

## ORIGINAL ARTICLE

# Intertranslatability, Theoretical Equivalence, and Perversion

Jack Woods

University of Leeds

I investigate syntactic notions of theoretical equivalence between logical theories and a recent objection thereto. I show that this recent criticism of syntactic accounts, as extensionally inadequate, is unwarranted by developing an account which is plausibly extensionally adequate and more philosophically motivated. This is important for recent anti-exceptionalist treatments of logic since syntactic accounts require less theoretical baggage than semantic accounts.

**Keywords** philosophy of logic; theoretical equivalence; anti-exceptionalism; revising logic; intuitionistic logic; non-classical logic

DOI:10.1002/tht3.269

It's increasingly common to claim that logic isn't special, that a logical theory is just a theory like any other (Hjortland 2017; Priest 2014; Russell 2014; Williamson 2017). These *anti-exceptionalist* views use theoretical virtues of alternative logics — simplicity, strength, etc. — in order to evaluate them analogously to how we evaluate scientific theories.<sup>1</sup> In doing so, it's helpful to know when two seemingly distinct theories are really differently formulated versions of the same theory, saying the same things in different ways. In other words, it's helpful to have an account of *theoretical equivalence* between logics.

For example, axiomatic and natural deduction presentations of classical propositional logic initially look quite different even though they're intuitively theoretically equivalent. Now, natural deduction is less metatheoretically tractable while axiomatic accounts do worse at revealing obvious entailments. Neither disadvantage is problematic once we're justified in treating them as theoretically equivalent; perceived disadvantages of a logical theory that are merely presentational can be ignored in calculating its overall virtues.

We have options when developing a notion of theoretical equivalence. One standard way uses semantic properties, like having the same set of models, to account for when two logical theories are equivalent. However, as the required model theory can only be developed by appealing to significant logical resources, many of which are in contention when choosing between logics, it's natural for anti-exceptionalists to prefer less theoretically taxing syntactic accounts of theoretical equivalence. Syntactic accounts require lighter metatheoretic resources (roughly a basic theory of syntax) which can often be developed

Correspondence to: E-mail: j.e.woods@gmail.com

in a neutral fashion with resources shared between a pair of logical theories.<sup>2</sup> Even if we can sometimes use a semantic notion of theoretical equivalence, it's useful to have a syntactic notion available for cases where the resources necessary to develop the semantic notion are in contention.

However, in a recent paper, John Wigglesworth argues that syntactic accounts of theoretical equivalence are simply too coarse-grained to explicate any reasonable notion of theoretical equivalence between logics. This because even the best syntactic accounts treat classical and intuitionistic logic as theoretically equivalent. Yet no two logical theories could be more obviously distinct than these (Wigglesworth 2017). His argument works against some syntactic accounts, but it doesn't work against them all. I'll show here how to avoid his objection by developing a more adequate and, importantly, more discriminating syntactic account of theoretical equivalence.<sup>3</sup>

A standard syntactic account of theoretical equivalence in the philosophy of science uses *intertranslatability* between theories—mappings between sets of sentences closed under classical logic—as the mark of theoretical equivalence. This doesn't work for comparing various logics. For one reason, we can't presume a uniform notion of closure. For another, logics centrally include claims about how different claims—including various hypothetical claims—about the world relate to one another. Whether one thinks about this enterprise as worldly or not, this requires us to modify the standard account of syntactic equivalence.<sup>4</sup> We need to modify the standard account of theoretical equivalence so that it captures equivalence between two *logical* theories instead of between two scientific theories.

The first thing to do is to generalize beyond particular closed theories to logics. We treat logics as sets of axioms and rules. For a sequent calculus and (some) presentations of natural deduction, this will include structural rules like identity, weakening, and cut as well as operational rules governing connectives like *modus ponens*. For axiomatic systems and (some) presentations of natural deduction, this will include axioms and operational rules. Since our concern is syntactic theories, we'll use derivability—repeated application of rules—as our closure condition, saying that  $\psi$  is in the closure of a set of sentences  $S$  in a logic  $L$  (writing this closure as  $L_S$ ) just in case  $\psi$  can be obtained by repeated application of the rules characterizing  $L$  to sentences in  $S$  and the sentences we obtain by such application.<sup>5</sup>

For Wigglesworth, two logical theories  $L$  and  $L^*$  are *intertranslatable* when, for every set of sentences  $S$ , there are mappings  $a_1: L_S \rightarrow L^*_S$  and  $a_2: L^*_S \rightarrow L_S$ , where these mappings commute with negation<sup>6</sup> (i.e., where  $a_1(\neg\psi) = \neg a_1(\psi)$  and analogously for  $a_2$ .) This generalized intertranslatability account looks initially plausible for explicating theoretical equivalence between logics.<sup>7</sup>

As Wigglesworth demonstrates, though, it faces extensional problems. In light of these, he claims that we should give a semantic characterization of theoretical equivalence instead of a syntactic one. This might be right if the generalization just mooted was the best syntactic account of theoretical equivalence, but it isn't. We can avoid Wigglesworth's conclusion by refining the intertranslatability account of theoretical equivalence in three ways: the first involves ensuring a common language, the second generalizing the

translations, and the third involves demanding that the translations preserve meaning, insofar as that can be done in a syntactic setting.<sup>8</sup>

These refinements, though they result in a more stringent notion of theoretical equivalence, are natural, philosophically motivated, and do not undermine the light-touch syntactic approach anti-exceptionalist should favor. Motivating these refinements involves discussing some important issues about theoretical equivalence between logics, so the ensuing discussion is important generally, not just as a way of avoiding an objection to a particular approach to theoretical equivalence between logics.

### Refinement 1: Ensuring common languages

The problems with Wigglesworth's target syntactic account start when the languages of the logics evaluated are either disjoint or overlapping. For example,  $p \vee \neg p$  occurs nowhere in the pure theory of conjunction ( $\&$ ) and negation ( $\neg$ ). Suppose we compared this theory with the pure theory of disjunction ( $\vee$ ) and negation. By the syntactic account given above, we'd need to compare the closure of every set of sentences involving  $\vee$  and  $\neg$  under both the pure theory of disjunction and negation and the pure theory of conjunction and negation. But since the latter doesn't contain any rules for disjunction, presumably  $\{p \vee \neg p\}$  has a high-trivial closure in the latter theory.

For a more worrisome example, let  $B$  be the pure theory of conjunction, disjunction, and negation and  $B^*$  the intuitive notational variant which treats  $\vee$  as conjunction and  $\&$  as disjunction (modifying the corresponding rules in the obvious way). Now consider the effects of closing a tautology (in  $B$ ) under  $B$  and  $B^*$ .  $B_{\{p \vee \neg p\}}$  contains all and only the classical tautologies in  $\&$ ,  $\vee$ , and  $\neg$ ;  $B^*_{\{p \vee \neg p\}}$  contains all sentences since we can derive both  $p$  and  $\neg p$  using "conjunction" elimination and then apply *reductio ad absurdum*.  $B$  and  $B^*$  thus aren't intertranslatable in Wigglesworth's sense. But they're intuitively notational variants and hence theoretically equivalent; so the lack of a shared language—in particular, a shared set of logical constants (and formation rules)—causes serious problems.<sup>9</sup>

The trouble, obviously, is that we can't assume a shared language with which to populate  $S$ . But the whole idea was to test the effect of each logic on arbitrary bits of common content (i.e. on various  $S$ ). There are two obvious solutions for this problem. First, we could treat intertranslatability as *relative* to an existing pair of mappings between the logical expressions of one language and the logical expressions of another (presuming for simplicity that the non-logical expressions are common.<sup>10</sup>) This allows us to formulate  $S$  in a neutral fashion, at least relative to our presumed mappings between logical expressions. We then say  $L$  and  $L^*$  are theoretically equivalent, relative to a pair of mappings  $t, u$ , just in case there are mappings  $f$  and  $g$  where for every set of sentences  $S$  in the language of  $L$ ,  $f: L_S \rightarrow L^*_{t(S)}$  and every  $S$  in the language of  $L^*$ ,  $g: L^*_{S'} \rightarrow L_{u(S')}$  (where these mappings commute with negation).

Second, we could apply our main translations to  $S$  as well, modifying our account accordingly. We'd then say that  $L$  and  $L^*$  are theoretically equivalent just in case there

are mappings  $f$  and  $g$  between  $L$  and  $L^*$  where for every set of sentences  $S$  in the language of  $L$ ,  $f: L_S \rightarrow L^*_{f(S)}$  and every  $S$  in the language of  $L^*$ ,  $g: L^*_S \rightarrow L_{g(S)}$ .

Either way, the key move is ensuring some type of common content for  $S$  so that the perhaps differential effects of closing  $S$  under each logic can be evaluated. Accordingly, we apply these mappings only to  $S$ , leaving untouched the role of each logic in closing the resulting sets of sentences under derivability. We require of translations between these results that they preserve the condition of commuting with negation so as to avoid entirely perverse pairs of translations and presumed mappings.

Each solution yields non-trivial notions of theoretical equivalence, but I prefer the first. To see why, note that  $B^*$  can be viewed two ways: as a notational variant of  $B$  or as the radical view permitting inferring disjuncts from disjunctions. The second solution doesn't distinguish these interpretations since there *are* mappings  $f$  and  $g$  between  $B$  and  $B^*$  such that for every  $S$  in  $B$ ,  $f: B_S \rightarrow B^*_{f(S)}$  and every  $S$  in  $B^*$ ,  $g: B^*_S \rightarrow B_{g(S)}$ . So they're theoretically equivalent by the second solution, no matter what we intended  $B^*$  to be. This is too coarse-grained; these distinct understandings of  $B^*$  should be distinguished.

The first solution distinguishes them. Let  $t$  map each connective in  $B$  to itself; let  $t^*$  swap  $\wp$  and  $\vee$ . The closure of  $S$  under  $B^*$  then depends on whether we're relativizing our notion of theoretical equivalence to  $t$  or  $t^*$  (note that we haven't modified these rules at all).<sup>11</sup> For example, consider the rule of "disjunction" elimination. As  $t^*$  translates  $p \wp \neg p$  as  $p \vee \neg p$ , disjunction elimination in  $B^*$  — "to infer  $\psi$  from  $\psi \vee \phi$ " — is just a notational variant of conjunction elimination presuming  $t^*$ . Presuming  $t$  instead, "disjunction" elimination permits inferring "disjuncts" from "disjunctions" *in one of the two theories*.<sup>12</sup>

So, for our first refinement, given mappings  $t, u$  between the logical expressions of  $B$  and  $B^*$ , let's require of theoretically equivalent logics  $B$  and  $B^*$  that there are mappings  $f, g$  such that for every set of sentence  $S$  in the language of  $B$ ,  $f: B_S \rightarrow B^*_{t(S)}$  and for every set  $S$  in the language of  $B^*$   $g: B^*_S \rightarrow B^*_{u(S)}$ , where  $f, g, t$ , and  $u$  commute with negation in the sense that  $f(\neg\psi) = t(\neg)f(\psi)$  and  $g(\neg\psi) = u(\neg)g(\psi)$ .<sup>13</sup> A little work shows that  $B$  and  $B^*$  aren't theoretically equivalent when we presume  $t$ . Even less work shows that they are when we swap  $t^*$  for  $t$ .

Since I prefer finer-grained approaches, I work with the first solution, treating theoretical equivalence as theoretical equivalence *relative* to mappings of the logical expressions in one language to the logical expressions of another.<sup>14</sup> In light of our target case below, we'll assume below that  $t$  and  $u$  are both the identity mapping unless otherwise specified.

## Refinement 2: Generalizing the translations

The second refinement involves the generality of  $f$  and  $g$ . Again, the usual treatment of theories as logically closed sets of sentences doesn't naturally apply to logic itself since logics centrally include claims about how claims about the world relate to each other. Wigglesworth accounts for this by treating two logics as theoretically equivalent when we have translations between the results of closing *any* set of sentences  $S$ , under both logics, into one another. Intuitively,  $S$  then plays the role of a set of premises and  $L_S$  the set of its consequences.<sup>15</sup>

This definition, though, permits using different translations for each  $S$ . These *local* translations might use special features of the content of  $S$ , instead of what each logic says about  $S$ , in translating between logics. For example, let  $D$  contain  $\psi \vee \neg\psi$  for every atomic formula  $\psi$  in our language.  $D$ , understood intuitionistically, is an extremely strong set of premises. It amounts to the *decidability* of all atomic formulas. Understood classically,  $D$  is trivial (since classical logic presumes that all formulas are “determinate”). So intuitively they “say” quite different things. However, since the classical and intuitionistic closures of  $D$  are the same, we can use the identity map to translate between the intuitionistic and classical closures of  $S$ , even though it isn’t generally adequate to translate between intuitionistic and classical logic. It’s only because of the presence of excluded middle for all atomic formulas in  $D$  that the identity translation suffices.

Permissive accounts of translation like Wigglesworth’s do reasonably well at capturing when two logically closed sets of sentences say the same thing, but badly at capturing when two logical theories generally “say” the same thing. This last, though, is what theoretical equivalence between logics should capture. We can capture this by using *general translations*: mappings  $a_1, a_2$  such that for every set of sentences  $S$  (in each language respectively),  $a_1: L_S \rightarrow L^*_S$  and  $a_2: L^*_S \rightarrow L_S$ .<sup>16</sup> We turn now to Wigglesworth’s particular counterexample to syntactic accounts as a warm-up to our third and most important refinement.

### Wigglesworth’s counterexample

Wigglesworth’s example exploits standard translations of classical logic ( $C$ ) into intuitionistic logic ( $I$ ) (due to Glivenko, Gödel, Gentzen, and others) where, for an eponymously labeled translation  $g$  and derivability relations  $\vdash_C$  and  $\vdash_I$ :

- 1:  $\vdash_C \psi \leftrightarrow g(\psi)$ .
- 2:  $\vdash_I g(\psi) \leftrightarrow \neg \neg g(\psi)$ .
- 3:  $\vdash_C \psi$  only if  $\vdash_I g(\psi)$

For concreteness, we’ll take as our  $g$  the mapping which takes  $p$  to  $\neg\neg p$ ,  $(p \vee q)$  to  $\neg(\neg g(p) \wedge \neg g(q))$ ,  $\neg p$  to  $\neg g(p)$ , and  $p \rightarrow q$  to  $g(p) \rightarrow g(q)$ . It’s an easy exercise to show that  $g$  commutes with negation, so it’s a general translation of classical logic into intuitionistic logic. As for the other direction, since (pure) intuitionistic (propositional) logic is a sublogic of classical, we can use the identity map as our translation. As  $g$  and the identity map translate, in the above sense, between the results of closing any set of sentences under intuitionistic and classical logic, Wigglesworth’s explication of syntactic accounts of theoretical equivalence treats intuitionistic and classical logic as equivalent. But they’re not. So it’s inadequate.

Should we conclude that syntactic explications of theoretical equivalence between logics are inadequate in general? This would be premature. To criticize all syntactic accounts this way, we’d need to show that Wigglesworth’s is the best syntactic account of theoretical equivalence. But it’s not; there’s a further natural refinement which solves Wigglesworth’s problem and results in a more satisfying overall account of theoretical equivalence. The key idea is requiring that two theoretically equivalent theories don’t

“say” quite different things. Even syntactic accounts of intertranslatability should aim to preserve content between translations. To make good on this requirement, of course, we need a syntactic criterion which plausibly captures when two logical theories say the same things. Happily, there is one.

### Refinement 3: Avoiding perverse “translations”

When we close a set of sentences  $S$  and its logical-expression-mapping image  $t(S)$  under  $L$  and  $L^*$ , two theoretically equivalent logics, we want the results of translating the  $L^*$ -closure of  $t(S)$  into  $L$  to be, in an intuitive sense, *L-equivalent* to the  $L$ -closure of  $S$  (analogously for translating  $L_{u(S)}$  into  $L^*_S$ ). Otherwise, assuming  $L$  and  $L^*$  really are theoretically equivalent, our translation from  $L^*_{t(S)}$  into  $L_S$  has perverted  $S$ 's content (relative to the mapping  $t$ ); we've gained or lost information in passing back and forth between the two logics. It's then only a translation in the *recherché* sense of a minimally adequate embedding of  $L^*_{t(S)}$  into  $L_S$ , not a translation in the fuller sense which would legitimate using intertranslatability as an explication of two logics saying the same thing.

Consider an analogy. Suppose we've what purports to be an English translation of *Saatleri Ayarlama Enstitüsü*. Suppose further we've a mapping between English and Turkish such that, for every sentence in *The Time Regulation Institute*, there's a corresponding sentence in *Saatleri Ayarlama Enstitüsü*. Does this guarantee that everything said in *Saatleri Ayarlama Enstitüsü* is said in *The Time Regulation Institute*? No, obviously not. For example, our mapping might send two subtly different ways of apologizing in Turkish, “Üzgünüm” and “Özür dilerim,” to “I am sorry”. This would be a perverted translation (though we often can't do much better in practice.)

For a truly adequate translation, we need that for every sentence in *Saatleri Ayarlama Enstitüsü*, there's a Turkish-equivalent sentence in the set of Turkish sentences obtained by applying our mapping to *The Time Regulation Institute*. If there isn't, then our translation has perverted the meaning of the Turkish original and, corresponding, the content of the two novels aren't the same. As we want our account of theoretical equivalence between logics to capture when their content is the same, we should demand that the translations between them aren't perverse.

Translations between classical and intuitionistic logic fails this constraint;  $g$ , for instance, effectively takes  $p \vee \neg p$  to  $\neg(\neg p \& \neg\neg p)$ , but while we can derive  $\neg(\neg p \& \neg\neg p)$  from  $p \vee \neg p$  intuitionistically, we cannot so derive  $p \vee \neg p$  from  $\neg(\neg p \& \neg\neg p)$ —that is, they're not intuitionistically equivalent. From an intuitionistic point of view,  $g$  is *perverse* (presuming reasonable mappings between logical expressions).

Let  $S = \{p \vee \neg p\}$ .  $I_S$  then contains all the classical tautologies involving  $p$ .  $p \vee \neg p$ , intuitionistically, is tantamount to the claim that  $p$  is decidable and thus that classical reasoning can be used concerning it.  $g$  purges every classical tautology in  $C_S$  by replacing it with an intuitionistic tautology in  $g(C_S)$ ; no solely classical tautologies remain. In particular,  $g(C_S)$  intuitively says nothing about decidability. So  $I_S$  and  $g(C_S)$  say quite different things, witnessed by the lack of intuitionistic equivalence between them.  $g$  is perverse.

I propose to lay lack of perversion down as a further constraint on theoretical equivalence. So two logics  $L, L^*$  are theoretically equivalent, relative to presumed mappings  $t, u$ , if there are general translations  $a_L: L_G \rightarrow L^*_{t(G)}$  and  $a_{L^*}: L^*_G \rightarrow L_{u(G)}$  where, given any set of sentences  $S$  in the language of  $L$ ,  $L_S$  is  $L$ -equivalent to  $a_{L^*}(L^*_{t(S)})$  and similarly for  $a_L$ . What is it to be  $L$ -equivalent in this context? It's to be such that for every sentence  $\psi$  in  $L_S$  there's a sentence  $\phi$  in  $a_{L^*}(L^*_{t(G)})$  such that  $L_{\{\phi\}} = L_{\{\psi\}}$ . Likewise for  $L^*$ .<sup>17</sup> Essentially, we're using the notion of logical closure to guarantee that the results of translation of the classical theory of  $G$  into an intuitionistic theory is intuitionistically equivalent to the intuitionistic theory of  $G$ .

We can call this constraint *Anti-Perversion*: any pair of logics  $L$  and  $L^*$  that further satisfy this condition will be such that by each of their lights, the  $L$  and  $L^*$  closure of any set of sentences (relative to the mapping between logical constants) will be the same, up to interderivability.<sup>18</sup> This means that any two theoretically equivalent logics  $L, L^*$  will be such that any theory formulated in  $L$  has an  $L$ -equivalent theory formulated in  $L^*$  and vice versa. This is a comforting result; it shows that two logics theoretically equivalent in our refined sense really do say the same thing about various bits of content.

## Upshot

As just pointed out, Wigglesworth's translations are perverse and, in fact, there is no unperverted pair of general translations  $a_I: I_G \rightarrow C_G$  and  $a_C: C_G \rightarrow I_G$ .

*pf.* Presume the identity mapping  $t$  between logical constants of  $C$  and  $I$ . Any general translation  $a_C$  needs to send the sentence  $p \vee \neg p$ , which is a member of  $C_A$  for any set  $A$ , to something intuitionistically weaker since  $C_{\{\}} contains  $p \vee \neg p$  and  $I_{\{\}}$  contains nothing as intuitionistically strong. This holds for all sentences intuitionistically equivalent to  $p \vee \neg p$ . Now let  $G$  be  $\{p \vee \neg p\}$ . Since  $C_{\{\}} = C_{\{p \vee \neg p\}}$  our general translation of  $C_G, a_C(C_G)$  contains only things intuitionistically weaker than  $p \vee \neg p$ . So there will be no sentence  $\phi$  in  $a_C(C_G)$  such that  $I_{\{\phi\}} = I_{\{p \vee \neg p\}}$ .$

Note this reasoning works *mutatis mutandis* against the first way out described in our first refinement or, analogously, when using  $g$  as our  $t$  as well as using it as our translation:  $g$  takes  $p \vee \neg p$  effectively to  $\neg(\neg p \text{ \& } \neg\neg p)$  and, of course,  $C_{\{\}} = C_{\{\neg(\neg p \text{ \& } \neg\neg p)\}}$ . So, again,  $a_C(C_{g(G)})$  doesn't contain any sentence  $\phi$  such that  $I_{\{\phi\}} = I_{\{p \vee \neg p\}}$ . We can also rule out unperverted *nice* translations using a result of Wojcicki (exploited for similar purposes by French (2017) and Humberstone 2000). 'Nice' here means that atomic nonlogical content is mapped to itself, that all logical expressions commute (relative to the presumed mapping), and that the translations are inverse to each other.<sup>19</sup> *Anti-Perversion* is thus robust enough to exclude Wigglesworth-type examples across a range of different refinements.

Nevertheless, there *are* intuitively equivalent pairs of logics which satisfy *Anti-Perversion*; consider  $B$  and  $B^*$  under the  $t^*$  mapping described above. For any



sentence in the translation of  $B^*_{T^*({p \vee \neg p})}$  ( $= B^*_{\{p \leftrightarrow \neg p\}}$ ) back into  $B$ , there will be a sentence in  $B_{\{p \vee \neg p\}}$  which is  $B$ -equivalent to it. And conversely. Consider also our opening example of natural deduction and axiomatic presentations of classical logic.

For a more involved and interesting example, we can use the De Morgan equivalences—for both our connective mapping and our general translation—in order to translate the conjunction negation fragment of classical logic into the disjunction negation fragment and conversely while satisfying *Anti-Perversion*.

Finally, note that we can take a range of interesting nonclassical logics, such as intuitionistic or relevance logic, and show that adding additional axioms yields theoretical equivalence with classical logic. For instance, intuitionistic logic with the addition of  $A \vee \neg A$ , as an axiom, for every piece of non-logical content  $A$  is theoretically equivalent, in our sense, to classical logic. As is the addition of  $\neg \neg A \rightarrow A$  for every piece of non-logical content  $A$ . And, of course, they're theoretically equivalent to each other.<sup>20</sup>

So there is a useful, interesting, and non-trivial syntactic notion of theoretical equivalence, captured by adding *Anti-Perversion* to general intertranslatability (relativizing the whole to a mapping between the logical expressions so as to guarantee that our sets of sentences can be formulated in a “neutral” language) which distinguishes clearly theoretically inequivalent logics like classical and intuitionistic logic while not distinguishing clearly theoretically equivalent logics like the disjunction negation and conjunction negation fragments of classical logic. It might not be a perfect notion of theoretical equivalence—there are well-known bizarre results to using syntactic notions like this to capture theoretical equivalence<sup>21</sup>—but there simply may be no perfect explication of our intuitive ideas about sameness of content. The notion I've articulated serves in many cases to do the work a notion of theoretical equivalence should do in comparing a pair of logics in an antiexceptionalist framework.

Correspondingly, none of this shows that Wigglesworth's suggested notion of theoretical equivalence for logical theories, in terms of categorical equivalence, isn't a useful and interesting one.<sup>22</sup> But I've shown that not all syntactic notions of theoretical equivalence are hopeless in explicating the some form of theoretical equivalence between various logical alternatives.

## Acknowledgments

Author thanks Rohan French, Simon Hewitt, Beau Madison Mount, Michael Rathjen, Gil Sagi, Jared Warren, John Wigglesworth, and Gözde Yıldırım for helpful discussion. Thanks also to discussion with Neil Dewar, whose simultaneously written Dewar (forthcoming), usefully covers similar issues to this article from a philosophy of science angle.

## Notes

- 1 See Woods (2017b) for other worries about anti-exceptionalist methodology.
- 2 See Woods (2017a, 2017b), and Meadows and Weber (2016) for discussion of these issues.
- 3 I'll stick with propositional logic as the problems I'm concerned with are already present there and quantificational theory poses additional difficulties.



- 4 A referee (not the one thanked elsewhere in this paper) suggests this is anti-anti-exceptionalist. Not so; it's rather the standard rejection of treating a logic-as a merely set of theorems. No one seriously conceives of a logic as a set of theorems anymore — since, for instance, some non-equivalent logics don't have any — so this point seems utterly uncontroversial, even given anti-exceptionalism.
- 5 Treating axioms as a zero-case of applying rules. This general notion of derivability is clear enough for our purposes (and can, of course, be spelled out precisely in the usual ways.) Thanks to a helpful referee for encouraging more generality here.
- 6 This constraint is a weakening of the typical requirement that translations commute with logical operations, which isn't obviously reasonable when comparing distinct logics. Adopting the weakened constraint doesn't eliminate all problems (Feferman 2000: A.4); for an immediately pressing one, note that not all logics contain primitive negation operators! We avoid this problem by means of the generalizable commutation condition mooted in the first refinement. See also endnote 13.
- 7 Of course, there's a wealth of material on translations between consequence relations which could also be exploited in this regard. I'll stick with developing Wigglesworth's account, but interested readers are encouraged to consult Wojcicki (1988) and references within.
- 8 French (2017) develops an account of "notational variance" — a slightly stronger notion of equivalence — requiring that an equivalent logics stay equivalent under uniform increases in background logical strength. This is an interesting additional condition to impose, but it's perhaps too stringent for theoretical equivalence. I hope to discuss this elsewhere.
- 9 I focus on correspondence between logical constants; the corresponding correspondence of formation rules is obvious.
- 10 We could also drop this assumption and relativize our translations to an existing pair of mappings between the languages in general; we ignore this complication by presuming a shared field of non-logical content.
- 11 We could narrow down which mapping to presume, drawing on something like the counterpart-theoretic story in Warren (2015a, §. 4) or that the resulting translations be *nice* in the sense spelled out below. Alternatively, we could let presumed mappings vary with our background purpose; after all, in practice some such mapping between logical expressions, often the identity mapping, will be implicitly presumed. 'Absolutized' versions, such as the notion of theoretical equivalence given by existentially or universally quantifying out the relativity of our preferred notion, are either extremely weak or strong, as our example of  $B$  and  $B^*$  shows. I hope to take up these matters elsewhere. Thanks to a helpful referee for discussion.
- 12 This solution, being syntactic, doesn't single out the bizarre theory; it just shows that there are validities in one theory which are countervalidities in the other. But that's all that's needed to carve the two apart. So even though there's a semantic-ish presumption of common content for  $S$  underlying our choice of a presumed mapping, our definition itself is still fully syntactic. This is good news for the anti-exceptionalist.
- 13 We could also require our presumed mapping between logical expressions be *nice* in the sense of preserving non-logical content and guaranteeing commutation for all logical expressions for a naturally more stringent notion of theoretical equivalence. See below.
- 14 Translations between sets of logical expressions can be complex; we might map  $p \vee q$  to  $\neg(\neg p \leftrightarrow \neg q)$ . Details would distract and aren't important here. What I say below is adaptable to the second solution; those who favor the coarse-grained are invited to read the discussion accordingly.

- 15 Again, I follow Wigglesworth's closure-based presentation throughout though we could reframe the discussion using the relevant derivability relation without any loss.
- 16 Wigglesworth's favored translation between classical and intuitionistic logic is general in this sense. Remember that we're suppressing the presumed mappings between logical expressions unless it matters.
- 17 This baroque formulation avoids explicitly appealing to entailment. Wigglesworth reasonably complains about such appeals by Glymour-style accounts of theoretical equivalence in terms of definitional extensions (see Barrett and Halvorson 2016; Glymour 2013; McSweeney 2016). We avoid Wigglesworth's objection by using the same technology he accepts in developing toy syntactic accounts of theoretical equivalence for criticism.
- 18 See Segerberg (1982: 43) for an analogous constraint on syntactic equivalence. As French (2017) notes, Segerberg's formulation of this condition doesn't presume any kind of compositionality (as I do in refinement 1 above), resulting in a weaker notion of equivalence. *Anti-Perversion* also resembles a constraint discussed by Barrett and Halvorson (2016: Definition 5). It differs by requiring that the equivalence between a sentence of intuitionistic logic and the translated version of the corresponding sentence in classical logic hold without any side-premises instead being provable "locally" from each  $I_S$  (as well as avoiding the use of the biconditional since our respective logics might not contain biconditionals, let alone biconditionals that mean the same thing.) This is appropriate to the setting of general translations, as discussed above, and in light of the application to logic itself instead of nonlogical theories. Note that  $p \vee \neg p \leftrightarrow \neg(\neg p \& \neg \neg p)$  is in  $I_{\{p \vee \neg p\}}$ . So the Barrett and Halvorson refinement is insufficient to avoid Wigglesworth's examples, even though it and *Anti-Perversion* are similarly inspired.
- 19 Roughly, Wojcicki shows that  $\neg p$  would have to be translated, by a nice translation, into one of six inequivalent formulas in a supra-intuitionistic logic. Since we can derive  $q$  from  $p$ ,  $\neg p$  classically, we need to be able to derive the translation of  $q$  from  $p$  and the translation of  $\neg p$  in that logic. But none of the six potential translations will do. See *op. cit.* for details.
- 20 See Woods (2017b) for details of this strategy in bolstering the strength of nonclassical logics.
- 21 See Niebergall (2000) and French (2017) for discussion. See also Warren (2015b: objection 4) for discussion of loosening the constraints on interpretation and problems lurking thereabouts.
- 22 Although I worry about the background metatheory for reasons similar to those I've explored in Woods (2017a).

## References

- Barrett, T. W. and H. Halvorson. "Glymour and Quine on Theoretical Equivalence." *Journal of Philosophical Logic* 45 (2016): 467–483.
- Dewar, N. "On Translating Between Logics." *Analysis* (forthcoming)
- Feferman, S. "Does Reductive Proof Theory Have a Viable Rationale?" *Erkenntnis* 53 (2000): 63–96.
- French, Rohan. "Notational Variance and Its Variants." *Topoi* (2017): 1–17.
- Glymour, Clark. "Theoretical Equivalence and the Semantic View of Theories." *Philosophy of Science* 80.2 (2013): 286–297.
- Hjortland, O. T. "Anti-Exceptionalism about Logic." *Philosophical Studies* 174 (2017): 631–658.
- Humberstone, Lloyd. "Contra-Classical Logics." *Australasian Journal of Philosophy* 78.4 (2000): 438–474.

- Mc Sweeney, Michaela. "An Epistemic Account of Metaphysical Equivalence." *Philosophical Perspectives* 30.1 (2016): 270–293.
- Meadows, Toby and Zach Weber. "Computation in Non-Classical Foundations?" *Philosophers' Imprint* 16.13 (2016): 1–7.
- Niebergall, Karl-Georg. "On the Logic of Reducibility: Axioms and Examples." *Erkenntnis* 53.1–2 (2000): 27–61.
- Priest, G. "Revising Logic," in *The Metaphysics of Logic*, edited by P. Rush. Cambridge: Cambridge University Press, 2014, 211–223.
- Russell, G. K. "Metaphysical Analyticity and the Epistemology of Logic." *Philosophical Studies* 171 (2014): 161–175.
- Seegerberg, Krister. *Classical Propositional Operators: An Exercise in the Foundations of Logic*. Oxford: Oxford University Press, 1982.
- Warren, Jared. "Conventionalism, Consistency, and Consistency Sentences." *Synthese* 192.5 (2015a): 1351–1371.
- Warren, Jared. "Talking with Tonkers." *Philosophers' Imprint* 15.24 (2015b): 1–24.
- Wigglesworth, John. "Logical Anti-Exceptionalism and Theoretical Equivalence." *Analysis* 77.4 (2017): 759–767.
- Williamson, T. "Semantic Paradoxes and Abductive Methodology," in *The Relevance of the Liar*, edited by B. Armour-Garb. Oxford: Oxford University Press, 2017.
- Wojcicki, R. *Theory of Logical Calculi: Basic Theory of Consequence Operations*. Dordrecht: Kluwer, 1988.
- Woods, J. "Against Reflective Equilibrium in Logical Theorizing." (2017a) (manuscript)
- Woods, J. "Logical Partisanship." *Philosophical Studies* (2017b) (forthcoming).