Limited Revisionism and Error Theory

Andrew B. Key

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Arts
In
Philosophy

James C. Klagge
Daniel Wodak
Lydia Patton

April 30th, 2019
Blacksburg, Virginia

Limited Revisionism and Error Theory

Andrew B. Key

ABSTRACT

In Joyce's *Myth of Morality*, Joyce proposes an error theory about morality. He then argues that, once we accept an error theory, we have three options: we can be abolitionists and jettison moral discourse, be conservationists and maintain our false moral beliefs, or be revolutionary fictionalists and assent to and act in accordance with moral discourse while believing it's false. In this paper, I argue that Joyce has ignored a fourth option—limited revisionism, or slightly changing our moral terms to avoid problematic commitments—and that this option is superior to the three aforementioned possibilities. Along the way, I show that Joyce has unfairly ignored limited revisionism because of faulty views about what makes a concept or term normative, and that limited revisionism ignores some expected pitfalls, such as overgeneralizing to legitimately error-theoretic discourses.

Limited Revisionism and Error Theory

Andrew B. Key

GENERAL AUDIENCE ABSTRACT

In this paper, I argue against Joyce's assertion that if our moral terms centrally implicate falsehoods, we should act as though morality is true (because it's useful) while also believing that it's false (because doing otherwise would be epistemically irresponsible). I show Joyce's position rests on a specific misunderstanding of how language operates, and argue instead that, if we accept that our current moral talk centrally implicates falsehoods, we should be open to slightly revising our moral terms to avoid implicating such falsehoods instead. I also show that my position, limited revisionism, avoids a few expected pitfalls—most importantly, it doesn't overgeneralize and apply to terms like "witch" and "phlogiston," which also centrally implicate falsehoods but seem like terms that shouldn't be open to conceptual change.

Table of Contents

Limited Revisionism and Error Theory

In Joyce's *Myth of Morality,* Joyce proposes an error theory about morality. He then

argues that, once we accept an error theory, we have three options: we can be *abolitionists* and

jettison moral discourse, be *conservationists* and maintain our false moral beliefs, or be

revolutionary fictionalists (hereafter, just fictionalists) and assent to and act in accordance with

moral discourse while believing it's false (Joyce 2001, 214). In this paper, I argue that Joyce has

ignored a fourth option—*limited revisionism*—and that this option is superior to the three

aforementioned possibilities.[1]

To begin, in Section I, I examine error theory and Joyce's motivations for pursuing

fictionalism. Then, in Section II, I use a naive but still intuitively compelling argument against

error theory to motivate limited revisionism. In Section III, I both refine limited revisionism, and

defend its viability against substantial objections by the error theorist. And finally, in Section IV,

I argue that once limited revisionism is on the table, it is a superior choice to fictionalism.

Section I: Error Theory and Fictionalism

Early on in Joyce's book, he formulates how error theories work:

An error theory... involves two steps of argumentation. First, it involves ascertaining just
what a term *means*. I have tried to explain this in terms of "non-negotiability," which in

---

1   I'd like to say two things here. Firstly, I'd like to thank my committee members and several others for their help
     and support—Daniel Wodak, Lydia Patton, Jim Klagge, Gregory Novack, and several others, all of whom
     provided valuable feedback and support.
       Secondly, I should note that the main thrust of this paper is similar to a point made in Nadeem Hussain's
     "Return of Moral Fictionalism" in Section 3.2 (Hussain 2004, 169–71). However, the point is hastily made and,
     both underestimates the move's import and the resources the error theorist has to attempt to block the move.
     I hope to give this idea the care it deserves.

turn I understood in terms of a translation test…[2] So, in artificially simple terms, the first step gives us something roughly of the form "For any *x*, Fx if and only if Px and Qx and Rx." We can call this step *conceptual*. The second step is to ascertain whether the following is true: "There exists an x, such that Px and Qx and Rx." If not, then there is nothing that satisfies "...is F." Call this step *ontological* or *substantive.* The concept of *phlogiston*—with its commitment to a stuff that is stored in bodies and released during combustion— and the concept of *tapu*[3]—with its commitment to a kind of contagious pollution—do not pass the test (Joyce 2001, 5).

In short, to propose an error theory, we identify some necessary conditions that are a "non-negotiable" part of a term's intensional definition,[4] and then show that there is nothing that meets those "non-negotiable" conditions. As a result, the term's extension is empty, and the discourse largely fails to state truths.

Applying this to the moral case, Joyce identifies a non-negotiable commitment to "regardless reasons"—reasons with overriding priority over all personal circumstances, desires, and so on—and argues they don't exist, which he says sinks the discourse (Joyce 2001, 42).

With these clarifications, we can move onto Joyce's answer to what I'll call his OUGHT QUESTION: "If there's nothing we [morally] ought to do, then what ought we to do (Joyce 2001, 175)?"

As stated in the introduction, Joyce believes that once we accept an error theory, we have three options: abolitionism, conservationism, and fictionalism. Joyce himself endorses

---

2   This requires a bit of explication that isn't present here. In essence, some part of the definition of a term is "non-negotiable" if it can't be adequately translated without maintaining it—if it is so integral that losing that part of the concept would render it no longer that concept.

3   *Tapu* is a term used by the Polynesians in a context vaguely similar to moral discourse, but it has ontological implications about an uncleanliness or pollution residing in objects that can spread through contact and be ritually cleansed. Joyce uses *tapu* and phlogiston as pet examples of error theoretic discourses.

4   For our purposes, a term's intensional definition is the necessary and sufficient conditions for something to be whatever is being defined (for example, a pen is a writing implement that marks with ink, etc.), while its extension is the set of all the things that meet the conditions of the intensional definition (for example, a list of all the pens in the world). This is not the only way to think about these terms, but it suffices here.

fictionalism, and he endorses it *strictly* because it wins a cost/benefit analysis between the three. Why reject abolitionism? He argues that moral discourse is useful in bolstering self control against the "perverse desire for present profit" and ensuring mutual cooperation, so abolitionism is undesirable. And conservationism? He argues that conservationism could not be a reliable strategy in the long run (given that it entails the promulgation of falsehoods or mass deception), and that it is epistemically irresponsible besides. So, Joyce opts for fictionalism. He believes that—while it cannot provide all the benefits of a sincerely believed morality— fictionalism essentially has the benefits of conservationism with the epistemic responsibility of abolitionism, so it's our best choice (Joyce 2001, 214–15).

This is a sufficient overview for our purposes. Onward, to limited revisionism.

Section II: Introducing Limited Revisionism via the NAIVE ARGUMENT

In this section, I explain limited revisionism and motivate it as a fourth response to the OUGHT QUESTION. To do so, I examine a naive argument against error theory that—even after we know it to fail—still has some intuitive pull, and then try to channel that pull into limited revisionism. At any rate, the naive argument goes something like this:

NAIVE ARGUMENT

(N1) If error theory is true, the intensional definitions of folk moral terms pick out the null set.

(N2) If the folk moral view is true, the intensional definitions of moral terms pick out recreational slaughter as wrong, and so on.

(N3) The same intensional definition cannot pick out two different extensions.

(N4) Error theory and our folk moral views give two different intensional definitions for our moral terms.

(N5) Error theory and folk morality aren't using the same definitions for moral terms.

(N6) Error theory can only refute folk morality if it's using the same definitions for moral terms.

(NC) Error theory doesn't refute folk morality, since it doesn't use the same intensional definitions as folk morality in the first place.

NAIVE ARGUMENT, if successful, would be disastrous for the error theorist. However, the error theorist has an easy response: **(N2)**, he would say, is false, because, while we *delusionally* *think* our intensional definitions pick out recreational slaughter and so on as wrong, they *actually* pick out the null set. (Most people are wrong about what the intensional definitions of moral terms pick out on an error theoretic account—it is an error theory, after all.) According to the error theorist, our folk moral view is **committed** to regardless reasons being in the intensional definitions of moral terms —it's *non-negotiable*—so our commonsense moral view is **committed** to the extension of moral terms being the null set, regardless of our delusions about the matter. So **(N2)** is false, premised on a failure to understand what our accepted intensional definitions actually pick out, and sinks the argument.

That response seems correct. However—and this feels like the point the argument wants make but doesn't—it's not clear why, *in the moral case*, we **must** stay committed to these non-negotiable intensions (NNIs), and the moral terms and moral discourse that uses them.[5] In fact, we can see that many would rather reject a morality that couldn't pick out things that they felt were obviously wrong. Take this passage from Shafer-Landau's *Moral Realism*:

> ...Aren't we still justified that, absent special circumstances, one's enjoyment doesn't make it right to hurt others? There certainly seems to be something intuitively plausible about such a claim... I would be more inclined to view a system that rejected it as

---

5   A few clarifications are in order here. Yes, I agree the NNIs are non-negotiable. And yes, I agree our discourse is committed to them. But this doesn't mean *every possible moral discourse* is committed to them, and some of those possible discourses come out very close to folk morality.

corrupt, as a counterfeit morality, than to abandon such a claim (Shafer-Landau 2003, 249).

Shafer-Landau's statement implies that, given the choice between a morality that doesn't say that hurting people is wrong, and a morality without regardless reasons, the latter—contrary to the error theorist—would be the better candidate for a good moral theory, and I agree. Now, this doesn't mean we need to reject the *spirit* of these NNIs; after all, there is no doubt that regardless reasons are central to our folk moral concepts. However, we needn't be beholden to regardless reasons in their uninstantiated forms. We could instead find a "next best thing:" a close replacement, serving most (if not all) the needs that regardless reasons did, without the trouble. If we did this, and altered our discourse to use similar but functional concepts incorporating these replacements, we could have a working—and very similar—morality to the one we already have. This idea is the thrust of limited revisionism.[6]

If this is still unclear, imagine a moral term—GOOD*, say—that has an intensional definition identical to 'good' *except* that it drops the problematic NNIs and replaces them with a "next-best thing." Unlike 'good,' GOOD* would have a non-empty extension, and pick out (almost?) all the things we thought 'good' picked out. Since GOOD* would be nearly identical to 'good', it could replace it in everyday discourse with minimal issue. If GOOD* so replaced 'good' (and *mutatis mutandis* for our other moral terms), our moral terms could successfully refer; our moral practice wouldn't be steeped in error. This would be a great success.

---

6   This is very similar to Frank Jackson's proposal about freewill (Jackson 1998, 1; especially p.44). My model differs on multiple points, but comparing may elucidate what I'm doing; the biggest difference is he takes commitments about freewill to be *hierarchical*, while I'm accepting the NNIs are, well, non-negotiable.

If this is viable, it seems that we have another answer to the OUGHT QUESTION: we could adopt a kind of *limited revisionism* instead of accepting abolitionism, conservationism, or fictionalism. The error theorist assumes we cannot give up regardless reasons, and therefore dismisses the prospects of limited revisionism. But it appears that we *could* look for a "next best thing" to objective prescriptivity and regardless reasons. And, given the choice, I think we would gladly take an option that lets us *truthfully* say that recreational slaughter is wrong, instead of just pretending. Limited revisionism, if tenable, would let us do exactly that.

Section III: Defending Limited Revisionism

However, it's unclear if limited revisionism is tenable; in fact, it looks outright question-begging. No matter how limited the changes, we *do* want to change the NNIs of moral terms, and that is precisely what the error theorist says we can't do.

So in this section, I show that—*contra* the arguments of the error theorist—our intuitions from Section II can hold water. I examine the argument by analogy (which I'll uncreatively call ARGUMENT BY ANALOGY) that the error theorist uses to block limited revisionism. Then, I argue the assumptions unpinning ARGUMENT BY ANALOGY are implausible, and further, that even if they were plausible, we shouldn't care *anyways*. Further—should the error theorist be concerned that limited revisionism overgeneralizes, in light of my arguments—I argue that we can provide specific criteria that make limited revisionism well-motivated for morality and

6

completely unmotivated for *tapu*, phlogiston, witches, and other legitimately error theoretic

discourses. Limited revisionism holds strong.

So, to begin understanding ARGUMENT BY ANALOGY, let's start with Joyce's own words:

On the one hand, we might have a discourse that centers on a predicate "…is P,"
involving the assertion of a variety of propositions—"a is P," "b is not-P," "For any x, if x
is P, then x is Q," etc.—and when we discover that we're *mistaken* about one or more of
these things—e.g., we discover that some things that are P are not Q—we don't decide
the whole "P discourse" has been a disastrous mistake; we simply change our minds
about one aspect of it: we stop making the conditional claim and carry on much as
before. On the other hand, there are some discourses regarding which the discovery
that one or more of the things we've been assenting to is mistaken leads us to throw in
the towel—to stop using the discourse altogether. The latter describes what happened
in the phlogiston case: the discovery that we had been wrong in thinking that there is a
stuff stored in combustible bodies and released during burning was sufficient for us to
decide that there is no phlogiston at all. When Lavoisier gave us the concept *oxygen*, it
wasn't available for Stahl to say "Well, this stuff that Lavoisier is calling 'oxygen' just *is*
what I've been calling 'phlogiston' all along—I was just mistaken about its being stored
and released during combustion." The belief that phlogiston is stored and released was
a *non-negotiable* part of the phlogiston discourse – the falsity of this belief was
sufficient to sink the whole thing (Joyce 2001, 3–4).

In short: in the case of phlogiston, that the stuff is stored and released during combustion is so

integral to the discourse that we must admit the discourse is committed to things that just

aren't there; **if it weren't so committed, it wouldn't be phlogiston discourse.** We accept that

regardless reasons are NNIs of morality, just as we accept that something being stored and

released during combustion is an NNI of phlogiston. So anything that doesn't involve talk of

non-existent regardless reasons "just isn't morality;" **if it weren't so committed, it wouldn't be**

moral discourse. A limited change of subject is still a change of subject. Limited revisionism is

off the table. So goes the ARGUMENT BY ANALOGY.

Frustratingly, this isn't very clear on what the underlying analogy between moral terms and phlogiston terms *is*. I think, however, it's that Joyce believes the "What is it in virtue of which a concept is about X?" question have similar answers for morality and phlogiston. I will operate under this assumption, since I see few, if any, alternatives.

Luckily for us, Matti Eklund asks what makes a concept normative in their *Choosing Normative Concepts.* And Eklund offers two possible answers: THE METAPHYSICAL VIEW, and THE NORMATIVE ROLE VIEW (EKLUND 2017, 71 and 79).[7] They go roughly as follows:

> THE METAPHYSICAL VIEW: A concept is normative in virtue of it ascribing a
> normative property. (Here think of ascribing Mackie's queer properties,
> or Joyce's regardless reasons.)
> THE NORMATIVE ROLE VIEW: A concept is normative because it's semantically
> associated with normative use.

Joyce's error theory obviously holds THE METAPHYSICAL VIEW. And I think the ARGUMENT BY ANALOGY exists to motivate intuitions that THE METAPHYSICAL VIEW is plausible—it's very intuitive that a concept is about phlogiston because it picks out something stored and released during combustion. However, THE METAPHYSICAL VIEW has a host of issues. As Eklund notes, it seems standard to think that ascribing the same referent doesn't make two concepts of same kind: the concepts of 'c-fibers firing' and 'pain' are physical and mental concepts, respectively, and (ignoring that this has been discredited) they have the same referent—yet this doesn't make 'c-fibers firing' a mental concept. The view seems implausible for thick normative concepts that *misevaluate* (such as 'chaste' or 'lewd'). And it also seems possible to refer to a normative

---

7 Some quick terminology: normative role is what's distinctive about the use of normative concepts in language
—be that their prohibiting or recommending things a certain way, expressing certain attitudes, etc. (Eklund
2017, 10). (We need not be committal here). An aside: Eklund also notes one might take a DISMISSIVE VIEW, and
think that there are no normative concepts (Eklund 2017, 65). I ignore this complication.

property using a non-normative concept. Eklund imagines an alien learning English who introduces a concept into their alien language, 'thgir;' 'thgir' ascribes the same referent 'right' does but doesn't play a role in deliberation and so on, and thus seems non-normative (Eklund 2017, 71–79). Although these arguments are only presented hastily—and I encourage looking at them in their entirety—they make it hard to endorse THE METAPHYSICAL VIEW. I think the error theorist should abandon it.[8]

At this point, I could hold that ARGUMENT BY ANALOGY fails because THE METAPHYSICAL VIEW is bad, but this would be disingenuous: the error theorist doesn't actually need to hold THE METAPHYSICAL VIEW. THE NORMATIVE ROLE VIEW is also compatible with error theory, provided the error theorist accepts a few additional premises:

> **MORAL ERROR THEORY WITH THE NORMATIVE ROLE VIEW**
> _____
> **EXCLUSIVITY:** that all possible concepts of good share a normative role;
> **REFERENCE:** that all possible concepts of good are referentially normative;[9]
> **CONCEPTUAL STEP:** that this shared normative role attempts to pick out things that give regardless reasons;
> **ONTOLOGICAL STEP:** that nothing gives regardless reasons;
> **ERROR:** therefore, all possible concepts of good are defectively referentially normative (and *mutatis mutandis* for other moral terms).

---

8   Perhaps the error theorist could say the folk view is committed to THE METAPHYSICAL VIEW even though it doesn't make sense, in the same way we're committed to regardless reasons even though they don't make sense. Then it wouldn't matter if it were flawed. What then?
    I don't think this response works. For one, as noted, THE METAPHYSICAL VIEW isn't problematic only for normative concepts; the pain/c-fibers firing example makes it clear it shouldn't hold up for concepts like phlogiston, either. This means the error theorist would have to make a much more sweeping claim to keep the ARGUMENT BY ANALOGY intact, one that seems terribly implausible. And even if the error theorist could, the very idea of assigning something as philosophically rigorous as THE METAPHYSICAL VIEW to folk belief seems absurd.

9   More quick terminology: a concept is *referentially normative* if its normative role determines its reference, such that concepts with the same normative role must have the same reference. A concept is *defectively referentially normative* if the concept picks out an unacceptable extension—i.e., one that's empty, or wildly indeterminate, or problematically trivial—and *non-defectively referentially normative* if the extension is acceptable.

New are EXCLUSIVITY and REFERENCE. Note that without them, CONCEPTUAL STEP falls flat, and ERROR

isn't true; and further, note that EXCLUSIVITY, REFERENCE and CONCEPTUAL STEP together block

limited revisionism, since they entail all possible moral terms having empty extensions.

However, EXCLUSIVITY and REFERENCE lack support, barring intuitions. A new ARGUMENT BY

ANALOGY* could secure such support. But to make it, the error theorist would need a similar

schema for phlogiston, and the normative role equivalent for phlogiston is unclear.

Nonetheless, let's assume there is one: call it *theoretical role,* the linguistic role characteristic of

entities posited by scientific theories. The resultant schema:

PHLOGISTON ERROR THEORY WITH THE THEORETICAL ROLE VIEW

EXCLUSIVITY$_P$: that all possible concepts of phlogiston share the same theoretical role;
REFERENCE$_P$: that phlogiston's theoretical role is referentially theoretical;
CONCEPTUAL STEP$_P$: that this shared theoretical role attempts to pick out something stored
    and released during combustion;
ONTOLOGICAL STEP$_P$: that something stored and released during combustion doesn't exist;
ERROR$_P$: therefore, all possible concepts of phlogiston are defectively referentially
    theoretical.

Now, given all this, the error theorist could present an ARGUMENT BY ANALOGY*:

ARGUMENT BY ANALOGY*

(A1) EXCLUSIVITY$_P$, REFERENCE$_P$, CONCEPTUAL STEP$_P$, ONTOLOGICAL STEP$_P$, and ERROR$_P$ are true.
(A2) ONTOLOGICAL STEP is true.
(A3) Normative roles and theoretical roles are relevantly similar—such that if
EXCLUSIVITY$_P$ and REFERENCE$_P$ are true, EXCLUSIVITY and REFERENCE are true.
(A4) EXCLUSIVITY and REFERENCE are true.
(A5) If EXCLUSIVITY and REFERENCE are true, CONCEPTUAL STEP is true.
(A6) CONCEPTUAL STEP is true.
(A7) If CONCEPTUAL STEP and ONTOLOGICAL STEP are true, then ERROR is true.
(AC) ERROR is true.

Our path to refute ARGUMENT BY ANALOGY* is clear**.** If we like having an error theory about

phlogiston, we cannot deny **A1,** and we've accepted **A2.** So we must deny **A3;** and **A3** can only

be false if **A1** is true and **A4** false. Luckily, we already accept **A1.** So we have to deny **A4.** But how can we motivate that EXCLUSIVITY or REFERENCE aren't true? What would motivate rejecting them, but not also **A1**?

Let's at least choose what to reject, first. On the one hand, Eklund talks about what could happen if we reject both REFERENCE AND EXCLUSIVITY, but I don't find it promising (and the same for rejecting only REFERENCE).[10] On the other hand, though, we could reject only EXCLUSIVITY. If *not all* possible moral concepts have the same normative role—and this is how I wish to reject EXCLUSIVITY—then not all moral concepts need involve regardless reasons, and not all of them need be in error. With EXCLUSIVITY out of the way, we could posit our concepts of GOOD* and so on from earlier, and with REFERENCE intact, we could trust their meaning be appropriately tied to their use. Limited revisionism could succeed.

Why reject EXCLUSIVITY? Because while EXCLUSIVITY seems true *descriptively,* it seems false *prescriptively*. Assume NNIs are a large part of how normative roles determine reference. It is far from clear that what we consider an NNI of morality in the future should be constrained by our opinions on what a NNI of morality is now. In earlier times, perhaps divine command theory was an NNI of morality. Today, this is clearly not so. Yet an error theorist at the time that divine command theory was an NNI could've argued that morality is in error because there is no god, and thus no one to give divine commands.[11] Does that mean that, today, morality is still in

---

10 At first, rejecting both seems a good move**.** However, I think this path dead-ends. We end up with profligate moral concepts with different extensions; all the concepts are *prima facie* on par unless you like ineffable questions; and it's not clear that even appeals to joint-carving or eliteness can salvage the situation (Eklund 2017, 26–32). This does grant us the possibility that, should we salvage the situation, the error theorist might have been attacking a "false" morality. But I think that's not much comfort, since it could just as easily turn out the folk were wrong instead.

11 If you don't share the relevant beliefs, consider hypothetically instead.

error because there is no god? This does not seem true. Constraining our future morality based on "regardless reasons" may well come to seem equally backwards. Why should we? The error theorist conflates the whole of moral practice with our current moral discourse, and thus assumes our if our moral discourse changes our practices and discourse aren't morality anymore. As the divine command theory example shows, this seems obviously wrong.[12]

Despite this, let's indulge the error theorist: assume EXCLUSIVITY is true. What then?

Even if EXCLUSIVITY were true of morality, *it would still be open to us to just talk about an importantly similar topic instead*—say, MORALITY*. Why care if MORALITY* isn't morality? Whence the error in discussing MORALITY*? We could effectively give up the name of morality while still making the limited revisionist move, and at most we'd have lost a label and moved onto a new discourse with interesting problems to solve. Perhaps MORALITY* isn't what the folk care about. So be it. But why do we need to be beholden to the folk if they're systematically wrong anyways? Error theory, right though it may be, just becomes a stepping stone to MORALITY*, and MORALITY* is where all the good philosophical action is at. And hell, isn't fictionalism just one idea for what MORALITY* could look like?[13]

---

12 An interesting thought occurs to me here: there may be an analogy here between DCT and morality writ large, and phlogiston and combustion writ large. Much like losing DCT doesn't sink morality, losing phlogiston doesn't sink combustion; and this seems true despite the fact that phlogiston was an integral part of explaining combustion. I am hesitant to say it was non-negotiable to combustion in the *exact* way DCT would've been non-negotiable to morality; but it may have been non-negotiable to the *then current conception* of combustion in the way DCT was non-negotiable to the *then current* conception of morality. There are some questions here about Moorean premises and the comparability of the situations based on that, though. Since combustion occurring is blatantly obvious, it makes sense that we would keep combustion even if we jettisoned phlogiston; however, morality "occurring" isn't similarly obvious. For relevant comments, see (McPherson 2009).

13 This sounds very similar to lines of argument I have heard in several different places, though I do not remember where. I credit whoever said these arguments, even if I myself cannot remember where I heard them. I think it's something like Sider's response to Thomasson's views on the existential quantifier?

This sounds flippant, but it's a serious issue. The error theorist's arguments might block changing the NNIs of morality if EXCLUSIVITY is true—but they can't block moving to MORALITY*, and MORALITY* can just be morality with our limited change of subject. There's nothing special about morality over MORALITY* other than it better representing folk-theory, which the error theorist has already shown us is bankrupt. MORALITY* even works *better* than morality, in the sense it's at least making true claims. The error theorist has guarded the fortress of "morality," but it seems a Pyrrhic victory.

Yet the error theorist can rightfully object this is too loose. "Let's make another argument by analogy: why not change phlogiston to PHLOGISTON*, *tapu* to TAPU*, and so on? Surely there is an error in *that.*" And I agree: something is wrong if the limited revisionism overgeneralizes. However, I think there is a principled way to preempt the overgeneralization worry. I propose two important criteria that a discourse must meet before limited revisionism is appropriate over error theory; morality meets these criteria, while *tapu*, phlogiston discourse, et al. do not. I discuss these criteria in turn.

The first: *limited revisionism is allowable if we need the discourse to support important societal practices or institutions.* Upon reflection, we find that regardless reasons help justify our practices of blaming and punishment; we tend to fail to find people blameworthy if we cannot show there was an objective standard they were being held against, or a reason they should have acted otherwise. Given the utter centrality of these practices to society, a "limited change of subject" here seems at least pragmatically justified.

13

Now, phlogiston discourse fails this first criterion nicely; it justifies nothing so central. *Tapu* discourse, though, might pass: it justifies certain important societal proscriptions and rituals. But an error theory about *tapu* seems appropriate. So there must be some difference, some second criterion we can appeal to, that differentiates the two.

Luckily, there is one ready at hand: there is nothing to change the subject *to* in the case of *tapu.* There is no "next-best thing," no veridical evolution of the concept that does everything we wanted of the original. Nothing gets enough of the critical claims right: no substitute can explain the need for the ritual washings, or how it spreads through contact. However, in the moral case, we seem to be able to find evolutions that get enough of the critical claims right (and at least aren't obviously false). Certainly, such views aren't flawless; but they often provide a "next-best-thing" that does most all we wanted of the folk conceptions of objective prescriptivity and regardless reasons. This is not possible for the phlogiston and *tapu* discourses, and it is for moral discourse.

Note this second criterion is especially important, because it preempts the overgeneralization worry. Take witch discourse, for example. Why is it that we cannot just have a "limited change of subject" in the case of witches? Certainly, the belief in witches underwrote many important (and awful) societal practices! However, what's non-negotiable about witches is magic and the supernatural—and there's nothing as good as that, no "next-best thing" that fills that role and keeps the critical claims of the discourse intact. And *mutatis mutandis* the other problem cases. As a result, limited revisionism doesn't overgeneralize.

So, there are two criteria for when limited revisionism is allowable: the subject in question must underwrite important societal practices, and there must *actually be* a veridical evolution of the concept that keeps the critical claims of the discourse true. This is what makes moral discourse special; it meets these two conditions, whilst phlogiston discourse, *tapu* discourse, and so on do not.

So there is a principled way to sort morality out from *tapu*, phlogiston, and witch discourse; and in fact, limited revisionism is *impossible* for the latter. The error theorist has no nontrivial way to stop limited revisionism once we've restrained its applicability. So limited revisionism is on the table with abolitionism, conservatism, and fictionalism, and should be included in Joyce's cost/benefit analysis. I take up that cost/benefit analysis now.

Section IV: Winning the Cost/Benefit Analysis

Let us assume that our arguments have paved the way for limited revisionism to be included with abolitionism, conservationism, and fictionalism as answers to the Ought Question. I can now defend that limited revisionism wins the cost/benefit analysis between the four.

To begin with, let us review the premises that Joyce uses to argue for fictionalism:

(F1) that abolitionism is undesirable because moral discourse is useful in bolstering self-control and ensuring mutual cooperation;
(F2) that conservationism is undesirable because it's epistemically irresponsible;
(F3) that fictionalism is not epistemically irresponsible and provides most of the benefits of morality listed in **(F1).**

I'll accept **(F1)** and **(F2);** our cost/benefit analysis is therefore limited to fictionalism and limited revisionism. However, I'll be attacking **(F3).** I think limited revisionism is more epistemically responsible, more rational, and provides the same alleged benefits. This is because I see an exploitable tension between **(F3)** and **(F2):** if we reject conservationism because it's epistemically irresponsible, *isn't acting in accordance with fictional beliefs[14] (that are contrary to our actual beliefs) irrational?* We can think of this as a question of rational justification: can any acts we perform because of our fictionalist beliefs be rationally justified?[15]

Now, the fictionalist does have a response: acting in accordance with fictional beliefs is *prudentially* justified since it has good outcomes. And presumably, acting prudentially is acting rationally. But this is too quick. It implies that the consistency between our behaviors and *our true beliefs about the outcomes of our behaviors* is what's rational. But if that consistency does the justificatory work, why not act because of those true beliefs instead of the fictional ones? We could just cut out the middleman, and then fictionalism seems to offer nothing.

To illustrate my point, consider Joyce's example of someone who's determined to exercise:

---

14 Note that Joyce ultimately caches this out in terms of noncognitive pretence belief, but this won't evade the problem; if we're acting against our true beliefs about morality in the way fictionalism seems to require, that's still irrational. It doesn't matter if we're acting in accordance with a noncognitive state if we're acting against a true belief state we have. And merely acting as if x is true when x is known to be false seems to invoke suspicions of acting against one's evidence.

15 Note that it is important that we talk of (subjective) rational justification, *but not moral justification*. Talking of moral justification would be problematically circular. I should also note that this is similar enough to a line of argument Hussain takes that I would be remiss to not mention it: he argues it's unclear how pretend must-be-doneness has any purchase over something like maximizing Railton's non-moral good, since neither *really* has must-be-doneness. Especially relevantly to us, he says "Joyce has to argue that make-believing that murder is wrong will result in more cooperation and coordination than believing that murder is wrong*" (Hussain 2004, 170–71); but I argue that it's more *rational* to act on the basis of believing murder is wrong* than to act because you pretend-believe murder is wrong while you *actually* believe murder has no moral valence.

Suppose I am determined to exercise regularly, after a lifetime of lethargy, but find myself succumbing to temptation. An effective strategy will be for me to lay down a strong and authoritative rule: I must do (say) fifty sit-ups every day, no less. I am attempting to form a habit, and habits are formed—and for the doggedly weak of will, maintained—by strictness and overcompensation. Perhaps in truth it doesn't much matter that I do fifty sit-ups every day, so long as I do more or less fifty sit-ups on most days. But by allowing myself the occasional lapse, by giving myself permission to *sometimes* stray from the routine, I pave the way for akratic sabotage of my calculations —I threaten even my doing more-or-less fifty sit-ups on most days. I do better if I encourage myself to think in terms of fifty daily sit-ups as a non-negotiable value.

However, to believe sincerely that fifty daily sit-ups are needed for me in order to achieve my end of health is to have a false belief (we'll assume), the holding of which will require other compensating false beliefs. If it is true that *more-or-less* fit sit-ups *nearly* every day is sufficient for my health, then that is what I ought to believe. On the other hand, to *pay attention* to this belief exposes me to self-subversion—a slippery slope to inactivity. This is precisely a case where my best interests are served by rehearsing thoughts that are false, and that I know are false, in order to fend off my own weaknesses. But in order to get the benefit from this strategy there is no necessity that I *believe* the thoughts, or attempt to justify them as true when placed in a philosophically critical context. When doing my sit-ups I think "Must do fifty!" but if, at some other time, you ask me whether I really *must* do fifty, then I will say "No, sometimes forty would suffice (Joyce 2001, 215).

Joyce's idea here is clear: our self-deception is prudentially justified, because we will do better if we have false but COMMITTED BELIEFS that *we must* do fifty sit-ups every day and likewise keep a corresponding COMMITTED HABIT—even if the health facts say this is unnecessary. This is so because we are WEAK-WILLED: we would inevitably fail to achieve our goals with CASUAL BELIEFS and a WEAK HABIT. As a result we do better to ignore the health facts, and have COMMITTED BELIEFS and a COMMITTED HABIT.

I think Joyce's whole argument is confused. First, an intuitive pass: if you are WEAK-WILLED, *then you need a COMMITTED HABIT to succeed;* your COMMITTED BELIEFS would, therefore, be no self-deception. Perhaps, in a world where you were not WEAK-WILLED, having your

17

COMMITTED BELIEFS would be a self-deception. But that isn't this world—the contents of your

COMMITTED BELIEFS are true *precisely because* you would fail without a COMMITTED HABIT. So there

is, in fact, no self-deception here. And now, a rigorous pass: this entire passage conflates all-

things-considered reasons and simple, overrideable reasons, and assumes we cannot keep

them straight. The health facts give a simple, overrideable reason to have a WEAK HABIT.

However, you have an *all-things-considered reason* (one that has already accounted for the

health facts, that you are WEAK-WILLED, and so on) to keep the COMMITTED HABIT. There is an

obvious *hierarchy* of reasons here, and the all-things-considered reasons very clearly win out.

There are various ways this could be cached out, but I'd rather not commit to any now, since

that's a sizable project. I only need you to agree to two things: first, that *it is true* that you have

overwhelmingly better odds of success with COMMITTED HABIT if you are WEAK-WILLED; and

second, that there's no need to pretend-believe the corresponding COMMITTED BELIEFS as a result

—you could just *actually* believe them instead. (If you know slipping for a single day would have

terrible consequences, letting yourself slip for a single day seems paradigmatically

instrumentally irrational.) There is no need for fictional belief here when true belief will do the

same work; and *mutatis mutandis,* I say, for moral beliefs. (E.G., perhaps I only need a WEAK

HABIT of honesty to reap the benefits. But almost all of us are WEAK-WILLED. So it is true that we

should have COMMITTED HABITS and the corresponding COMMITTED BELIEFS. This gives us a very

substantive reason to be honest.)[16] Since the contents of the fictionalist beliefs *are true*

16 I should note the kind of belief and behavior patterns I'm recommending here are quite familiar. Consider, for
    example, the man who does not drink at all because he knows he's prone to substance abuse. Certainly, the
    health facts are that there's a middle road for him between being a teetotaler and being a drunkard. But he
    knows his own psychology and is a teetotaler precisely because he will fail attempting to follow that middle
    road. Given his psychology, being a teetotaler is more—not less—rational. The same seems true here.

anyways, fictionalism gains no benefit from this line of argument. If anything, it seems to give us fuel for a limited revisionist move that reduces some categorical reasons to extremely common prudential ones.

So, at this point, we can see that the reasons why Joyce championed fictionalism—promoting epistemic responsibility and rational behavior, bolstering self-control—seem to support limited revisionism as well if not better. We can have true beliefs about what's GOOD*; there is no irrationality in acting on something being GOOD*; and how Joyce justifies fictional beliefs justifies the corresponding true ones. Therefore, I conclude: if fictionalism and limited revisionism are the only competitors in our cost/benefit analysis, limited revisionism wins.

Conclusion

In this paper, I explained error theory, and argued that when Joyce asked the OUGHT QUESTION, he provided three answers—abolitionism, conservationism, and fictionalism—but left out the best answer, limited revisionism. I showed how limited revisionism is well-supported by our intuitions, and defended limited revisionism from the ARGUMENT BY ANALOGY. And finally, I made the case that limited revisionism wins a cost/benefit analysis against fictionalism.

Where to go from here? I think, perhaps, an interesting metaphilosophical point results: it's possible many metaethical positions share common ground with limited revisionism, given that any theorist who thinks regardless reasons are central to folk morality, acknowledges they don't exist, and then produces an account of morality without them is essentially engaging in a limited revisionist project. If this is true, error theory loses much of its bite and uniqueness;

19

many theorists we consider realists may well be limited revisionists. And perhaps limited

revisionism implies another metaphilosophical point: we could see the traditional error theorist

as having been "misguided all along," or "playing a different game," as so much of metaethics

was concerned with making consistent intensions to match our extensions while the error

theorist wanted our extensions to conform to our faulty intensions. But these points are vague

and sketchy. One clear takeaway: you can be an error theorist about morality, but a realist

about MORALITY*—and being a REALIST* is probably the "next best thing" to being a realist.

Bibliography

Eklund, Matti. 2017. *Choosing Normative Concepts*. Oxford University Press.

Hussain, Nadeem J. Z. 2004. "The Return of Moral Fictionalism." *Philosophical Perspectives* 18 (1): 149–188.

Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Vol. 1. Oxford University Press.

Joyce, Richard. 2001. *The Myth of Morality*. Cambridge University Press.

McPherson, Tristram. 2009. "Moorean Arguments and Moral Revisionism." *Journal of Ethics and Social Philosophy*, no. 2: 1–25.

Olson, Jonas. 2011. "Getting Real about Moral Fictionalism1." *Oxford Studies in Metaethics* 6: 181.

Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. Oxford University Press.