

# THE SELF-EFFACEMENT GAMBIT

Jack Woods

**Abstract:** Philosophical arguments usually are and nearly always should be abductive. Across many areas, philosophers are starting to recognize that often the best we can do in theorizing some phenomena is put forward our best overall account of it, warts and all. This is especially true in areas like logic, aesthetics, mathematics, and morality where the data to be explained are often based in our stubborn intuitions. While this methodological shift is welcome, it's not without problems. Abductive arguments involve significant theoretical resources which themselves can be part of what's being disputed. This means that we will sometimes find otherwise good arguments suggesting their own grounds are problematic. In particular, sometimes revising our beliefs on the basis of an argument can undermine the very justification we used in that argument. This feature, which I'll call self-effacingness, occurs most dramatically in arguments against our standing views on the subject matters mentioned above: logic, mathematics, aesthetics, and morality. This is because these subject matters all play a role in how we reason abductively. This isn't an idle fact; we can resist some challenges to our standing beliefs about these subject matters exactly because the challenges are self-effacing. The self-effacing character of certain arguments is thus both a benefit and a limitation of the abductive turn and deserves serious attention. I aim to give it the attention it deserves.

Philosophical arguments usually are and nearly always should be abductive. From Lewis's claim that philosophy is really a game of weighing costs,<sup>1</sup> to the rise of anti-exceptionalism in logic,<sup>2</sup> to the near universal reliance on methods like reflective equilibrium in ethics and metaethics,<sup>3</sup> philosophers are starting to recognize that often the best we can do in theorizing some

---

<sup>1</sup> See the introduction to his 1983 for a classic statement.

<sup>2</sup> See Hjortland 2017; Priest 2006, 2016; Russell 2015; and Williamson 2017, Unpublished. There are still serious problems for implementing anti-exceptionalist credos about logic (Woods Forthcoming *a*, Forthcoming *b*).

<sup>3</sup> See Scanlon 2014 for trenchant defense and McPherson 2015 for worries.

phenomena is put forward our best overall account of it, warts and all. This is especially true in areas like logic, aesthetics, mathematics, and morality where the data to be explained are often based in our stubborn intuitions. While this methodological shift is welcome, it's not without problems. Abductive arguments involve significant theoretical resources which themselves can be part of what's being disputed. This means that we will sometimes find otherwise good arguments which suggest their own grounds are problematic. In particular, sometimes revising our beliefs on the basis of such an argument can undermine the very justification we used *in* that argument.

This feature, which I'll call *self-effacement*, occurs most dramatically in arguments against our standing views on the subject matters mentioned above: logic, mathematics, aesthetics, and morality. This is because these subjects all play a role in how we reason abductively (on how, see below). This isn't an idle fact; we can resist some challenges to our standing beliefs about these subject matters exactly because the challenges are self-effacing. The self-effacing character of certain arguments is thus a benefit and limitation of the "abductive turn" and deserves serious attention. I aim to give it the attention it deserves.

I'll start by distinguishing two types of self-effacement, giving detailed examples of each. The first type occurs when accepting a conclusion undermines the inputs to our abductive machinery; the second type occurs when accepting a conclusion undermines abductive methodology itself. Logic, mathematics, aesthetics, and morality each allow self-effacement in at least one of these two ways. In fact, it's likely that otherwise good self-effacing arguments actually occur for each subject matter.

Many of these arguments take the form of challenges to all of our beliefs from one (or more) of these subjects. Recent examples include debunking arguments against all our moral beliefs like those articulated by [Harman \(1977\)](#), [Joyce \(2001\)](#), and [Street \(2006\)](#), as well as more general skeptical challenges like those posed by [Cohen \(2000\)](#). I'll focus on these global challenges in what follows, leaving discussion of the less interesting case of local self-effacement largely to the side.

I'll clarify how, why, and most importantly when self-effacement blocks these global challenges to our standing beliefs; I call this way of blocking challenges *the self-effacement gambit*. Unfortunately, we can still use many blocked challenges to build new challenges to our standing beliefs. These blocked challenges are still cases where our beliefs about *T* say that they're not the best overall beliefs about *T*. That's a significant vice for any set of beliefs. So, we can sometimes use the viciousness of *T* to argue, *without* self-effacement, that we ought to revise *T*.

For this reason, it's plausible that our beliefs about morality and aesthetics can only gain limited protection from the self-effacement gambit. This is especially plausible when we focus on theoretical costs of wholesale revision or rejection of our moral and aesthetic beliefs, such as costs to our

abductive practice itself.<sup>4</sup> The ability to use abduction survives wholesale revision or rejection of our views on these subjects, so we cannot clearly block the new challenges. Though my main focus is elsewhere, these new challenges clearly raise problems for “self-defeat” strategies for avoiding evolutionary debunking (Vavova 2014).

Mathematics and logic fare better. We use logic and mathematics not only in justifying the “inputs” to an abductive argument, but also in aggregating the facts about theoretical virtues so input. This fact shows that rejecting our logical and mathematical beliefs *en bloc* would be tantamount to abandoning the abductive method entirely. So, while the cogency of any self-effacing argument is problematic, the costs of coming to doubt our beliefs about mathematics and logic are extreme enough that we can reasonably ignore the otherwise problematic existence of self-effacement.

This gives rise to a more nuanced “abductivized” strategy for blocking all these challenges, old and new. We can call this *the revised self-effacement gambit*. The enormous cost of rejecting our current mathematical and logical beliefs overwhelms the viciousness of self-effacing arguments. The costs of rejecting our current moral and aesthetic beliefs aren’t clearly high enough to do so. So, mathematics and logic are on firmer ground than morality, even if certain challenges to morality are also self-effacing.

In short, if we use abduction to decide what to do when our logical and mathematical beliefs are self-effacing, we get a clear answer. Since the costs of abandoning these beliefs include the costs of abandoning abductive methodology entirely, we should shrug off challenges to our basic logical and mathematical beliefs which arise from self-effacement. But we cannot, as we should not, put aside analogous challenges to our moral and aesthetic beliefs.

## 1 Abductive Arguments

Abductive arguments can be analyzed into a series of content premises articulating relevant bits of information and that these are the bits of information that matter, a “structural” premise which aggregates the information from the content premises, a linking principle taking us from the aggregated information to what we ought to believe, and a defeasible conclusion drawn on the basis of the previous. The particular cases we’re interested in are comparisons of different theories where the conclusion tells us to accept one of these theories and not the other.

The point of giving such abductive arguments is *conferring justification* on revising or rejecting our current beliefs in line with this conclusion. It’s

---

<sup>4</sup> I’ll put aside practical and theoretical costs unrelated to abductive methodology. The costs of abandoning our standing moral or aesthetic beliefs might be different when we factor these in. But, since the costs to abduction of abandoning mathematics and logic are enormous, stable, and relatively uncontentious, we can simplify. I aim to defend mathematics and logic from certain challenges and for this we only need some of the theoretical costs.

this context that we focus on here, leaving questions about whether the resulting beliefs themselves are otherwise *justifiable* or *justified* to the side. This is a familiar notion of justification, albeit one that gets less attention than it should in the literature.<sup>5</sup> This is especially striking since it's this notion of justification that's important for the case of revising our most basic beliefs about matters like mathematics, logic, morality, and aesthetics.

I'll not worry overmuch about differences between rejection, wholesale revision, or complete doubt of our views.<sup>6</sup> This is largely because both rejecting and doubting all our beliefs about some subject matter are severe kinds of wholesale revision. We could explicitly represent them as such, but I'll simply trust the reader not to get confused.

Even the particular abductive arguments we're concerned with can be represented in many ways, including suppressing some of these premises into the act of inference itself. As it will make certain points easier to articulate, we'll be pedantic about representing premises and linking principles and we'll "deductivize" the argument.<sup>7</sup> We'll also make clear how to extend these arguments by the claim that the defeasing conditions *are not met*, yielding the further conclusion that we ought to revise.

Our basic abstract argument compares the theoretical virtues of two incompatible theories of some phenomena (the generalization to  $n$  many theories is obvious):

CONTENT <sub>1</sub> :	Theory <sub>1</sub> of $A$ has theoretical virtue $v$ to degree $j$ .
CONTENT <sub>2</sub> :	Theory <sub>2</sub> of $A$ has theoretical virtue $v$ to degree $k$ .
	...
STRUCTURAL:	Theory <sub>1</sub> is more theoretically virtuous than theory <sub>2</sub> . (from CONTENT <sub>1</sub> , CONTENT <sub>2</sub> , . . .)
THEORY CHOICE:	We ought, <i>ceteris paribus</i> , to believe the most theoretically virtuous theory of $A$ ;
SUB-CONCLUSION:	We ought to believe theory <sub>1</sub> unless we have sufficient additional reason to believe theory <sub>2</sub> . (from STRUCTURAL and THEORY CHOICE)
UNDEFEATED:	We do not have sufficient additional reason to believe theory <sub>2</sub> .
CONCLUSION:	We ought to believe theory <sub>1</sub> and not theory <sub>2</sub> . (from SUB-CONCLUSION and UNDEFEATED)

<sup>5</sup> See Engel 1992 for this point, sensible puzzlement about the lack of attention, and an analysis of this kind of justification.

<sup>6</sup> These differences are important, but they're not important for the point I make below.

<sup>7</sup> My pedantry has limits. I do not, for example, exhibit the justification of each content premise.

We can fill in this schema to obtain particular abductive arguments.<sup>8</sup> For example, let ‘theory of *A*’ mean “non-pragmatic explanation of our particular moral beliefs, intuitions, and perceptions.”<sup>9</sup> Suppose that it’s theoretically vicious for an explanation to posit brute connections between facts. Then it’s problematic to hold that our moral beliefs and intuitions are largely accurate while also holding that there’s no explicable systematic connection between our particular moral views and what makes them accurate.

After all, in the paradigmatic cases where we take our intuitions and perceptions to be accurate, we can explicate this connection. Our method for explaining why we have particular sensory perceptions uses the things and events in the world which make our perceptions accurate to causally explain why we have these perceptions. The connection there is explicable. If the analogous connection is brute for an explanation of our moral beliefs, intuitions, and perceptions, that’s a strike against it.

Many non-naturalist accounts of morality are thus going to have a strike against them, at least in their present form. Letting theory<sub>1</sub> of our particular moral beliefs be that they’re the product of social conditioning and letting theory<sub>2</sub> of our particular moral beliefs be that they somehow brutally track real moral facts and properties, we get our content premises: theory<sub>1</sub> is relatively virtuous, theory<sub>2</sub> relatively vicious.

If we add as content premises that the rest of the relevant virtues and vices of these explanations are roughly on a par, the weighing procedure is easy. By THEORY CHOICE we then ought to adopt theory<sub>1</sub>, rejecting theory<sub>2</sub>, unless we have sufficient reason to hold onto theory<sub>2</sub>. If we don’t have sufficient reason to resist the argument’s intuitive conclusion (i.e., we have the relevant version of UNDEFEATED), then we ought to adopt theory<sub>1</sub> and reject theory<sub>2</sub>.

Adding the further assumption that we currently believe theory<sub>2</sub>, we obtain a loose version of Harman’s (1977) debunking argument against stark-raving moral realism.<sup>10</sup> Whether Harman’s argument works is up for grabs; in particular, which theoretical virtues are relevant and what it would take to defeat SUB-CONCLUSION are contentious, as I discuss elsewhere

---

<sup>8</sup> This schema is just an indicative working example. The reader will easily be able to rejigger it for other uses of abduction.

<sup>9</sup> ‘Non-pragmatic’ in the sense of a causal, grounding, or “real explanation” explanation of our particular moral beliefs. I take the existence of non-pragmatic explanations for granted here.

<sup>10</sup> I’ve gone with a simple “no brute connections” construal of Harman for simplicity. To reconstruct his actual argument, we’d need a lot more detail. For instance, Harman argues that the non-naturalist realist explanations targeted contain explanatory dross and then implicitly invokes the claim that explanations without dross are better than otherwise equivalent explanations. Spelling this out carefully takes work. See Sayre-McCord 1988 for some of that work. My prior article (2018b) contains a historically accurate reconstruction of Harman’s argument which draws on Sayre-McCord.

(2018b).<sup>11</sup> Nevertheless, it's a clear instance of schema above, though an instance that doesn't look self-effacing. Section 2.1 sketches a debunking-style argument against our moral views that does look self-effacing.<sup>12</sup>

Now that we've given an example, let's return to describing our schema. It can be usefully broken into two nested arguments, as indicated by the horizontal line. The intuitive conclusion of the "inner" argument—that we ought to believe theory<sub>1</sub>—is defeasible. Here we've made the *literal* conclusion of the inner argument indefeasible by listing the defeasing conditions in an antecedent. This is unusual, but it forestalls certain confusions. This structure makes clear that it's in light of UNDEFEATED that we can conclude outright that we ought to believe this or that theory. Since it will be useful for the below formulations, we'll call the claim that we ought to believe theory<sub>1</sub> and not believe theory<sub>2</sub> the *directive* of both the "outer" argument to CONCLUSION and the "inner" argument to SUB-CONCLUSION.

We can rationally accept SUB-CONCLUSION without rejecting theory<sub>2</sub>. To do so, we just need sufficient justification for rejecting UNDEFEATED. This blocks moving to CONCLUSION and thereby blocks the unconditional obligation to revise. In fact, we'll see below that the failure of certain inner abductive arguments against our logical and mathematical beliefs to extend to outer arguments helps to insulate our logical and mathematical beliefs from certain challenges.

Our overall question is how seriously to take abductive arguments for revising our epistemic states when these arguments are self-effacing. That is, when the justification for revising would be undermined by adopting the very epistemic states recommended. I'll argue shortly that when an argument for revision of our standing beliefs is self-effacing, we cannot rationally use it to justify revising our standing beliefs—at least not directly. To see this, we need to get clear on what self-effacement is.

## 1.1 Self-Effacement

So what exactly is self-effacement? It's when an argument directs us to cease believing or otherwise reject things which are part of its support. To make this precise, we'll define it with respect to the argument schema given above (trusting the reader to make the necessary changes for other cases.) An argument  $\sigma$  of that form is *self-effacing* when the following two conditions are satisfied:

<sup>11</sup> The issue is what counts as a good reason to resist revising (2018b, §3.1). Unsurprisingly, the usual targets of these kinds of challenges, non-naturalist realists, tend to be much more epistemically permissive than anti-realists, error theorists, and naturalist realists. But see Barkhausen 2016 for arguments that naturalists aren't in a much better position than non-naturalists.

<sup>12</sup> Slightly different debunking arguments are found in Street 2006 and Joyce 2001. These can also be fit to our argument schema—which is a good exercise for the interested reader. See Schafer 2010, Vavova 2014, Lutz 2018, and Isserow Forthcoming for useful discussion of Street-Joyce-style evolutionary debunking.

- We have sufficient support for  $\sigma$ 's premises preceding the horizontal line.<sup>13</sup>
- Revising our beliefs about  $A$  in accord with  $\sigma$ 's directive<sup>14</sup> would undermine our *actual* support for at least one of  $\sigma$ 's premises.

This definition allows that instances of the “inner” argument, stopping at SUB-CONCLUSION, can be self-effacing. This is because accepting SUB-CONCLUSION amounts to accepting that our actual theory of  $A$  says of some theory  $T$  of  $A$  that it's not the best. When  $T$  is *our* theory, then our theory of  $A$  says we've significant reason to reject itself and thereby some of our content premises. And this is the case regardless of whether we can move forward to CONCLUSION.<sup>15</sup> See also [section 1.5](#).

There are three ways for self-effacement to arise:

- Revising our  $A$ -theory in line with  $\sigma$ 's directive would undermine our justification for CONTENT $_j$ , for some  $j$ .
- Revising our  $A$ -theory in line with  $\sigma$ 's directive would undermine our justification for STRUCTURAL or our understanding of one of the theoretical virtues used in CONTENT $_j$ , for some  $j$ .<sup>16</sup>
- Revising our  $A$ -theory in line with  $\sigma$ 's directive would undermine our justification for THEORY CHOICE.

We ignore the third way; it happens, but it's rare and it's complicated.<sup>17</sup> We'll call self-effacement arising the first way *content self-effacement* and self-effacement arising the second way *structural self-effacement*. We'll now give examples of each, working with the case of abductive reasoning about logical theories.

---

<sup>13</sup> We could give a definition of self-effacement that didn't presume we already had support for the premises. This would require modifying the second condition, but in obvious ways. I didn't do this because the most interesting cases of self-effacement are categorical arguments where we have justification for the premises. Hypothetical self-effacement is also far too easily confused with other phenomena. See [section 1.6](#).

<sup>14</sup> That is, coming to believe theory $_1$  and rejecting theory $_2$ , as suggested by SUB-CONCLUSION and demanded by CONCLUSION.

<sup>15</sup> I suppress any argument that if we ought to reject some view, then we've got significant reason to reject it. Whatever view of reasons and ought you have, some such connection seems obvious. I'll leave it to the reader to situate my points in their favored framework.

<sup>16</sup> It wouldn't be unreasonable to divide this type of self-effacement into two, even though both have to do with the mechanics of abductive theory choice. I won't do so here for reasons of simplicity.

<sup>17</sup> So I won't discuss revising our commitment to the evaluative normativity expressed by the 'ought' in 'you ought to believe such and so.' But to put my cards on the table: such a notion of theoretical obligation is a fundamental part of any cognitive project and thereby immune from many challenges. There, I said it. See also [Woods 2018b](#), §5.

## 1.2 Logical Self-Effacement

These examples are driven by the proof-theoretic weakness of some deviant logics. Consider transitivity, the property that a logic has when a proof of  $\psi$  from  $\phi$  and a proof of  $\phi$  from  $\Delta$  guarantees that there is a proof of  $\psi$  from  $\Delta$ .<sup>18</sup> Some logics do not have this property. This is problematic since ordinary mathematical proof presumes transitivity and, in fact, it can be quite difficult to see how to avoid making use of transitivity *even when* it can be avoided in principle. Showing that we can carry out non-trivial mathematical proofs in non-transitive logics is thus essential to an abductive case for adopting a non-transitive logic.

We can show that non-transitive logics<sup>19</sup> can prove claims of ordinary mathematics. Non-transitive logicians gesture informally toward this result when making their case. Our argument, though, makes use of transitivity. As transitivity isn't a basic feature of non-transitive logics, gesturing toward our argument is clearly insufficient. Of course, as this "recapture" result is a bit of ordinary mathematics, if we're right (we are), then there will be a non-transitive proof. Yet this isn't the kind of thing the *non-transitive logician* can take for granted given where they start; they need a rigorous proof of recapture, not a hand-wavy bootstrap.<sup>20</sup>

There are potential ways of defending a bootstrapping approach, but they either require special pleading or are currently underdeveloped.<sup>21</sup> We'll follow standard justificatory practice in mathematical logic and the foundations of mathematics when developing our examples; we'll presume that we're only entitled to claim logical and mathematical results when we can justify them rigorously using logical or mathematical resources we actually accept (or rigorously justify that we can justify them, etc.). Standard justificatory practice, combined with the problems with the problems just mentioned, yield the examples we're after:

- **Content Self-Effacement:** Suppose the target phenomena, our  $A$ , is logical consequence. Let theory<sub>2</sub> be our current transitive theory of logical consequence and let theory<sub>1</sub> be a weak, non-transitive logic. Changing which theory we accept can undermine our justification for claims about the virtues of these theories; this is because the theory we accept constrains what we can legitimately prove. In particular, changing which theory we accept can undermine our justification for our content premises.

<sup>18</sup> There are various ways of formulating these properties precisely. I'll speak rather loosely here as I've dotted my  $i$ 's and crossed my  $t$ 's elsewhere (Forthcoming *b*).

<sup>19</sup> See Tennant 2017 for the most rigorous version of such a logic and Cobreros et al. 2012 for an interesting variation.

<sup>20</sup> This objection dates back to Burgess (2005). Tennant 2017, 12.4 is his current rejoinder. See Woods Forthcoming *b* for why I think Tennant's reply is still insufficient.

<sup>21</sup> See Woods Forthcoming *a*, Forthcoming *b* for discussion. See also Meadows 2015 for useful criticism of similar bootstrapping approaches to non-classical set theory.

For example, it's a minimal virtue of a theory of logical consequence that standard mathematical results are entailed by standard mathematical starting points. Showing what a particular theory entails, though, is often highly non-trivial as it involves *generalizations* about what we can prove from various starting points; in short, a theory of proof.<sup>22</sup> Demonstrating, using a theory of proof, that non-trivial mathematical starting points have their usual known consequences depends on having a rather strong proof theory. Yet the only sufficiently strong extant proof theories for non-transitive logics are developed in classical or intuitionistic logic—logics which are fully transitive.

We *can* demonstrate that theory<sub>1</sub>, a non-transitive logic, can recapture ordinary mathematics when we use theory<sub>2</sub>, a transitive logic, to evaluate what theory<sub>1</sub> can prove. This underwrites a strong *prima facie* case for accepting theory<sub>1</sub>. But rejecting theory<sub>2</sub> as a consequence undermines our justification for accepting CONCLUSION. Why? Because using the transitivity of consequence, as we can when we accept theory<sub>2</sub>, is essential to our way of showing that theory<sub>1</sub> recaptures ordinary mathematical reasoning. Changing from theory<sub>2</sub> to theory<sub>1</sub> on the basis of this fact would thus undermine our justification for the claim that theory<sub>1</sub> had the virtue of recapturing ordinary mathematics; *that* argument made use of transitivity. So we'd lose our justification for one of our content premises. So, such an argument, given the facts supposed, would be content self-effacing.

- **Structural Self-Effacement:** Now let theory<sub>1</sub> be a logical theory so weak that we can't cleanly prove basic facts of elementary arithmetic while using it. Forget about the content premises and focus on the relevant instance of STRUCTURAL. Justifying STRUCTURAL requires assigning weights to theoretical virtues like 'being able to recapture ordinary mathematical reasoning,' aggregating various virtues and vices, then weighing it all out. If we hold that a theoretically virtuous explanation of *A* is a minimally plausible explanation of *A*, as we should, we also need to justify the minimal plausibility of theory<sub>1</sub>.

All of this—comparing weights of virtues, both explicating and justifying minimal plausibility, and so on—requires a fragment of mathematics and logic at *least* strong enough to recapture elementary arithmetic. When theory<sub>1</sub> is too weak to prove elementary arithmetical facts, we're not entitled to make elementary arithmetical comparisons. So, if we adopted theory<sub>1</sub> on the directive of an

---

<sup>22</sup> There are alternative methods, but none that are available to the standard non-classical logician given other weaknesses of their view. Focus on the proof-theoretic approach.

argument like the above, we'd lose our justification for STRUCTURAL. Voila, structural self-effacement.<sup>23</sup>

These cases establish that both kinds of self-effacement can happen. Of course, comparing the theoretical virtues of alternative logics with the vices of our own is a rather special case. We'll shortly look at more humdrum ones.<sup>24</sup> But we'll first briefly discuss the badness of self-effacement and put to rest a potential confusion.

### 1.3 What's Wrong with Self-Effacement?

There's surely something problematic with justifying a conclusion by means of premises which would be undermined by accepting it. This is especially true for the purpose we've fixed: generating a justification *we can use* in justifying a revision of our current beliefs (see [section 1](#)). But I shouldn't just thump the table and claim that:

BASIC FACT: It's irrational to revise our beliefs<sup>25</sup> on the basis of conclusions where so revising would destroy the justification for that conclusion.<sup>26</sup>

So here's a quick and dirty argument for BASIC FACT. Imagine how you'd justify your epistemic action to someone *after* acting on the basis of the conclusion of a self-effacing argument. You'd have to cite support to which you are no longer entitled, but you can't do that. Maybe you could instead cite support which is only now available from your new perspective. But even then you would have to view your prior *actual* justification, and thus your actual move from old to new perspective, as irrational. This situation seems epistemically vicious.<sup>27</sup>

If self-effacement undermines rational retrospective endorsement of our epistemic action (holding fixed our actual reasons at the time of revision), we should treat revising using a self-effacing argument as irrational. After all, in the cases we're imagining, we can easily see that even if we're entitled

---

<sup>23</sup> For readers who doubt that some relatively common non-classical logics are this weak, I invite them to try to prove something non-trivial in elementary number theory using only a weak non-classical logic. This means proving the result *without* invoking recapture results or, if they're invoked, proving recapture as well while using only that non-classical logic.

<sup>24</sup> For excruciating details of the cases both described here and to be described below, see again [Woods 2018b](#), [Forthcoming a](#), [Forthcoming b](#).

<sup>25</sup> I mean revision to be implicitly restricted to the right sorts of updating methods; I'm thus not worried about fancy "belief pill" cases.

<sup>26</sup> In our cases, we destroy the justification for the premises which entail the conclusion, but since justification is transitive in at least this case, BASIC FACT rules out rationally revising on the basis of self-effacing arguments.

<sup>27</sup> There could be independent reasons to believe a theory which undermines our current justificatory standpoint; perhaps it's elegant, perhaps it's intrinsically intuitive, perhaps not doing so means putting up with a self-effacing theory and that's too high a cost to bear in a particular case (see [section 1.6](#) and [section 3](#)). The point here is that we can't justify moving to that theory by using the undermined parts of our actual justificatory standpoint.

to the premises of our argument, we won't be after accepting and acting on the conclusion. But it seems bad to doom ourselves to irrationality; it seems reasonable to treat such epistemic actions as irrational.<sup>28</sup>

So, on the basis of an intuitive "reflection" principle for belief revision, it's irrational to revise on the basis of self-effacing arguments since we can see in advance that doing so would undermine our justification for so revising. This *does not* mean that we're not currently entitled to the premises of a self-effacing argument; we sometimes are entitled to them, especially the ones preceding SUB-CONCLUSION. Rather, we're not justified in acting on their consequences when that amounts to revising away our support. Since accepting CONCLUSION without revising or doubting our actual beliefs would be paradigmatically irrational, we're not rationally warranted in accepting it, even though we have an otherwise good argument for it.<sup>29</sup>

Again, after irrationally revising, we may find *new* reasons to have acted as we have, but even then we can't justify why we acted as we did, although we *can* justify why we should have. So there's good reason to avoid accepting and acting on conclusions when so doing would undermine our reasons for accepting and acting on them. In particular, the presumptive truth of the materials necessary for our abductive investigations isn't open to challenge by means of these methods since this would undermine the project of using abductive methods at all. This is the basis of how to block these challenges; it's the basis of the self-effacement gambit.

#### 1.4 The Self-Effacement Gambit

Start with Wright's (2004) discussion of warrant-transmission and "pre-suppositions of cognitive project." Warrant-transmission was introduced to diagnose what's wrong with Moorean anti-skeptical arguments. Wright argues, contra Moore, that we can't gain justification or warrant for the belief that there's an external world from the output of the cognitive project of using perception to limn the world. This is because the existence of an external world is presupposed by the cognitive project of using perception

<sup>28</sup> In [Woods Forthcoming a](#) and [Forthcoming b](#), I spotted the logical anti-exceptionalist that we can rationally revise from within a logical system even if our reasons for doing so essentially depend on our current logical theory. I was more generous in those essays since we can easily construct severe problems for revising logic abductively *even on this assumption*. Rewriting the current argument in a generous way can be done, although it would involve significant complications. I'll leave it to someone more generous than me to do so.

<sup>29</sup> Contentiously, some of these arguments could be sound (which raises problems for accounts of good reasoning like that of McHugh and Way [2018]). The key question is how to evaluate UNDEFEATED. In [Woods 2018b](#), I suggested that the self-effacement of debunking arguments against logic and mathematics was so costly that we had a conclusive reason against revising our beliefs (blocking the move from SUB-CONCLUSION to CONCLUSION in that instance). This could be resisted by someone less conservative about justification. Either way, my earlier discussion conflated this move with the related one involving structural self-effacement I develop later.

to limn the world. Given where that project starts, *that the world exists* is outside its remit; presupposing that it does is required to make sense of perceptual warrant.

We can sensibly invert Wright's point; it's plausible that we can't lose warrant or justification for a theory *T* *directly* (though see [section 1.6](#)) from the fact that challenges to *T* can be justified *using T* or when engaged in *T*-presupposing cognitive projects. This is particularly acute in the case of *T*-presupposing abductive investigations into the best systematic account of our *T* beliefs. Given what this project is and where it starts, systematic doubt about our *T* beliefs is outside its remit.

So this is our simple anti-skeptical strategy. When an argument is self-effacing, as many skeptical arguments turn out to be, then we're not rationally entitled to draw their conclusion because so doing would undermine our justification for their premises (by BASIC FACT.) Shortly, we'll see that this simple defense against self-effacing challenges doesn't go far enough, but it does allow us to shrug off certain *particular* arguments. That's not nothing.

Some will worry that if we can argue from within *T* that we'd be better off with a different account of *T*, then even if our argument for this depends on our current account of *T*, we should nevertheless revise that account.<sup>30</sup> I think this worry comes from two mistakes.

One mistake is not distinguishing whether we can rationally revise a view in line with the conclusion of a *particular* self-effacing argument from whether we can revise this view in line with the conclusion of a related non-self-effacing argument. We can often rejigger self-effacing arguments to avoid self-effacement. For instance, the most natural arguments against particular logical principles often make use of those very principles. We can usually modify these arguments so that they proceed via distinct logical principles at the cost of some complexity. See [section 1.6](#) for another example of extracting a cogent argument against a particular subject matter from a self-effacing one.

The other mistake is a bit more impressionistic, but pervasive. There's a tendency to see arguments "dynamically," treating ourselves as moving from premises we already accept to a conclusion we now do in a stepwise fashion.<sup>31</sup> This makes it seem as if we're moving stepwise from premises to SUB-CONCLUSION to CONCLUSION in a warrant-preserving process of belief updating.

Actual belief revision doesn't work this way ([Harman 1986](#)). When we act on CONCLUSION, treating it as rational, we're actually accepting that accepting both the argument's premises and CONCLUSION is cogent (among

---

<sup>30</sup> Perhaps these handwringers will mention the famed ladder we kick away. But it's hard to see how ladder-kicking in this case is rational, whatever other virtues it might have.

<sup>31</sup> I suspect overfamiliarity with stepwise deductive calculi conjoined with slightly less familiarity with actual mathematical proof is a large part of this mistake.

other things), *then* updating our epistemic state accordingly. A quick inspection shows that it's paradigmatically irrational to accept all the premises of a self-effacing argument, along with their support, while accepting its conclusion; it's tantamount to accepting ' $p$  and  $p$  isn't justified.'<sup>32</sup>

### 1.5 Problems with the Self-Effacement Gambit

We've been focusing on the "outer" argument. Turn now to the "inner" argument. We can easily accept its premises without revising our view about some subject matter as there may be significant reason to avoid revising our view. In other words, we can accept the SUB-CONCLUSION, deny UNDEFEATED, and thereby avoid drawing CONCLUSION. Is this a comfortable resting place?

No. It's already problematic to accept SUB-CONCLUSION without drawing CONCLUSION. Accepting SUB-CONCLUSION means holding that there's significant reason to revise our beliefs about some subject matter or phenomena. Moreover, it means accepting that these reasons are demonstrable from within our current view of that subject matter or phenomena.

This is an uncomfortable position to be in even when we don't actually revise our beliefs; we're accepting that our beliefs aren't entirely epistemically cogent from within. Even if we accept that revising would be severely problematic, we shouldn't be happy that our beliefs about some subject matter or phenomena are so down on themselves.

It helps here to distinguish between when a view or theory is self-effacing and when a particular argument is self-effacing.<sup>33</sup> Say that a view or theory is self-effacing when there are self-effacing arguments whose premises are justified by that view or theory which direct us to revise that very view or theory (some of this definition is redundant). Accepting a particular self-effacing argument to SUB-CONCLUSION commits us to treating our standing theory of  $A$  as self-effacing. It's *this* which is the significant theoretical vice of  $A$ .

Summing up, there's a strong case to be made that revising our views in line with the conclusion of a self-effacing argument is irrational. It seems to violate an intuitive reflective principle for justification as well as being tantamount to simultaneously accepting  $p$  and that  $p$  is unjustified. This observation generates a strategy for blocking a certain challenges to our standing views: the self-effacement gambit.

Yet there are sound arguments whose directive (note 'directive,' not 'conclusion') is revision of our standing beliefs about certain subject matters

---

<sup>32</sup> Strictly speaking, there's a few moves between accepting that we ought to revise our beliefs about something and being rationally committed to the consequences of this acceptance for particular beliefs. I've suppressed these moves since the extremely modest amount of closure required here seems unproblematic. I hope to explore elsewhere the general relationship of self-effacement to both Moore's Paradox and other cases of pragmatic incoherence.

<sup>33</sup> Thanks to Daniel Wodak for encouraging me to distinguish these explicitly.

and phenomena. This very fact witnesses problems with our accounts of various subject matters and phenomena; it underwrites reasons to revise, even if these reasons can be outweighed.

It's unclear how strong these reasons are. In what follows, I'll suggest that although we plausibly find such reasons in the context of logic and mathematics, they're not quite strong enough prop up challenges to our logical and mathematical views (at least a fragment thereof, at least when we take abduction as our primary methodology). I'll also suggest that when we turn to morality and aesthetics, the analogue reasons might be strong enough to underwrite challenges. But, in order to see this difference, we need to do a bit more clarification.

### 1.6 What Self-Effacement Is Not

Self-effacing arguments are easily confused with forms of inference like *reductio ad absurdum*. On the basis of this confusion, some have argued that there's no problem with self-effacing arguments as there's no problem with *reductio ad absurdum*. But this really is a confusion.

I suspect that this mistake comes from failing to cleanly distinguish *categorical* from *hypothetical* reasoning. Categorical reasoning involves moving from claims we accept to claim which are supported by the claims we accept. Hypothetical reasoning involves reasoning *under a hypothesis* to establish facts about what follows from that hypothesis. Sometimes we mix the two together, such as when we infer that some proposition is false on the basis of the fact that we can deduce an absurdity from it. This inference involves *categorically* using a fact established by *hypothetical* reasoning—that  $\varphi$  yields absurdity—to conclude that  $\varphi$  is false.

There are ways of representing proof rules like *reductio* where it seems that we reject a premise on the basis of a conclusion we categorically reach on the basis of that premise, but this is misleading. The most representative form of an actual argument by *reductio ad absurdum* isn't:

$$\begin{array}{l} \varphi \\ \psi \\ \hline \chi \wedge \neg\chi \end{array}$$

where we reject either  $\varphi$  or  $\psi$  on the basis of the fact that they jointly entail a contradiction. The most representative form is rather:

$$\begin{array}{l} [\varphi, \psi] \Rightarrow (\chi \wedge \neg\chi) \\ \hline \neg\varphi \vee \neg\psi \end{array}$$

where  $\Rightarrow$  represents deducibility under the hypotheses within the square brackets. Neither  $\varphi$  nor  $\psi$  function as a premise in our reasoning by

reductio; they are not the *categorical* consideration mustered in favor of our conclusion. Rather, the categorical consideration used is the *deducibility of a contradiction from  $\varphi$  and  $\psi$* .

Yet there's a grain of truth in most tempting objections, including this one. It's vicious for a theory to be self-effacing—to say of itself that it's not entirely correct.<sup>34</sup> This is true *even when* we can resist revising our theory because the otherwise good argument that it's not entirely correct is self-effacing. Although the analogy between reductio arguments and self-effacing arguments is incorrect, that a theory generates self-effacing challenges to itself can play a role in *different* categorical arguments for revising our views. This is rather analogous to how the deducibility premise plays a role in reductio proofs and therefore the temptation to run the two together isn't entirely surprising.

We can use the viciousness of self-effacing views as part of an abductive argument against those views. These new arguments aren't typically self-effacing since the demonstration of self-effacement doesn't usually rest on any categorical use of content from the views criticized. Instead, they're arguments built on *top* of a self-effacing argument.<sup>35</sup> I'll show shortly that how problematic such arguments are depends on what would happen if we acted on them. For now, though, it's enough to recognize that revising one of our views in line with the directive of a self-effacing argument is rationally problematic even though the cogency of that self-effacing argument is itself a theoretical vice of our view.

## 2 Locating Self-Effacement

We've focused so far on examples drawn from the theoretical side of rationality, in particular from logic. More practical considerations might also play a role in whether a particular theory or explanation counts as best. Perhaps an explanation of my moral beliefs which has it that it's moral for me to have my moral beliefs is better than one which treats having those beliefs as immoral. Perhaps more beautiful theories are just intrinsically better.

If moral or aesthetic considerations play a role in determining explanatory goodness, then challenges to our views of morality and aesthetics can also be self-effacing. We've given examples of self-effacing challenges to our logical and mathematical views. I'll do the same now for morality and aesthetics, showing how content self-effacement can occur for challenges to our aesthetic and moral views. These challenges can be blocked by the self-effacement gambit. Unfortunately, we cannot so easily dispense with related challenges, as I'll show shortly.

<sup>34</sup> Thanks to Josh Schechter for discussion of this and related points. See his 2018 for further useful remarks.

<sup>35</sup> There may be cases where any such argument would itself be self-effacing. Set such cases aside.

## 2.1 Moral Self-Effacement

While it's plausible that moral facts play a role in the best pragmatic explanation of our moral beliefs—I believe Roy Cohn was a monster because he was, in fact, a monster—it's controversial whether moral facts play a role in the best non-pragmatic explanation of our moral beliefs.<sup>36</sup> It's tempting to say that only physically realized phenomena like our *beliefs* about morality, the testimony of others about morality, and the like could explain our moral beliefs (Harman 1977).<sup>37</sup> For instance, one standard defense of abductive methodology claims that the best explanation is just the one most likely to be true. On many metaethical views, what makes psychological facts true is independent from what makes moral facts true. So how could the latter play a role in explaining the former?

On the other hand, aesthetic considerations like naturalness and elegance play a clear and plausibly non-instrumental role as theoretical virtues. This is obvious for mathematical explanation (Hardy 1967), but it's easy to find instances elsewhere. So aesthetics seems to play a role in certifying certain explanations as the best. If aesthetic considerations play such a role, why not moral considerations?<sup>38</sup>

Some philosophers have argued, in particular, that the morality of a moral theory counts as evidence for its truth:

This is a curious type of argument, for it has the form that *P* is true because it would be better if it were true . . . That is not in general a cogent form of argument . . . it may have a place in ethical theory, where its conclusion is not factual but moral. . . . It may be suitable to argue that one morality is more likely to be true than another, because the former makes for a better world than the latter—not instrumentally, but intrinsically. (Nagel 1995, 92)

There being a difference between one's theory of the best normative X (the best morality, the best standards of inference, the best rules of justification . . .) and one's (so far) best theory of X, necessarily provides a reason (though perhaps not a decisive reason) to think one's (so far) best theory is wrong. (Sayre-McCord Unpublished, 1)

---

<sup>36</sup> See Sturgeon 1986 for discussion.

<sup>37</sup> Paulina Sliwa suggests (personal communication) that norms of truthfulness and trust like those undergirding testimony are an essential part of the explanation of our beliefs. She goes on to suggest that they thereby play a non-pragmatic role in explanations of our moral beliefs. Since the entanglement of these norms in our abductive methodology is quite deep, it would be a rather important case of potential self-effacement if the relevant norms were moral ones. I don't think the relevant kinds of truthfulness and trust are moral notions, but I don't have space to address the issue here.

<sup>38</sup> See Hanson 2018 for one set of reasons to treat these two subject matters analogously. See Woods 2018a for another set of reasons.

The view Nagel is flirting with is stronger than Sayre-McCord's. Sayre-McCord permits morality's explanatory work to be instrumental. That is, his view permits immorality to be necessarily indicative of falsity, but not *why* the immoral view of morality is false. Nagel's view, as I understand it, eliminates this possibility. It's *because* it makes for a better world than alternatives that the true moral view is true.

To illustrate this difference, consider "third factor" views like Enoch's (2011) where moral beliefs and moral facts are explanatorily independent, both being instead explained by a common *explanans*. There's then a necessary connection between our beliefs about morality and the moral facts since they're explained by the same thing. This is consistent with Sayre-McCord's claim, but not Nagel's. It's not the fact that it's more moral that *makes* it the case that it's more likely true, but whatever the third factor is.

So there are strong and weak ways for morality to contribute to explanatory goodness.<sup>39</sup> It's even possible to avoid routing through the likeliness of a theory or explanation entirely, letting the morality of a view be an intrinsic part of its explanation. If we accept any of this, then the moral or immoral character of a moral theory can make a difference to what we ought to believe about morality.

The weak account is enough to generate self-effacement. Even if morality (or aesthetics) plays only an instrumental role in justifying our moral (or aesthetic) views, the link between the moral character of our moral views and their likeliness can be undermined when coming to significantly revise our moral (or aesthetic) views. That's enough to produce cases of self-effacement, just as our inability to *see* or *prove* that a certain weak logic had certain undesirable properties was sufficient to generate logical self-effacement. However, since I'm tempted to think that morality and aesthetics can play a more intrinsic role in making an account best, I'll focus on the stronger account.

Suppose that the second-best purely theoretical explanation of morality is that it's an effective fiction imposed by the ruling classes. That is, this Hobbesian view is the second-best explanation of morality once we *ignore any moral or practical virtues* or their effect on the plausibility of an explanation. Suppose, *per impossibile*, that some version of stark-raving (non-naturalist) moral realism is a slightly better purely theoretical explanation.

Now suppose further that the Hobbesian view is a morally superior explanation of morality than realism. Perhaps it's morally better if we aren't beholden to mind-independent facts about how we ought to behave. Maybe this is because it's morally best to have an explanation of morality

---

<sup>39</sup> There's a litany of other defenders, such as Enoch (2009), Preston-Roedder (2014), and Quinn (1994). See Sayre-McCord Unpublished for detailed and compelling defense of the argumentative strategy.

that allows that we morally shouldn't be so slavish to moral concerns.<sup>40</sup> Or perhaps the type of moral realism in question requires us believe, on pain of immorality, that the immoral are intrinsically inferior; clearly that would be a pretty morally pernicious account of morality. Whatever the reason, if the Hobbesian explanation is sufficiently morally superior to believing in some kind of moral realism, it will unseat moral realism as the best *overall* explanation of our moral beliefs.

We can then argue for rejecting moral realism on the grounds that we ought to believe the best explanation of morality. Yet the best explanation of morality, the Hobbesian one, entails that our judgments about morality are systematically unjustified; they're actually capitalist fictions.

Since coming to reject our actual moral beliefs would immediately undermine our justification for this rejection—since part of why the Hobbesian view was the best overall explanation was its *moral* character—we have a case of content-based self-effacement. The example is merely illustrative; we could argue like this in several different ways.<sup>41</sup> It's enough here to demonstrate plausible ways such an argument could go.

Morality can only be self-effacing in the content way though—once we've settled the virtues of a good explanation or theory, there's no further *moral* question which arises in explicating these virtues or aggregating them. We'll draw on this below to argue that there's still a determinate sense in which morality is more vulnerable to the kind of challenges we're concerned with, even though it permits self-effacement. But first aesthetics.

## 2.2 Aesthetic Self-Effacement

Whether or not morality partially determines the quality of some explanations, aesthetics surely plays such a role. *Elegance*, *naturalness*, and *non-ad-hoc-ness* are often treated as important virtues of explanations and theories.<sup>42</sup> This is most glaring in mathematics as noted by Russell (1984), but it's also clear on inspection that we use these virtues in nearly all actual applications of abductive machinery.

Similarly to moral considerations, aesthetic considerations might be relevant because they evidence the explanation most likely to be true. Perhaps the world is more beautiful than it appears. Perhaps it won't tolerate ugly explanations or theories—the beautiful can be really horrid that way. We need not treat aesthetic considerations as so flatly instrumental though.

<sup>40</sup> See Schafer 2016 for an account of Hume which has this rather appealing upshot.

<sup>41</sup> We could use claims about the essential role of agency in our cognitive projects in place of the Nagel or Sayre-McCord strategy. We could use of Hayward's (Forthcoming) recent defense of the immorality of moral realism. And so on.

<sup>42</sup> See Lipton 2004, 66 where these virtues are explicitly labeled aesthetic.

Lipton (2004) treats ‘loveliness,’ explicated in terms of ability to convey understanding, as a primary theoretical virtue, albeit one connected non-trivially to likelihood. This is a view on which aesthetic virtues play an important indirect role in justifying that some explanation is the best. He goes on to cite the ability to capture the aesthetic virtues of theories in terms of loveliness as part of the attraction of ‘inference to the loveliest explanation.’

Non-instrumental views are also possible. We might just flat out treat being pretty and natural as part of what it is to be the best explanation of some phenomenon. So, just as with morality, we can formulate stronger and weaker roles for aesthetics to play. Each role seems to be *prima facie* reasonable and familiar from our actual abductive practice.

The virtuousness of aesthetically pleasing explanations is more independent of subject matter than the virtuousness of morally pleasing explanations. While a morally better explanation of morality is perhaps a better moral explanation, a morally better explanation of some physical phenomenon is by no means a better physical explanation. Besides lacking the pithy expression, it simply seems irrelevant to the explanatory goodness of a physical explanation or the virtues of a physical theory whether or not it conforms to our moral scruples.<sup>43</sup>

Of course, not all of aesthetics plays a role in explanation. It’s not germane to the quality of an explanation of a particular mathematical theorem whether that explanation is funny. But it seems hard to deny that theoretically-oriented aesthetic virtues are always or nearly always relevant; it’s a constant source of interest in science to show that a particular explanation is not only plausible, but also elegant, simple, and natural. So at least the theoretically-oriented aesthetic virtues listed above play a role in abduction regardless of the subject area of the explanandum.

Aesthetic properties also only give rise to cases of content self-effacement (we’ll not construct such cases here; just follow the recipe above.) After all the virtues are in and weighed, there’s no further important aesthetic question about which theory wins out.<sup>44</sup> It’s also unlikely that aesthetic facts play a significant role in explicating particular virtues, aesthetic or no, in the way that mathematics and logic explicate notions of prediction, commitment, and probability. So aesthetic self-effacement seems restricted to content self-effacement.

---

<sup>43</sup> What we should all things considered *believe* about physics is perhaps a different story, but even here, it’s *prima facie* implausible that morality will play a *significant* role as it might for our moral beliefs. See Maguire and Woods Unpublished for a survey of cases.

<sup>44</sup> There’s an aesthetic question for any actual calculation. We shouldn’t bobble around when we can calculate cleanly and efficiently, but that’s orthogonal.

### 2.3 Logical and Mathematical Self-Effacement

We've already seen self-effacing arguments against our logical and mathematical theories. The reader won't have trouble generating more on their own since mathematics and logic play an especially entangled role in our abductive reasoning: both justifying and making any sense of typical theoretical virtues involves making use of something like a theory of entailment and a minimal fragment of arithmetic. Our focus in this section is thus on structural self-effacement.<sup>45</sup>

Consider theoretical strength. This virtue<sup>46</sup> can—in fact, must—be explicated in terms of what a particular theory predicts. And prediction is at least largely explicated in terms of what a theory entails. Theoretical simplicity needs a notion of complexity which is typically developed in a fragment of arithmetic. Ontological simplicity needs a notion of cardinal comparison which itself requires a significant chunk of either mathematics or logic.<sup>47</sup>

Even more fundamentally, the aggregation implicit in STRUCTURAL needs a background where we assign relative weights to virtues, and then compare them. This involves a minimal theory of aggregation, which isn't entirely trivial.<sup>48</sup> Finally, in order for an explanation or theory to be best in the relevant sense for abductive justification, it shouldn't just be the best of the worst. It should also be minimally plausible. This seems to require the notion of being minimally probable. Which, of course, is theorized mathematically. So, unlike in aesthetics and morality, we can have *both* content and structural self-effacing challenges to our logical and mathematical beliefs. This is what makes the case of abductive challenges to logic and mathematics distinctive (though see [footnote 37](#).)

We don't need all of mathematics and logic to employ abductive methodology. We don't even need much mathematics and logic. What we need is a minimal fragment of theoretical power from some combination of mathematics and logic which we can use to compare various explanations.<sup>49</sup> Let me explain.

---

<sup>45</sup> Content self-effacing challenges to our logical and mathematical views isn't especially different from the case of aesthetics and morality.

<sup>46</sup> Though Russell ([Forthcoming](#)) suggests that theoretical strength isn't clearly a theoretical virtue.

<sup>47</sup> Finite cardinal comparisons can be done in an insignificant chunk of logic. Interesting theories are rarely finite.

<sup>48</sup> I'm skipping the justification of aggregation functions since we're now flogging a dead horse. Adding this in would add yet another place where appeal to a minimal fragment of mathematics and logic is necessary.

<sup>49</sup> For the cognoscenti, what we need is a theory which contains something like a theory of inductive definitions, the ability to do arbitrary cardinal comparisons, etc. It would be shocking if our needs exceeded the logical strength of a very weak subsystem of arithmetic, such as  $WKL_0$ .

We have a choice when developing our account of these background notions. We can work in a small fragment of mathematics, such as a fragment of arithmetic, or we can work in a suitably strong background logical theory. For example, we could situate the theory of cardinal comparisons in a logic augmented with a binary quantifier  $M(X, Y)$  expressing that there are more  $X$ s than  $Y$ s. Alternatively, we could define the relevant notion, on minimal background assumptions, in second-order logic. We could also just define it in first-order set theory. Any of these methods and many others beside would suffice, although obviously each has virtues and vices. What we need is *some* system with the appropriate amount of theoretical strength. It doesn't matter so much what it is. Call this "disjunction" of mathematical and logical theories *the common core*.

Correspondingly, we get structural self-effacement of an interesting sort only when we formulate a challenge to the truth of *some minimal fragment or other*—to the common core. We cannot get a direct justification for holding onto any *particular* way of developing the common core since we can always use one development of the common core to run an abductive argument against another way (exercise for the reader).

We can't exploit structural self-effacement to defend any *particular* disjunct of the common core against reasonable challenges. What we can do, in the fashion gestured at above, is justify the common core itself since this is what we need to carry out abductive reasoning. In other words, structural self-effacement gives us reason to reject wholesale skepticism about the resources (but no particular instantiation of such) for carrying out abductive investigations. And this is a good thing! It would be beyond dogmatic to defend particular sets of logical or mathematical beliefs against reasonable objections on these kinds of grounds.<sup>50</sup>

Summing up, we need some minimal fragment of logic and mathematics to carry out the *machinery* of abductive reasoning. This common core of "theoretical power" is necessary to make sense of theoretical virtues and to justify premises like STRUCTURAL. Abandoning this common core amounts to abandoning abduction altogether. In contrast, we could purge our abductive reasoning of moral and aesthetic content without doing damage to the machinery of abduction, even if we'd do damage to the content premises germane to particular cases.

### 3 Types of Self-Effacement

We've now seen several different cases of self-effacement covering a number of different areas. We summarize the epistemically relevant differences as follows:

---

<sup>50</sup> There are cases of content self-effacement that allow the defense of our particular logical and mathematical beliefs. These are not difficult to construct using the recipes already given.

- Morality can be content self-effacing, but morality's role in abduction is context-sensitive. Abductive reasoning doesn't usually employ moral content, even if abductive reasoning about morality does. Consequently, we can abandon our moral views without abandoning a recognizable notion of abductive methodology, even one applicable to morality. So morality's role in determining which explanation or theory of morality is best is shallow.
- Aesthetics also permits content self-effacement, but aesthetics' role in abduction is fairly context independent. Still, we could get rid of our affection for pretty theories without enormous loss, even though the resulting abductive methodology would look pretty different from our actual abductive methodology.
- Logic and mathematics permit both content and structural self-effacement, and they play a context-independent role in abduction. This is because we require the common core of mathematics and logic to explicate and compare theoretical virtues for any application of abductive reasoning.<sup>51</sup> That is, logic and mathematics are required for any recognizable notion of abductive methodology as applied to any context.

### 3.1 Weighing the (Theoretical) Costs

These differences mean that the theoretical costs of acting on the directive of a self-effacing argument differ quite heavily between different types of self-effacement.<sup>52</sup> These costs are reckoned from our *current* standpoint; we're costing wholesale, not piecemeal, revision, rejection, or doubt of our mathematical, logical, aesthetic, or moral views given our current views.

In the case of moral self-effacement, the costs don't seem especially high. Although we might have to revise our actual abductive practice once we revise morality, our successor notion wouldn't be wildly far away from our current one. At worst, if we came to systematically doubt, reject, or massively revise our moral beliefs, we'd stop treating morality as a theoretical virtue of moral explanations and theories. But the rest of our

---

<sup>51</sup> Some particular arguments might involve materials which are context-dependent as well, but that's not important for the points below.

<sup>52</sup> Again, I'm focusing on the *theoretical* costs of wholesale revision of our moral, aesthetic, logical, and mathematical views. There are also *practical* costs which might need to be weighed for an account of what we all things considered should believe. Since it's a contentious issue how to weigh practical and theoretical costs and since we only need theoretical costs for the defense of logic and mathematics I'm going to offer shortly, I'll put these issues to the side. It's enough to see that there is a question for morality and aesthetics about how costly wholesale doubt would be—though, as many error theorists have shown, it's not nearly as high as people seem to think.

abductive methodology, and our actual abductive methodology in most cases, would be untouched.

Purging our abductive methodology of aesthetic virtues, in contrast, would amount to a quite dramatic overhaul of our abductive machinery. We use theoretically-oriented aesthetics constantly in evaluating explanations and theories. Junking them or significantly revising our aesthetic views would thus amount to an enormous change to our abductive practices. This seems rather costly. So weighing out the reasons for and against a challenge to our aesthetic beliefs, it's not clear what to do. Potentially, we might decide that this revision of our abductive methods isn't costly enough to outweigh the fact that our best abductive methods instruct us to wholesale revise, reject, or doubt our aesthetic beliefs.

Abandoning the common core of our logical and mathematical views would be catastrophic;<sup>53</sup> we cannot escape using some fragment of logic and mathematics when engaging in abductive reasoning. So, at least for the common core thereof that we've been interested in, it is clear that we've sufficient reason to resist revising, even when our logical and mathematical views are self-effacing.

Abductive challenges to all of our logical and mathematical beliefs threaten the common core. They thereby threaten our abductive practice generally, and thus threaten the general cognitive project of using abductive methods to justify our beliefs and views. This isn't true of aesthetics, much less morality. Consequently, we have obviously sufficient reason to resist revising for logic and mathematics, less for aesthetics, and even less for morality.

### 3.2 Persistent Worries about Morality and Aesthetics

Returning to our argument schema, this means that UNDEFEATED—that we don't have sufficient reason to resist revising—is most plausible for challenges to our moral views, less so for the challenges to our aesthetic views, and clearly implausible for the challenges to our logical and mathematical views. So we cannot extend SUB-CONCLUSION to CONCLUSION when challenging the common core of our mathematical and logical views although we might be able to do so when challenging our moral and aesthetic views. This is important because accepting CONCLUSION is extremely problematic, as I'll now show.

By BASIC FACT, we cannot rationally justify revising our views by following out a self-effacing challenge. Yet if our view about a subject matter endorses a self-effacing argument of the form above, it thereby says of itself that it's not the best view. When we can move from SUB-CONCLUSION to CONCLUSION, our view further says that we ought to revise it. We

<sup>53</sup> The case of content self-effacement for logic and mathematics is trickier. We put it to the side for now, noting that Woods 2018b, Forthcoming *a*, and Forthcoming *b* address closely related issues.

thereby ought to revise our view, by the lights of our view, even though doing so would be irrational (again, by BASIC FACT). The situation for our standing moral and aesthetic beliefs might thus be much worse than mere self-effacement.

When our current moral and aesthetic beliefs license such arguments and thereby commit us to irrationality, we have an *independent* reason to reject our moral and aesthetic beliefs.<sup>54</sup> This is an even stronger form of the challenge gestured at in section 1.6. Even without the strengthened form of the challenge, we may have sufficient reason to revise simply on the grounds that our moral and aesthetic beliefs are self-effacing and thereby vicious—that is, on the grounds that they permit a good argument to SUB-CONCLUSION and thereby say they’re not so hot.

As pointed out in section 1.6, these new challenges shouldn’t be conflated with the self-effacing arguments they’re built on. They’re rather different arguments against our standing moral and aesthetic beliefs which use as a premise that our moral and aesthetic beliefs permit self-effacement or that they commit us to irrationality by means of self-effacement. Note also that they’re not self-effacing. Neither self-effacement nor irrationality is a moral or aesthetic vice.

What should we do in the face of these challenges? We need to once again weigh the costs of revising our moral and aesthetic beliefs against the costs of sticking to our guns and tolerating good self-effacing arguments. It’s not immediately clear what the outcome of this weighing will be, although I suspect it looks pretty bad for morality and aesthetics. But these new challenges exist and need to be dealt with, whatever the ultimate outcome of the weighing turns out to be.

These new challenges cannot be run against our logical and mathematical beliefs. We have *pro tanto* reason to revise our logical and mathematical views, sure. But our case for revising is easily outweighed by the catastrophe of losing our ability to use abductive arguments at all. As we also cannot move from SUB-CONCLUSION to CONCLUSION in this case—given the catastrophe of revision—we don’t get the more vicious commitment to irrationality either. So we have sufficient reason to resist revising. This is the hull of the revised anti-skeptical argument I mentioned at the outset.

#### 4 Conclusion: The Revised Anti-Skeptical Gambit

Drawing together the preceding discussion, our revised “abductivized” self-effacement gambit goes as follows. For the common core of logical and mathematical beliefs, we cannot reason to CONCLUSION since there’s overwhelming reason to not act on SUB-CONCLUSION. This means that we’re ultimately just weighing *pro tanto* reasons to revise our faith in the

---

<sup>54</sup> Or, anyways, so it seems given how bad our epistemic situation would be in such a case.

common core against the catastrophe of abandoning abduction altogether. This is a calculation whose outcome is easy to see in advance.

The corresponding calculation for morality and aesthetics is much less clear; revising our moral and aesthetic views doesn't bring (theoretical) catastrophe in its wake. So we have no robust defense against challenges to morality and aesthetics arising from the existence of good self-effacing arguments, even though we can't use those self-effacing arguments themselves to reject morality and aesthetics directly.

Put slightly differently, it's a presupposition of engaging in a cognitive project at all (in the relevant sense) that enough of our background mathematical and logical beliefs are correct. Rejecting them is rejecting the possibility of such cognitive projects. This cost is far too high to be counterbalanced by the irritation of our logical and mathematical views saying that they're not the best views to have. We can thus fend off challenges built on this irritation. But our moral and aesthetic beliefs are not part of what's presupposed by engaging in any cognitive projects, so they're not similarly protected.

It would be both fascinating and surprising if, in contrast to what I've argued, aesthetics or morality could be shown to play a deeply entangled role in our abductive methodology—one, say, analogous to that played by mathematics and logic.<sup>55</sup> If morality or aesthetics were that deeply entangled, then a novel anti-skeptical defense is available against abductive challenges to them—just modify the revised self-effacement gambit. But if their involvement in our abductive methodology is shallow, as I suspect, then they are vulnerable to abductive challenges even when these challenges are self-effacing. Either way, the roles that morality, aesthetics, logic, and mathematics play in abductive reasoning reveals quite a bit about which of our views we are, and more importantly, *are not* in a position to challenge.

Jack Woods  
University of Leeds  
E-mail: [j.e.woods@gmail.com](mailto:j.e.woods@gmail.com)

---

<sup>55</sup> Though see [footnote 37](#).

**Acknowledgments** Thanks to Tim Button, Justin Clarke-Doane, Catarina Dutilh-Novaes, Billy Dunaway, Edward Elliot, Daniel Elstein, Graeme Forbes, Marcus Giaquinto, Jessica Isserow, Simon Kirchin, Mary Leng, Jimmy Lenman, Barry Maguire, Erum Naqvi, Huw Price, Mark van Roojen, Josh Schechter, Paulina Sliwa, Robert Stern, Mike Titelbaum, Robbie Williams, Daniel Wodak, Gözde Yıldırım, Claire Field, and Rob Knowles for helpful discussion of earlier versions. Thanks also to audiences at the Epistemology of Mathematics: Knowledge, Proof, and Explanation meeting at Leeds, the Maths and Morals meeting at Kent, the Cambridge Moral Sciences Club, the Sheffield Undergraduate conference, and the Saint Louis conference on Reasons and Rationality for useful discussion.

## References:

- Barkhausen, Max. 2016. "Reductionist Moral Realism and the Contingency of Moral Evolution." *Ethics* 126 (3): 662–689. <https://doi.org/10.1086/684708>.
- Burgess, John P. 2005. "No Requirement of Relevance." In *The Oxford Handbook of Mathematics and Logic*, edited by Steward Shapiro, 727–750. Oxford: Oxford University Press.
- Cobrerros, Pablo, Paul Egré, David Ripley, and Robert van Rooij. 2012. "Tolerant, Classical, Strict." *Journal of Philosophical Logic* 41 (2): 347–385. <https://doi.org/10.1007/s10992-010-9165-z>.
- Cohen, Gerald. 2000. *If You're an Egalitarian, How Come You're So Rich?* Cambridge, MA: Harvard University Press.
- Engel, Mylan. 1992. "Personal and Doxastic Justification in Epistemology." *Philosophical Studies* 67 (2): 133–150. <https://doi.org/10.1007/BF00373694>.
- Enoch, David. 2009. "Wouldn't It Be Nice If  $p$ , Therefore,  $p$  (for a moral  $p$ )." *Utilitas* 21 (2): 222–224. <https://doi.org/10.1017/S0953820809003495>.
- Enoch, David. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.
- Hanson, Louise. 2018. "Moral Realism, Aesthetic Realism, and the Asymmetry Claim." *Ethics* 129 (1): 39–69. <https://doi.org/10.1086/698732>.
- Hardy, G. H. 1967. *A Mathematician's Apology*. Cambridge: Cambridge University Press.
- Harman, Gilbert. 1977. *The Nature of Morality: An Introduction to Ethics*. Oxford: Oxford University Press.
- Harman, Gilbert. 1986. *Change in View: Principles of Reasoning*. Cambridge, MA: MIT Press.
- Hayward, Max Khan. Forthcoming. "Immoral Realism." *Philosophical Studies*. First Published Online 2019.
- Hjortland, Ole Thomassen. 2017. "Anti-exceptionalism about Logic." *Philosophical Studies* 174 (3): 631–658. <https://doi.org/10.1007/s11098-016-0701-8>.
- Isserow, Jessica. Forthcoming. "Evolutionary Hypotheses and Moral Skepticism." *Erkenntnis*. First Published Online 2018.
- Joyce, Richard. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.
- Lewis, David. 1983. *Philosophical Papers*. Oxford: Oxford University Press.
- Lipton, Peter. 2004. *Inference to the Best Explanation*. London: Routledge.
- Lutz, Matt. 2018. "What Makes Evolution a Defeater?" *Erkenntnis* 83 (6): 1105–1126. <https://doi.org/10.1007/s10670-017-9931-1>.
- Maguire, Barry and Jack Woods. Unpublished. "Explaining Epistemic Normativity."
- McHugh, Connor and Jonathan Way. 2018. "What Is Good Reasoning?" *Philosophy and Phenomenological Research* 96 (1): 153–174. <https://doi.org/10.1111/phpr.12299>.
- McPherson, Tristram. 2015. "The Methodological Irrelevance of Reflective Equilibrium." In *The Palgrave Handbook of Philosophical Methods*, edited by Christopher Daly, 652–674. London: Palgrave Macmillan.
- Meadows, Toby. 2015. "Unpicking Priest's Bootstraps." *Thought: A Journal of Philosophy* 4 (3): 181–188. <https://doi.org/10.1002/tht3.172>.
- Nagel, Thomas. 1995. "Personal Rights and Public Space." *Philosophy and Public Affairs* 24 (2): 83–107. <https://doi.org/10.1111/j.1088-4963.1995.tb00024.x>.
- Preston-Roedder, Ryan. 2014. "A Better World." *Philosophical Studies* 168 (3): 629–644. <https://doi.org/10.1007/s11098-013-0154-2>.
- Priest, Graham. 2006. *Doubt Truth to Be a Liar*. Oxford: Clarendon Press.
- Priest, Graham. 2016. "Logical Disputes and the A Priori." *Logique et Analyse* 59 (236): 347–366.
- Quinn, Warren. 1994. *Morality and Action*. Cambridge: Cambridge University Press.
- Russell, Bertrand. 1984. *Theory of Knowledge: The 1913 Manuscript*. Edited by Elizabeth Ramsden Eames and Kenneth Blackwell. London: George Allen & Unwin.

- Russell, Gillian. 2015. "The Justification of the Basic Laws of Logic." *Journal of Philosophical Logic* 44 (6): 793–803. <https://doi.org/10.1007/s10992-015-9360-z>.
- Russell, Gillian. Forthcoming. "Deviance and Vice: Strength as a Theoretical Virtue in the Epistemology of Logic." *Philosophy and Phenomenological Research*. First Published Online 2018.
- Sayre-McCord, Geoffrey. 1988. "Moral Theory and Explanatory Impotence." *Midwest Studies in Philosophy* 12 (1): 433–457. <https://doi.org/10.1111/j.1475-4975.1988.tb00177.x>.
- Sayre-McCord, Geoffrey. Unpublished. "On a Theory of a Better Moral Theory and a Better Theory of Morality."
- Scanlon, T. M. 2014. *Being Realistic About Reasons*. Oxford: Oxford University Press.
- Schafer, Karl. 2010. "Evolution and Normative Scepticism." *Australasian Journal of Philosophy* 88 (3): 471–488. <https://doi.org/10.1080/00048400903114219>.
- Schafer, Karl. 2016. "The Modesty of the Moral Point of View." In *Weighing Reasons*, edited by Errol Lord and Barry Maguire, 241–256. New York: Oxford University Press.
- Schechter, Joshua. 2018. "Is There a Reliability Challenge for Logic?" *Philosophical Issues* 28 (1): 325–347. <https://doi.org/10.1111/phils.12128>.
- Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127 (1): 109–166. <https://doi.org/10.1007/s11098-005-1726-6>.
- Sturgeon, Nicholas. 1986. "Harman on Moral Explanations of Natural Facts." *The Southern Journal of Philosophy* Supplement 24: 69–78. <https://doi.org/10.1111/j.2041-6962.1986.tb01597.x>.
- Tennant, Neil. 2017. *Core Logic*. Oxford: Oxford University Press.
- Vavova, Katia. 2014. "Debunking Evolutionary Debunking." In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, Vol. 9, 76–101. Oxford: Oxford University Press.
- Williamson, Timothy. 2017. "Semantic Paradoxes and Abductive Methodology." In *Reflections on the Liar*, edited by Bradley Armour-Garb, 325–346. Oxford: Oxford University Press.
- Williamson, Timothy. Unpublished. "Abductive Methodology and Applied Mathematics."
- Woods, Jack. 2018a. "Footing the Cost (of Normative Subjectivism)." In *Methodology and Moral Philosophy*, edited by Jussi Suikkanen and Annti Kauppinen, 166–190. London: Routledge.
- Woods, Jack. 2018b. "Mathematics, Morality, and Self-Effacement." *Noûs* 52 (1): 47–68. <https://doi.org/10.1111/nous.12157>.
- Woods, Jack. Forthcoming a. "Against Reflective Equilibrium in Logical Theorizing." *Australasian Journal of Philosophy*.
- Woods, Jack. Forthcoming b. "Logical Partisanship." *Philosophical Studies*. First Published Online 2018.
- Wright, Crispin. 2004. "Warrant for Nothing (and Foundations for Free)?" *Aristotelian Society Supplementary Volume* 78 (1): 167–212. <https://doi.org/10.1111/j.0309-7013.2004.00121.x>.