LOGICAL ANTI-EXCEPTIONALISM AND THEORETICAL EQUIVALENCE

JOHN WIGGLESWORTH

Anti-exceptionalism about logic takes logical theories to be continuous with scientific theories. Treating logical theories and scientific theories on a par allows one to provide a justification for logic that is compatible with standard criteria for theory choice in the sciences. But if the anti-exceptionalist position understands logical theories in this way, then it becomes open to more general questions about what logical theories are and when logical theories should be treated as equivalent. Though these are familiar questions that arise in philosophical discussions of scientific theories, they are rarely directed toward logical theories. It is the latter question of theoretical equivalence in logic that this paper addresses. To answer this question, while maintaining the spirit of anti-exceptionalism, one should appeal to standard criteria for theoretical equivalence recognized in the philosophy of science. This paper compares two accepted criteria of equivalence for scientific theories — one syntactic and one semantic — in the context of logical anti-exceptionalism, and argues that the syntactic approach leads to undesirable consequences. The anti-exceptionalist should therefore take a semantic approach when evaluating whether logical theories, understood as scientific theories, are equivalent. This paper argues for a particular semantic approach, in terms of categorical equivalence, to determine whether logical theories are equivalent.

There are several varieties of logical anti-exceptionalism, but we focus on a recent formulation of the view given by Timothy Williamson (2017). According to this version of anti-exceptionalism, scientific theories are sets of sentences closed under a relation of logical consequence. Given a set of sentences, S, the *theory* that S generates is the set of sentences $T_S^{\mathcal{L}}$ obtained by closing S under a relation of logical consequence, as given by the logic $\mathcal{L}$. For example, $T_S^{\mathcal{C}}$ is the theory generated by closing S under classical logic. $T_S^{\mathcal{I}}$ is the theory generated by closing S under intuitionistic logic. And so on. The sentences contained in $T_S^{\mathcal{C}}$ are sentences that hold according to the classical theory of S. Similarly for the intuitionistic theory, and for any theory generated from S by closing it under a relation of logical consequence.

Scientific theories are subject to certain criteria. Given two scientific theories, one can compare them with respect to their theoretical virtues, such as strength, simplicity, unifying power, accuracy to the data, etc. How these theories compare with respect to the criteria can help one decide which scientific theory to believe over the other. Anti-exceptionalists argue that logical theories are continuous with scientific theories. As such, they can also be evaluated according to the same criteria for theory choice. For example, given a set S, one could evaluate the classical theory of S against the intuitionistic theory of S to see how they compare with respect to the theoretical virtues. The anti-exceptionalist argues that by allowing for these kinds of comparisons, the standard criteria for scientific theory choice can help one to decide which logical theory to believe. A recent focus of logical anti-exceptionalism has been to articulate precisely how this kind of evaluation can be carried out.[1]

The focus of this paper concerns other criteria by which theories can be evaluated, namely, criteria for theoretical equivalence. If logical theories are continuous with scientific theories, and scientific theories are subject to criteria for theoretical equivalence, it follows that logical theories should also be subject to such criteria. There are two broad approaches to criteria for theoretical equivalence in the philosophy of science, one syntactic and the other semantic. Our question is: Which approach to theoretical equivalence should the logical anti-exceptionalist adopt? We start with the syntactic approach.

Very roughly, two theories are syntactically equivalent iff they are intertranslatable. The idea can be traced at least back to Quine (1975), who also happened to be an early proponent of logical anti-exceptionalism (1951, 1981). Quine's original proposal was flawed, but one way to capture the spirit of Quine's account uses the syntactic notion of relative interpretability. Filling in some of the details, an interpretation is a mapping τ from the language of theory T to the language of theory T' that satisfies certain conditions. Under the mapping, each formula ϕ in the language of T is associated with another formula $\tau(\phi)$, which is ϕ's interpretation in the language of T'. Of course, not just any mapping will

[1] See Hjortland (2017), Priest (2014, 2016), Russell (2014, 2015), Williamson (2017), among others.

do. For the mapping to be considered an interpretation, it must commute with the logical connectives and quantifiers, so that $\tau(\phi \wedge \psi) := \tau(\phi) \wedge \tau(\psi), \tau(\exists x \phi) := \exists x \tau(\phi)$, etc. Theory T is interpretable in T' iff there is an interpretation τ such that if a sentence $\phi \in T$, then $\tau(\phi) \in T'$. In other words, if ϕ holds according to theory T, then ϕ's interpretation, $\tau(\phi)$, holds according to theory T'. If there are interpretations $\tau : T \to T'$ and $\sigma : T' \to T$, then T and T' are mutually interpretable.[2]

There are other syntactic approaches to theoretical equivalence. Clark Glymour (1970, 1977, 1980) proposes an account in terms of definitional equivalence.[3] Definitional equivalence tries to capture an equivalence between theories that are formulated in different signatures. For a simple example, consider equivalent formulations of group theory with either a binary function symbol, $\cdot$, and a constant symbol, e, or with a binary function symbol, $\cdot$, and a unary function symbol, $^{-1}$. Different formulations of group theory can be given in these signatures, but they should yield equivalent theories of groups. Definitional equivalence tries to capture this idea by extending each theory with definitions of the symbols not in its signature, but which occur in the other theory. The original theories are then definitionally equivalent iff their definitional extensions are logically equivalent.

Definitional equivalence is an interesting and powerful notion of theoretical equivalence in its own right. For example, if one has already settled on classical logic as the background logical theory, and then considers two classical theories $T_S^{\mathcal{C}}$ and $T_{S'}^{\mathcal{C}}$ (formed by closing two sets of sentences, S and S', under classical logic), one can show that if these theories are definitionally equivalent, then they are mutually interpretable. That is, if the definitional extensions of $T_S^{\mathcal{C}}$ and $T_{S'}^{\mathcal{C}}$ are (classically) logically equivalent, then $T_S^{\mathcal{C}}$ and $T_{S'}^{\mathcal{C}}$ are mutually interpretable. In fact, if the theories have disjoint signatures, then they are definitionally equivalent iff they are mutually interpretable. So, when classical logic is assumed, there is an interesting correspondence between definitional equivalence and mutual interpretability.[4] However,

[2]This is still somewhat imprecise, but will do for our purposes. For the full details, see Barrett and Halvorson (2016), §4.

[3]Though Glymour was one of the first to apply definitional equivalence to the philosophy of science, the notion was already familiar to logicians. Many thanks to an anonymous referee for noting that both a syntactic version and a semantic correlate of definitional equivalence can be found in de Bouvère (1965), though the discussion there is limited to classical theories.

[4]For proofs of these results, see Barrett and Halvorson (2016), Theorems 1 and 2.

the criterion that definitional equivalence provides is arguably off limits to the logical anti-exceptionalist, who is trying to settle on a logical theory. The anti-exceptionalist cannot appeal to definitional equivalence because it directly invokes the notion of logical equivalence, which is something that is at issue in the anti-exceptionalism debate. So we focus on the mutual interpretability criterion.

The logical anti-exceptionalist may want an account of when two logics, $\mathcal{L}_1$ and $\mathcal{L}_2$, yield equivalent logical theories. A natural syntactic condition for theoretical equivalence, given in terms of mutual interpretability, presents itself: $\mathcal{L}_1$ and $\mathcal{L}_2$ yield equivalent logical theories iff for all S, $T_S^{\mathcal{L}_1}$ is mutually interpretable with $T_S^{\mathcal{L}_2}$. This syntactic account of theoretical equivalence is fine-grained in the sense that the mutual interpretability condition must hold for *every* set of sentences S. It is therefore able to distinguish between logical theories that overlap to a high degree. It is hard to imagine syntactic criteria that are more fine-grained, short of requiring that $T_S^{\mathcal{L}_1} = T_S^{\mathcal{L}_2}$ for all S. However, in the context of logical anti-exceptionalism, both the identity criterion and the weaker mutual interpretability criterion set the bar too high, as they classify formulations of, e.g., classical logic, given in terms of different logical constants, as distinct. Under the identity criterion, this is straightforward. In the case of mutual interpretability, it happens because an interpretation τ is required to commute with all of the connectives and quantifiers. But this is inappropriate when comparing logical theories that use different logical constants. A standard way to weaken the notion of interpretability is to require only that the mapping τ commutes with negation, so that $\tau(\neg\phi) := \neg\tau(\phi)$ (see Feferman 2000). Call a mapping $\tau : T \to T'$ that satisfies this weaker condition a translation, and let T and T' be equivalent iff there are translations $\tau : T \to T'$ and $\sigma : T' \to T$. In other words, two theories are equivalent when they are intertranslatable.

While mutual interpretability is too fine-grained, it turns out that intertranslatability isn't fine-grained enough. Intertranslatability, as a sufficient condition for theoretical equivalence, forces the anti-exceptionalist to say that some logical theories are equivalent when, intuitively, they are not. A clear example is given by classical and intuitionistic logic. Given a set of sentences, S, one can compare the classical and intuitionistic theories generated by S, respectively, $T_S^{\mathcal{C}}$ and $T_S^{\mathcal{I}}$. Every intuitionistic consequence of S is also a classical consequence of S, and so $T_S^{\mathcal{I}} \subseteq T_S^{\mathcal{C}}$. But the reverse usually isn't true. There are trivial cases where the classical and intuitionistic theories of S coincide, for example when $S = T^{\mathcal{C}}_S$. Ignoring these trivial cases, however, $T^{\mathcal{I}}_S$ is usually a *proper* subset of $T^{\mathcal{C}}_S$. In most cases, there are some sentences that hold according to the classical theory that do not hold according to the intuitionistic theory.

But, as is well-known, there is a mapping, γ, from the classical theory to the intuitionistic theory. In fact there are several, but since we only need one, we can take the familiar Gödel-Gentzen mapping from classical logic to intuitionistic logic, which commutes with negation (see Troelstra and Schwichtenberg 2000, §2.3). For any S and any ϕ, if $\phi \in T^{\mathcal{C}}_S$, then $\gamma(\phi) \in T^{\mathcal{I}}_S$. For the classical and intuitionistic theories to be intertranslatable, we need a mapping back from $T^{\mathcal{I}}_S$ to $T^{\mathcal{C}}_S$. As $T^{\mathcal{I}}_S \subseteq T^{\mathcal{C}}_S$, the identity mapping, ι, such that $\iota(\phi) = \phi$ will do. For any S and any ϕ, if $\phi \in T^{\mathcal{I}}_S$, then $\iota(\phi) = \phi \in T^{\mathcal{C}}_S$.

For any base set of sentences, the classical and intuitionistic theories generated from those sentences are intertranslatable. So classical logic and intuitionistic logic satisfy the syntactic criterion for theoretical equivalence given in terms of intertranslatability. If the anti-exceptionalist, who takes logical theories to be continuous with scientific theories, evaluates theoretical equivalence in terms of intertranslatability, then the anti-exceptionalist must say that classical and intuitionistic logic are equivalent logical theories. But, historically at least, these logics have dominated the debate as competitors for the title of *the* correct notion of logical consequence, a debate that continues today.[5] If there are two logics that should not be considered equivalent, it's these two.

In fact, it's even more serious, as the Gödel-Gentzen translation is also a translation from classical logic to minimal logic. Again, the identity mapping translates the other way, so given any set of sentences, the classical, intuitionistic, and minimal theories generated from that set are intertranslatable. The logical anti-exceptionalist would then be forced to say that all three are equivalent logical theories, though intuitively this should not be so.[6]

[5]See, e.g., Rumfitt (2015). It should be noted that anti-exceptionalism is a position that is available to the logical pluralist as well (see Hjortland 2017). The main argument of this paper applies to both the logical monist and the logical pluralist.

[6]There are also interesting translations between other logics, including modal logics, though we do not focus on these examples here. See Pelletier and Urquhart (2003).

The syntactic approaches to theoretical equivalence, as given by the notions of identity, mutual interpretability, and intertranslatability between theories, are therefore unavailable to the anti-exceptionalist. There are, however, semantic alternatives that do a better job of distinguishing between logical theories. The most promising semantic account, given by the notion of categorical equivalence, focuses on categories of models of theories. Two theories are categorically equivalent iff their respective categories of models are equivalent. Generally, two categories, **C** and **D**, are equivalent iff there are functors $F : \mathbf{C} \to \mathbf{D}$ and $G : \mathbf{D} \to \mathbf{C}$ between them whose compositions are naturally isomorphic to the categories' respective identity functors.

There are other semantic approaches to theoretical equivalence, approaches that do not appeal to category theory. For example, one could take theories to be equivalent when they have identical classes of models, or equinumerous classes of models, or pointwise isomorphic classes of models. But Hans Halvorson (2012) has shown that these criteria are inadequate in different ways. The first criterion is too strict, failing to allow intuitive cases of equivalent theories. The second criterion is too liberal, failing to distinguish between theories that are intuitively not equivalent. And the third criterion fails in both ways. In search of an adequate semantic notion of theoretical equivalence, we focus on categorical equivalence, which is becoming a standard notion of theoretical equivalence in the philosophy of science (see, e.g., Halvorson 2016 or Weatherall 2016).

Filling in some details of the categorical account, consider a set of sentences S and the theories, $T_S^{\mathcal{L}_1}$ and $T_S^{\mathcal{L}_2}$, generated by S according to the logics $\mathcal{L}_1$ and $\mathcal{L}_2$. Assuming that there is a suitable model theory for the logic $\mathcal{L}_1$, the theory $T_S^{\mathcal{L}_1}$ has a category of models, $\mathbf{Mod}_{\mathcal{L}_1}(S)$. The objects of the category are the models of S, or equivalently, the models of $T_S^{\mathcal{L}_1}$ (as understood by the model theory for $\mathcal{L}_1$), and the arrows are homomorphisms between models. Similarly for $\mathcal{L}_2$. There is then a categorical equivalence between $T_S^{\mathcal{L}_1}$ and $T_S^{\mathcal{L}_2}$ iff there are two functors $F : \mathbf{Mod}_{\mathcal{L}_1}(S) \to \mathbf{Mod}_{\mathcal{L}_2}(S)$ and $G : \mathbf{Mod}_{\mathcal{L}_2}(S) \to \mathbf{Mod}_{\mathcal{L}_1}(S)$ such that $FG \cong 1_{\mathbf{Mod}_{\mathcal{L}_1}(S)}$ and $GF \cong 1_{\mathbf{Mod}_{\mathcal{L}_2}(S)}$. That is, two theories are categorically equivalent iff their respective categories of models are equivalent. For the anti-exceptionalist, the notion of categorical equivalence suggests a general criterion for determining when two logics yield equivalent logical theories: $\mathcal{L}_1$ and $\mathcal{L}_2$ yield equivalent logical theories iff for all S, $\mathbf{Mod}_{\mathcal{L}_1}(S)$ is equivalent to $\mathbf{Mod}_{\mathcal{L}_2}(S)$.

The question is whether categorical equivalence is strong enough to discriminate between logical theories that should intuitively be distinct. In the case of classical and intuitionistic logic, the answer is yes. To see this, consider a set of sentences S and the classical and intuitionistic theories that S generates. The classical theory $T_S^{\mathcal{C}}$ has a category of models, $\mathbf{Mod}_{\mathcal{C}}(S)$. The objects of the category are the Tarskian models of S, familiar from standard model theory for first order classical logic; the arrows of the category are homomorphisms between Tarskian models. The intuitionistic theory $T_S^{\mathcal{I}}$ also has a category of models, $\mathbf{Mod}_{\mathcal{I}}(S)$, though in this case the models can be thought of as Kripke models, familiar from standard Kripke semantics for first order intuitionistic logic.[7]

For most sets of sentences, S, $\mathbf{Mod}_{\mathcal{C}}(S)$ and $\mathbf{Mod}_{\mathcal{I}}(S)$ are not categorically equivalent. Essentially, the reason is that, for most S there are "more" Kripke models of S than there are Tarskian models of S. The excess is due to the fact that there are Kripke models that fail to satisfiy some classical logical truths, e.g., some instances of $\phi \vee \neg\phi$. There are no such Tarskian models. However, there are some cases where an equivalence does hold. In fact, it can be shown that, for any S, a functor $F : \mathbf{Mod}_{\mathcal{C}}(S) \rightarrow \mathbf{Mod}_{\mathcal{I}}(S)$ is an equivalence iff there are no non-trivial homomorphisms between classical models of S. In other words, $\mathbf{Mod}_{\mathcal{C}}(S)$ and $\mathbf{Mod}_{\mathcal{I}}(S)$ are equivalent iff the category $\mathbf{Mod}_{\mathcal{C}}(S)$ of classical models of S is discrete. Theories with discrete categories of classical models include, for example, the theory of the n element linear order, which effectively just says "$1 \leq 2, 2 \leq 3, ..., n-1 \leq n$". It follows that the classical theory of the n element linear order and the intuitionistic theory of the n element linear order are categorically equivalent.

Nevertheless, the existence of such S does not threaten the anti-exceptionalist's ability to distinguish between classical and intuitionistic logic as logical theories. The categorical account of theoretical equivalence requires only that there be some S such that $\mathbf{Mod}_{\mathcal{C}}(S)$ and $\mathbf{Mod}_{\mathcal{I}}(S)$ are not equivalent, which there clearly is (a particular example is discussed below). The categorical account is therefore

[7] It is, however, slightly more complicated. $\mathbf{Mod}_{\mathcal{I}}(S)$ is a topos of presheaves from a category **K** of possible worlds (together with an accessibility relation) to the category **Set** of sets. The details are not important for our purposes here.

a more suitable account of theoretical equivalence for the logical anti-exceptionalist, as it is able to distinguish between classical and intuitionistic logic as logical theories, a distinction that the intertranslatability account is unable to make.

In fact, the logical anti-exceptionalist might be able to adopt a weaker categorical condition for the equivalence of logical theories. The proposed criterion allows the anti-exceptionalist to distinguish between classical and intuitionistic logic because there are sets S such that $\mathbf{Mod}_{\mathcal{C}}(S)$ and $\mathbf{Mod}_{\mathcal{I}}(S)$ are not equivalent. An interesting case of this is when S is empty. The classical theory generated by the empty set is the set of all classically valid sentences. Similarly, the intuitionistic theory generated by the empty set is the set of all intuitionistically valid sentences. One might call these the "pure" theories of classical and intuitionistic logic.

Looking at these theories categorically, the category of classical models of the empty set is simply the category of *all* classical models. This category is not discrete. It follows from the above that the category of models of the pure theory of classical logic and the category of models of the pure theory of intuitionistic logic are not equivalent. The logical anti-exceptionalist might then consider adopting a weaker condition for equivalence between logical theories: $\mathcal{L}_1$ and $\mathcal{L}_2$ yield equivalent logical theories iff $\mathbf{Mod}_{\mathcal{L}_1}(\emptyset)$ is equivalent to $\mathbf{Mod}_{\mathcal{L}_2}(\emptyset)$. Of course, the weaker condition may be more vulnerable to counterexamples, logics that are judged equivalent under this condition, but which intuitively are not (and which may not be judged equivalent by the stronger categorical criterion that considers all S). One might also find the weaker condition unsatisfying, because, as Williamson (2017, §2) points out, comparing pure theories only compares theoremhood as opposed to full consequence relations. Nevertheless, the weaker condition does allow the anti-exceptionalist to distinguish between classical and intuitionistic logic, if a logical theory is understood as the "pure" theory generated by that logic from the empty set.

In summary, though there are two general approaches — one syntactic and one semantic — to theoretical equivalence in the philosophy of science, the logical anti-exceptionalist should prefer the semantic approach. The semantic approach is preferable, because the standard syntactic approach in terms of intertranslatability forces the anti-exceptionalist to say that classical logic and intuitionistic logic are equivalent logical theories, though intuitively they are not. Alternative semantic approaches can be given in terms of categorical equivalence. The logical anti-exceptionalist may choose to formulate a categorical criterion for theoretical equivalence using either a strong or a weak condition. According to the strong condition, two logics yield equivalent logical theories iff for every set of sentences, S, the categories of models of S (as understood by the model theories of the two logics) are equivalent. Alternatively, there is a weaker condition that only compares the categories of models generated by the empty set, thereby comparing the categories of models of the "pure" theories of the relevant logics. Both formulations entail that classical and intuitionistic logic are not equivalent logical theories, a welcome consequence for the logical anti-exceptionalist.[8]

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft (DFG), under the research project *Mathematics: Objectivity by Representation* [grant number LE 2500/4-1].

REFERENCES

Barrett, T. W. and Halvorson, H. (2016). Glymour and Quine on theoretical equivalence. *Journal of Philosophical Logic*, 45:467–483.

de Bouvère, K. (1965). Synonymous theories. In Addison, J. W., Henkin, L., and Tarski, A., editors, *The Theory of Models: Proceedings of the 1963 International Symposium at Berkeley*, pages 402–406. North Holland, Amsterdam.

Feferman, S. (2000). Does reductive proof theory have a viable rationale? *Erkenntnis*, 53:63–96.

Glymour, C. (1970). Theoretical realism and theoretical equivalence. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Association*, pages 275–288.

Glymour, C. (1977). The epistemology of geometry. *Noûs*, 11:227–251.

Glymour, C. (1980). *Theory and Evidence*. Princeton University Press, Princeton, NJ.

Halvorson, H. (2012). What scientific theories could not be. *Philosophy of Science*, 79:183–206.

[8] I wish to thank Steve Awodey, Laurenz Hudetz, Hannes Leitgeb, Luis Rosa, Georg Schiemer, Dimitris Tsementzis, and several anonymous referees for helpful comments and discussions of this material.

Halvorson, H. (2016). Scientific theories. In Humphreys, P., editor, *Oxford Handbook of Philosophy of Science*, pages 585–608. Oxford University Press, Oxford.

Hjortland, O. T. (2017). Anti-exceptionalism about logic. *Philosophical Studies*, 174:631–658.

Pelletier, F. J. and Urquhart, A. (2003). Synonymous logics. *Journal of Philosophical Logic*, 32:259–285.

Priest, G. (2014). Revising logic. In Rush, P., editor, *The Metaphysics of Logic*, pages 211 – 223. Cambridge University Press, Cambridge.

Priest, G. (2016). Logical disputes and the a priori. *Logique et Analyse*, 59:347–366.

Quine, W. V. (1951). Two dogmas of empiricism. *The Philosophical Review*, 60:20 – 43.

Quine, W. V. (1975). On empirically equivalent systems of the world. *Erkenntnis*, 9.

Quine, W. V. (1981). *Philosophy of Logic*. Harvard University Press, Cambridge, MA, 2nd edition.

Rumfitt, I. (2015). *The Boundary Stones of Thought*. Clarendon Press, Oxford.

Russell, G. (2015). The justification of the basic laws of logic. *Journal of Philosophical Logic*, 44:793–803.

Russell, G. K. (2014). Metaphysical analyticity and the epistemology of logic. *Philosophical Studies*, 171:161–175.

Troelstra, A. S. and Schwichtenberg, H. (2000). *Basic Proof Theory*. Cambridge University Press, 2nd edition.

Weatherall, J. O. (2016). Are Newtonian gravitation and geometrized Newtonian gravitation theoretically equivalent? *Erkenntnis*, 81:1073–1091.

Williamson, T. (2017). Semantic paradoxes and abductive methodology. In Armour-Garb, B., editor, *The Relevance of the Liar*. Forthcoming from Oxford University Press, Oxford.