is the latter, wilful, variety of problem that presents the greater hazard, as it is apt to be incorrigible. As Godden explains, it lends itself to a cynical, bad faith approach to argument (reminiscent of that proposed ironically by Schopenhauer):

the thought that we may keep two sets of rational books, one private and one public: where acceptability is established, and entitlement secured, in some private ledger of ‘reason’ which is unaccountable to public audit, while the books of public reason (e.g., reasoned discourse in the publics of deliberative democracy) merely record exercises of rhetorical power whereby acceptance is coerced (Godden 2022, 58).

Thus Godden is concerned about arguers who prioritize winning at all costs. Elsewhere, I have distinguished between “*real* winners, who win the argument deservedly, because they are in the right, and *mere* winners, who win the argument undeservedly, because they are in the wrong”, where winning the argument consists in convincing the audience, an achievement which is presumptively independent of the facts that determine whether the arguer is in the right or in the wrong (Aberdein 2016a, 1). With such priorities, arguers may settle for mere winning, even when they know they are in the wrong. This is the perspective identified by Katharina Stevens and Daniel Cohen as the *adversarial attitude*: “arguers with adversarial attitudes behave in ways that enhance their prospects of winning, including behavior intended to prevent reasons which might work against their goals from being recognized” (Stevens and Cohen 2021, 901). Stevens and Cohen distinguish the adversarial attitude from the less compromising *adversarial stance* required of an arguer fulfilling the *adversarial function* of argument. Nonetheless, they deprecate the overemphasis of the adversarial function within what Cohen has termed the Dominant Adversarial Model (DAM) (Cohen 2015, 259). As Cohen complains, this model is substantially incomplete: “The DAM account cannot make any sense of arguers who walk away from an argument having had their positions changed, either by winning or losing or listening and learning, and declaring it a good argument on that account” (Cohen 2015, 264). In Rawlsian terms, DAM is an ideal theory: “Modern DAM theories come with the promise that following the rules will bring reality close enough to the ideal, but rules cannot replace genuine cooperative attitudes, nor can they preclude the impacts of adversarial attitudes” (Stevens and Cohen 2021, 909). Realistic engagement with argumentative practice requires non-ideal theory.

#### 4 Vice-Charging

The Owl of Minerva Problem does not only afflict theories of argument; it is a potential problem for any theory that involves explicit reflection on the norms of our practices. Of course, that includes virtues and vices, so it is reasonable to anticipate that this apparatus can give us the resources to create hitherto unthinkable varieties of vice. Ian James Kidd provides a helpful context for such reflection in his discussion of vice-charging. For Kidd, a vice-charge is a threefold relationship between a critic, a target, and an audience: the critic accuses the target of some vice, typically in front of an audience. Each of the three might, but need not, benefit from the charge. Hence he can draw a distinction between two species of vice-charge, which I will term “eristic” and “robust”:<sup>4</sup> “[Eristic] charges involve reportage of one’s negative judgments, but not the presentation of any reasons, evidence, or feelings in support of them, so they do not do any real critical work—they might simply let off steam, vent frustration, or register disapproval” (Kidd 2016, 183); “A robust charge involves primarily an active and intentional attempt to persuade others [perhaps the target, perhaps the audience] for the ultimately ameliorative reason of making things better” (Kidd 2016, 184). For the latter, benign variety of vice-charge to succeed requires a suitable account of epistemic responsibility and a broad consensus, critically including the target, as to what the vice comprises and whether the case at issue exemplifies it. None of this is easily done, so Kidd unsurprisingly concludes that “robust critical practice of epistemic vice-charging is possible in principle, but very difficult in practice” (Kidd 2016, 194). As Quassim Cassam observes, “the biggest pitfall of vice-charging [is] the ever-present temptation to see epistemic vice every time one encounters a group of non-conformists whose views or conduct are fundamentally different from one’s own” (Cassam forthcoming, 12). That is, it lends itself to the same sort of type-one error noted above as underpinning the fallacy fallacy.

Godden’s subdivision of the pernicious effects of the Owl of Minerva problem extends to eristic vice-charging: a vicious vice-charge consisting of comparatively harmless venting and the like is naïve; a more sinister attempt to discredit the other party is cynical. The former includes the more benign varieties of eristic dialogue or quarrel discussed by Douglas Walton and Erik C. W. Krabbe, whereby “powerful, repressed feelings that could otherwise cause constant disruptions and violent physical demonstrations of unrest and frustration” are released (Walton and Krabbe 1995, 77).<sup>5</sup> The latter is perhaps exemplified by the dominance aspect of what Justin Tosi and Brandon Warmke call “moral grandstanding”, the use of moral talk to enhance one’s social status: “some grandstanders use moral talk for darker purposes. They grandstand to dominate others. They use moral talk to shame or silence others and create fear. They verbally threaten and seek to humiliate. They try to impress people by derogating their rivals” (Tosi and Warmke 2020, 17). Just as cynical cases of fallacy mongering are likely to prove more incorrigible than naïve ones because they further the dialectical goals of the arguer, vice-charging pursued for the cynical goal

<sup>4</sup> Resisting Kidd’s pejorative use of “rhetorical” for the eristic charge.

<sup>5</sup> More lighthearted instances include “ludic flying” (Parks 1986, 441) and the dozens (Henning 2021, 22).

of improving the social status of the vice-charger through an exercise of dominance should be expected to be harder to remedy than more careless cases.

Vicious vice-charging is related to many more familiar vices in the long history of vice and virtue talk. It has echoes both of self-righteousness, in the implicit assumption that the critic is free of the vice at issue, and of hypocrisy, in the likelihood that the critic is, in fact, also guilty of the selfsame vice. But the vice which matches it most closely is perhaps *moralism*. As defined by Alfred Archer, this comprises “an inflated sense of the extent to which moral criticism is appropriate” (Archer 2018, 343). Critically, as with other offspring of the Owl of Minerva, it depends on a reflective awareness of our own normative practices and it works to undermine those practices; it is “a vice that threatens our ability to persuade others that their behaviour is unacceptable” (Archer 2018, 348). So, where the fallacy fallacy is a fallacy that arises from misapplication of fallacy-charging, moralism is a vice which arises from misapplication of vice-charging. If, as I have suggested elsewhere, “fallaciousness may be grounded in the vices of arguer and audience” (Aberdein 2016b, 418), moralism should have a counterpart applicable to argumentation that grounds the fallacy fallacy. Moralism is not among the argumentative vices that I have hitherto proposed but it is closely related to two of them: misplaced zeal, a subtype of undue willingness to engage in argumentation, and unfairness to others in evaluating their arguments, a subtype of unwillingness to listen to others (Aberdein 2016b, 416). There is no reason to expect each fallacy to correspond to a unique vice; if anything we should expect the contrary (Aberdein 2016b, 420). So I tentatively ascribe the fallacy fallacy to the conjunction of these two vices.

## 5 The Lark of Arete

What’s to be done? Godden’s account of the Owl of Minerva problem provides grounds for optimism as well as pessimism: the problem may be “technically ineliminable, [but] it is typically remediable” (Godden 2022, 54). The same self-awareness which enables the cynical to devise new ways to exploit our normative practices also enables the well-intentioned to develop their argumentative virtues. Godden thus proposes a beneficent sibling for the Owl of Minerva, the Lark of Arete: where “the pessimistic Owl of Minerva makes possible a kind of vicious cycle of incompetence, the optimistic Lark of Arete makes possible a virtuous cycle of autonomous mastery and virtuosity” (Godden 2022, 57). The Lark of Arete thereby opens up the prospect of “converting a behavior into an action, a habit into a practice, a creature into an agent, an agent into a virtuoso, and a public of arguers into a culture of argumentative virtuosity” (Godden 2022, 58). What are the implications for argumentative virtues and vices? As discussed in §2, virtues should be seen as situation-responsive. This sensitivity to circumstance requires help from other virtues, either as part of a web of interacting virtues or by appeal to regulatory virtues. With respect to argumentation, there are several candidates for the latter role, including humility (Aberdein 2016c, 2020) and phronesis (Aberdein 2021). In either case, this suggests that becoming argumentatively virtuous enough to routinely argue in virtuous ways is no small task. Indeed, the naïve arguer is unlikely even to fully understand the scale of the task.

So, with argumentative vices an ever present temptation, how does anyone ever come close to meeting the standards? Here the Lark of Arete comes to our aid. Godden echoes Agnes Callard's concept of acting on a proleptic reason: by virtue of values which we aspire to but do not yet possess (Callard 2016). As he explains, even a naïve arguer may pursue "improvement through the increasingly purposive self-regulation of future action" (Godden 2022, 58). The fallacy fallacy is thus a symptom of how reflection on our own norms can be abused but such reflection is also integral to the acquisition of the virtuous character necessary for good argument.

## References

- Aberdein, Andrew. 2010. Virtue in argument. *Argumentation* 24(2): 165–179.
- Aberdein, Andrew. 2014. In defence of virtue: The legitimacy of agent-based argument appraisal. *Informal Logic* 34(1): 77–93.
- Aberdein, Andrew. 2016a. Arguments with losers. *Florida Philosophical Review* 16(1): 1–11.
- Aberdein, Andrew. 2016b. The vices of argument. *Topoi* 35(2): 413–422.
- Aberdein, Andrew. 2016c. Virtue argumentation and bias. In *Argumentation, Objectivity and Bias: Proceedings of the 11th International Conference of the Ontario Society for the Study of Argumentation (OSSA), May 18–21, 2016*, eds. Patrick Bondy and Laura Benacquista. Windsor, ON: OSSA.
- Aberdein, Andrew. 2020. Intellectual humility and argumentation. In *The Routledge Handbook of Philosophy of Humility*, eds. Mark Alfano, Michael Lynch, and Alessandra Tanesini, 325–334. London: Routledge.
- Aberdein, Andrew. 2021. Was Aristotle a virtue argumentation theorist? In *Essays on Argumentation in Antiquity*, eds. Joseph Bjelde, David Merry, and Christopher Roser, 215–229. Cham: Springer.
- Aikin, Scott F. 2020. The Owl of Minerva problem. *Southwestern Philosophical Review* 36(1): 13–22.
- Aikin, Scott F. and Robert B. Talisse. 2020. Disagreement, polarization, and the flight of Minerva's Owl: A reply to Vaidya. *Argument* 10(2): 473–482.
- Archer, Alfred. 2018. The problem with moralism. *Ratio* 31(3): 342–350.
- Callard, Agnes. 2016. Proleptic reasons. *Oxford Studies in Metaethics* 11: 129–154.
- Cassam, Quassim. 2016. Vice epistemology. *The Monist* 99(2): 159–180.
- Cassam, Quassim. forthcoming. Misunderstanding vaccine hesitancy: A case study in epistemic injustice. *Educational Philosophy and Theory* DOI: 10.1080/00131857.2021.2006055.
- Cohen, Daniel H. 2015. Missed opportunities in argument evaluation. In *Proceedings of ISSA 2014: Eighth Conference of the International Society for the Study of Argumentation*, eds. Bart J. Garssen, David Godden, Gordon Mitchell, and A. Francisca Snoeck Henkemans, 257–265. Amsterdam: Sic Sat.
- Cotton, Christian. 2019. Argument from fallacy. In *Bad Arguments: 100 of the Most Important Fallacies in Western Philosophy*, eds. Robert Arp, Steven Barbone, and Michael Bruce, 125–127. Chichester: John Wiley & Sons.
- Damer, T. Edward. 2009. *Attacking Faulty Reasoning: A Practical Guide to Fallacy-Free Arguments*. Belmont, CA: Wadsworth, sixth ed.
- Davies, Richard. 2015. Eristic dialectic: The fencing master's judgment. *Eris* 1(1): 20–33.
- Fischer, David Hackett. 1970. *Historians' Fallacies: Toward a Logic of Historical Thought*. New York, NY: Harper.
- Gascón, José Ángel. 2018. Virtuous arguers: Responsible and reliable. *Argumentation* 32(2): 155–173.
- Godden, David. 2022. Getting out in front of the Owl of Minerva problem. *Argumentation* 36(1): 35–60.
- Harman, Gilbert. 1999. Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society* 99: 315–331.
- Hedden, Brian. 2019. Hindsight bias is not a bias. *Analysis* 79(1): 43–52.
- Henning, Tempest. 2021. "I said what I said"—Black women and argumentative politeness norms. *Informal Logic* 41(1): 17–39.
- Iyengar, Shanto and Sean J. Westwood. 2015. Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science* 59(3): 690–707.
- Joshi, Hrishikesh. 2020. What are the chances you're right about everything? An epistemic challenge for modern partisanship. *Politics, Philosophy & Economics* 19(1): 36–61.

- Kidd, Ian James. 2016. Charging others with epistemic vice. *The Monist* 99(2): 181–197.
- Lemanski, Jens. 2021. Discourse ethics and eristic. *The Polish Journal of Aesthetics* 62: 151–162.
- Lycan, William G. 1996. *Consciousness and Experience*. Cambridge, MA: The MIT Press.
- Miller, J. Steve and Cherie K. Miller. 2015. *Why Brilliant People Believe Nonsense: A Practical Text for Critical and Creative Thinking*. Acworth, GA: Wisdom Creek Academic.
- Parks, Ward. 1986. Flyting, sounding, debate: Three verbal contest genres. *Poetics Today* 7(3): 439–458.
- Pollock, John L. 1986. *Contemporary Theories of Knowledge*. Totowa, NJ: Rowman and Littlefield.
- Powers, Lawrence H. 1998. Ad hominem arguments. In *Argumentation and Rhetoric*, eds. H. V. Hansen, C. W. Tindale, and A. V. Colman. Newport News, VA: Vale.
- Ross, Lee. 1977. The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology* 10: 173–220.
- Schopenhauer, Arthur. 1896. *The Art of Controversy and Other Posthumous Papers*. London: Swan Sonnenschein. Trans. T. Bailey Saunders.
- Stevens, Katharina and Daniel H. Cohen. 2021. Angelic devil's advocates and the forms of adversariality. *Topoi* 40(5): 899–912.
- Tosi, Justin and Brandon Warmke. 2020. *Grandstanding: The Use and Abuse of Moral Talk*. New York, NY: Oxford University Press.
- Walton, Douglas N. and Erik C. W. Krabbe. 1995. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. Albany, NY: State University of New York Press.
- Webber, Jonathan. 2006. Virtue, character and situation. *Journal of Moral Philosophy* 3(2): 193–213.
- Whyte, Jamie. 2005. *Crimes Against Logic*. New York, NY: McGraw-Hill.
- Woods, John. 2007. Lightening up on the ad hominem. *Informal Logic* 27(1): 109–134.