Philosophical Papers
Vol. 34, No. 1 (March 2005): 44 74

What is Metaphysical Equivalence?

Kristie Miller

Abstract: Theories are metaphysically equivalent just if there is no fact of the matter that could render one theory true and the other false. In this paper I argue that if we are judiciously to resolve disputes about whether theories are equivalent or not, we need to develop testable criteria that will give us epistemic access to the obtaining of the relation of metaphysical equivalence holding between those theories. I develop such 'diagnostic' criteria. I argue that correctly inter-translatable theories are metaphysically equivalent, and what we need are ways of determining whether a putative translation is correct or not. To that end I develop a number of tools we can employ to discern whether a translation is a correct one.

1 Introduction

What does it mean to say that two theories are equivalent; that they amount to the same thing; that we have a case of mere verbal disagreement? How do we know when this is the case? Let us call the strong equivalence relation that holds between such theories metaphysical equivalence. Then intuitively, theories are metaphysically equivalent just if they use different language to describe the same underlying reality. That is, if the features of the world described by one theory are identical to the features of the world described by the other theory—there are no facts that could render one theory true and the other false.

Frequently it is argued, or at least hinted, that certain theories, including metaphysical theories, are equivalent. Presentists are sometimes accused of espousing a metaphysics that is either trivially false, or equivalent to eternalism. It is suggested that although it *appears* that presentists and eternalists are making substantially different claims about what exists, specifically about whether any temporal locations other than the present exist, in fact this might be mere verbal disagreement. Perhaps the presentist and the eternalist mean something

different by 'everything' so that when the presentist says that everything that exists, exists in the present, and the eternalist denies this, they are not making contradictory claims at all.[1]

Equally, there are those who wonder whether what appear to be incompatible ontological claims are really incompatible at all. Putnam famously argues that in a world in which there exists three simples, the mereological universalist, who holds that there exist seven objects in that world, and the mereological nihilist, who holds that there exist only three objects in that world, hold views that are equivalent. In terms of the universalist language, it is true that there exist seven objects, and false that there exist three objects, and the reverse is true in terms of the nihilist language. Each of these languages are equally good ways of talking about the world, and since there is no 'absolute' framework from which we can talk about the world as it is in itself, it makes no sense to try to say that there are 'really' seven objects not three, or three objects not seven.[2] So, following Putnam, we might argue that mereological universalism, mereological nihilism[3]—the view that no combination of simples composes any composite object—and mereological non-universalism—the view that only some combinations of simples compose objects—are really equivalent theories.

Perhaps these theories are equivalent because 'object' is defined by its role in the entire ontological theory in which it features, and thus universalists, nihilists and non-universalists simply mean something different when they deny or affirm that certain arrangements of simples compose some object.[4] Or perhaps, as has recently been argued by Eli Hirsch, the existential quantifier can have multiple meanings. On some interpretations of 'there exists a thing' certain sentences will be true, and on other interpretations they will be false,[5] thus explaining how

1 Ted Sider raises this possibility in his (1999).
2 Putnam, H. (1987).
3 For a discussion of nihilist or 'near-nihilist' views, see Merricks (2000 and 2001); van Inwagen (1990) and Unger (1979).
4 Cf. Putnam, H. (1988).
5 Hirsch (2002).

universalists, non-universalists and nihilists can all speak truly.

That's not all. Recently Storrs McCall and E.J. Lowe have claimed that two competing accounts of persistence over time, three dimensionalism and four dimensionalism, are equivalent.[6] Three dimensionalism is the view according to which objects have only three spatial dimensions and persist through time by being wholly present whenever they exist. Four dimensionalism is the view that objects have both spatial and temporal dimensions, and persist through time by being the mereological fusion of temporal parts. McCall and Lowe argue that objects can correctly be described in either three or four dimensional language, and that these descriptions are 'equivalent' in the sense that they are inter-translatable and there is no fact of the matter in the world that makes one of the descriptions true and the other false.[7]

Alan Sidelle goes further. He argues that in matters of persistence and ontology there are various 'packages' of views each of which preserves a different set of folk intuitions and theoretical ideals. According to Sidelle, there is no fact of the matter which of these packages truly describes the world. Rather, he argues, they are merely different ways of making coherent our various intuitive judgements and theoretical ideals.[8]

Finally, in *Platonism and Anti-Platonism in Mathematics*, Mark Balaguer controversially argues that there is only one viable version of Platonism: full-blooded Platonism, and one viable version of non-Platonism: fictionalism, and that there is a sense in which these two theories are equivalent.[9] Balaguer argues that for all practical purposes there is no difference between Platonism and fictionalism: both offer exactly the same vision of mathematical practice. The only difference between the two theories is with respect to ontology: Platonists maintain that mathematical objects exist, and fictionalists maintain that they do not.

6 McCall (1994); McCall and Lowe (2003).
7 Lowe and McCall (2003) pg 118.
8 Sidelle (2002).
9 Balaguer (1998).

Balaguer argues, however, that this is no real difference at all, for there is simply no fact of the matter as to whether mathematical objects exist or not.[10] As Balaguer puts it 'the metaphysical question of whether there exist any abstract objects is empty, but the two conclusions [Platonism and fictionalism] cash this out in different ways.'[11] But then if Platonism and fictionalism provide the same account of mathematical practice, and differ only in matters of ontology, then if the ontological debate is empty, it would seem that Platonism and fictionalism are equivalent.

In all of these examples we find claims about theories being in some sense equivalent. In each case the underlying intuition is that theories are equivalent in this sense if somehow they are describing the same underlying reality: if there is no fact that could make one theory true and the other false. This suggests that we might define metaphysical equivalence in terms of sets of worlds being identical. We might say that any two theories x and y are metaphysically equivalent iff the set of the worlds in which x is true, is identical to the set of worlds in which y is true, and the set of worlds in which x is false, is identical to the set of worlds in which y is false. But this will not do. For then any two necessarily true theories will turn out to be metaphysically equivalent. If we are considering theories about the concrete elements of the world, we could instead appeal to features of worlds in virtue of which theories are true. For theories of the concrete, we might say the following:
Two theories x and y are metaphysically equivalent iff:

(i) the set of worlds in which x is false is identical to the set of worlds in which y is false and

(ii) the set of worlds in which x is true is identical to the set of worlds in which y is true and

(iii) in every world w in which x is true, the features of w in virtue of which x is true, are identical to the features of w in virtue of which y is true and

10 Balaguer (1998) pg 179.
11 Balaguer (1998) pg 151.

(iv) in every world w1 in which x is false, the features of w1 in virtue of which x is false, are identical to the features of w1 in virtue of which y is false.

Ultimately though, when we are considering theories about necessarily existing abstracta or concreta, talk of 'features of worlds' boils down to talk about truth makers (choose your favourite account of truth makers). A theory about God and a theory about the number 3 are not equivalent, and they are not equivalent because they have different truth makers. At their heart, claims about metaphysical equivalence are claims about truth makers: any two theories are metaphysically equivalent just in case they have the same truth makers. Now, this is not very revealing. Knowing that two theories are equivalent just in case they have the same truth makers does not seem to help in determining whether any two particular theories are equivalent or not. But then, we should not expect a definition to do that job for us. What we need is some sort of apparatus within which either to argue that certain theories are metaphysically equivalent, or to dispute such a claim.

In this paper I develop what I will call diagnostic criteria of metaphysical equivalence. These are criteria against which putatively equivalent theories can be measured, and it can be determined whether they are equivalent or not. They are also criteria that allow points of dispute between parties who disagree as to whether certain theories are metaphysically equivalent or not, to be clearly located and their significance to be noted. I formulate these general criteria in part by considering a number of the cases described above. Though I consider these claims in order to explore the issues surrounding metaphysical equivalence, I do not, in general, attempt to come down on one side or the other with respect to any of these particular debates. Rather, I want to use these cases to raise questions and formulate diagnostic criteria.

2 Inter-translatability

One obvious criterion of metaphysical equivalence is inter-translatability.

Putnam, for instance, holds that any two theories[12] that are empirically adequate and inter-translatable, such as, in his view, universalism and nihilism, are metaphysically equivalent. The thesis of quantifier variance—the claim that the existential quantifier has, or could have different meanings—goes some way towards explaining how this translation could work. Since we define the logical constants by describing their roles in determining the truth conditions of sentences, we can define different meanings of the existential quantifier by stipulating different truth conditions for sentences containing the quantifier. So, for instance, Hirsch argues that the meaning of the quantifier employed by the universalist is such that sentences of the form 'there exists something composed of the F-thing and the G-thing' is true just in case 'the F-thing' refers to something and 'the G-thing' refers to something. On the other hand, the non-universalist's sense of the quantifier will be such that sentences of the form 'there exists something composed of the F-thing and the G-thing' will be true just in case 'the F-thing' and 'the G-thing' refer to things that are connected in certain special ways.[13] While for the nihilist, of course, a sentence of the form just described will never be true, since no F-thing or G-thing ever compose anything.

McCall and Lowe too place the weight of their argument on the claim that three and four dimensionalism are inter-translatable (though they do not show exactly how this translation is to work).[14] Roughly speaking, the idea is that if today I point to a rabbit, described in four dimensional language I point to a temporal part of some four dimensional rabbit. Described in three dimensional language I point to a wholly present

12 I will talk of 'theories' being inter-translatable. I understand theories to be sets of sentences (which make various claims about, or purport to describe the world) in some sub-language. Thus the claim that theories are inter-translatable is just the claim that the sentences of the theories are inter-translatable.

13 Hirsch (2002) pg 54.

14 For a defence of the claim of inter-translatability and an account of this translation see my 'The metaphysical equivalence of three and four dimensionalism' forthcoming in *Erkenntnis*.

rabbit that is strictly identical to some rabbit I pointed to yesterday. In three dimensional language, the claim that the wholly present rabbit is strictly identical yesterday and today, is equivalent to the claim in four dimensional language, that the temporal part of the rabbit yesterday is topologically connected to the temporal part of the rabbit today.[15]

We can see the same notion of translatability arising in Balaguer's arguments. Zalta and Colyvan suggest that one way to understand Balaguer's claim that Platonism and fictionalism are equivalent with respect to all matters but ontology, is to understand Platonism and fictionalism as two interpretations of a single formalism: 'xAx'.[16] Then the disagreement between the two resides in the fact that the Platonist reads the formalism such that the quantifier has existential import, and reads the predicate 'A' as 'abstract', while the fictionalist reads the quantifier as lacking existential import, and reads the predicate 'A' as 'fictional'. Hence 'xAx' under one interpretation reads, 'there exist abstract objects', and under the other interpretation reads, 'there are fictions'. Since the debate about whether there exist abstract objects or not is, according to Balaguer, an empty one, we can see these two interpretations as two ways of explicating this ontological emptiness, thus explaining how Platonism and fictionalism provide the same understanding of mathematical practice and yet appear to differ so radically on ontological matters.

The question is whether Putnam is right to hold that if two theories are inter-translatable and empirically adequate, this is sufficient for concluding that they are metaphysically equivalent. There seems to be a sense in which Putnam is right. Consider the intuitive sense of a 'correct' translation, according to which a translation is correct just if it truly 'gets it right'. In this intuitive sense, it is difficult to see how a correct inter-translation between theories could fail to be sufficient for metaphysical equivalence. For if the translation is truly correct in this strong sense,

15 McCall (1994) pg 216.
16 Zalta and Colyvan (1999).

then surely the two theories are indeed describing the same underlying reality in different terminology.

That this is so is strongly supported by the fact that those who reject some particular claim of metaphysical equivalence between two theories invariably argue that the theories in question are not correctly inter-translatable, rather than arguing that despite the fact that they are so translatable, this is insufficient grounds to conclude that they are equivalent. For instance, Ted Sider, Peter van Inwagen and Trenton Merricks all resist the idea that the existential quantifier has multiple interpretations, and thus resist the claim that nihilism, non-universalism and universalism are equivalent.[17] And presumably fictionalists and Platonists who hold that their theories are not equivalent think that this is so because they think that the Platonist's sentences that quantify over abstract objects cannot correctly by mapped to the fictionalist's sentences that do not so quantify.

So I think we can say that if two theories are correctly inter-translatable, then this entails that they are metaphysically equivalent. The real question then, is whether we can define some function that maps the sentences of one theory onto the sentences of the other theory such that that mapping counts as a correct mapping, and thus counts as a correct translation between those theories.[18] Further, given that we can define such a function, what epistemic access can we have to whether or not on any particular occasion, a translation is indeed a correct one. Let us begin with the first question first.

A translation function is a function that maps sentences of one theory onto sentences of some other theory. One sort of translation function is a function that maps the sentences of one theory onto the sentences of another theory, just when those sentences are assertible under the same possible situations. Let us call this an assertibility mapping. Now, presumably everyone can agree that, for instance, where the four

17 Sider (2001) see the introduction; Merricks (2001) chapter 1 and van Inwagen (2002).
18 Quine (1975).

dimensionalist will utter 'there is a rabbit stage', under the same conditions the three dimensionalist will utter 'there is a wholly present rabbit'. So too under the same conditions the nihilist will utter 'there exists a dog-wise arrangement of simples', while the universalist (and most non-universalists) will utter 'there exists a dog.' In each of these cases we have an assertibility mapping: a function that maps the sentences of theory A that are, by the lights of theory A correctly assertible, onto the sentences of theory B that are, by the lights of theory B correctly assertible, when and only when those sentences are assertible under the same possible situations.

Of course, the existence of an assertibility mapping does not show that the theories in question are correctly inter-translatable and thus metaphysically equivalent. Nihilists think that universalists and non-universalists are wrong when they assert that 'there exists a dog' just as fictionalists think that Platonists are wrong to assert that there exists abstract objects. But any two theories can only be metaphysically equivalent if the sentences they assert under the same circumstances have the same truth values, that is, if the assertibility mapping is truth preserving. This suggests that we say that an assertibility mapping is a correct mapping—a correct translation—only if it preserves the truth values of the sentences in each of the theoretical languages. Now prima facie we might think that an assertibility mapping that preserves truth values ought to count as a correct translation. If the nihilist and the universalist assert 'there is a dog-wise arrangement of simples' and 'there is a dog' on all and only the same actual and possible occasions, and if on those occasions either both sentences are true, or both false, then don't we have a correct translation between those sentences?

Well perhaps we do. But if we are right that we have a correct translation just when we have a case of metaphysical equivalence, then it cannot be that an assertibility mapping that preserves truth is sufficient for a correct translation. For consider. Suppose that God necessarily exists. Suppose further that whenever you assert 'x' I assert 'x and God exists'. (Or if you are sure that God is contingent, suppose when you

assert 'x' I assert 'x and p or not p'.) Then it would seem that there is an assertibility mapping between our sentences that is truth preserving. But surely we would be wrong to conclude that 'x' in your mouth, is a correct translation of 'x and God exists' in my mouth.

The problem here, of course, is that we not only want the assertibility mapping to be truth preserving, we want it to be truth preserving in virtue of the same truth makers. That is why a genuinely correct translation between theories entails that they are equivalent. So let us say that we have a correct translation between theories just if there is an assertibility mapping that is truth preserving and where it preserves truth in virtue of the same truth makers. But now we are again faced with the problem that since frequently we have no access to truth makers, it follows that we do not always have access to whether some assertibility mapping is a correct translation. Thus in these cases we will have no access to whether or not the theories in question are equivalent.

Clearly what we need are some tools to help us in deciding whether on any occasion we have a correct translation. So let us begin by defining what I will call a *practical translation*, where a practical translation is an assertibility mapping that is truth preserving. Since a translation is correct only if it is also a practical translation, on any occasion if we can show that some mapping is not a practical translation, then we can conclude that it is not a correct translation and thus that there is no equivalence. That is, we can at least go some way towards showing how we would falsify some claim about putative metaphysical equivalence. Of course, that we have a practical translation does not entail that we have a correct translation. In section 2.4 and 3 we will consider what if anything licences the move from holding that a translation is practical, to that it is correct.

2.1 Diagnosing a Practical Translation: Empirical Equivalence

How are we to know whether or not we have a practical translation? Well, we have a practical translation just if we have an assertibility mapping that is truth preserving. So one way of diagnosing the existence of a

practical translation is via empirical equivalence. Now, if we had defined an assertibility mapping as a function that maps sentences of different theories that are assertible under all and only the same actual situations, then an assertibility mapping would entail a weak empirical equivalence. That is, it would entail that the theories in question make all of the same observational predictions in the actual world. In fact though, we defined an assertibility mapping as a function that maps sentences of different theories that are asserted under all and only the same possible situations. Thus we have an assertibility mapping only if the theories in question are strongly empirically equivalent, that is, only if they make the same observational predictions in all worlds—there is no actual or possible piece of evidence that could render one theory true and the other false.[19]

So we can use (strong) empirical equivalence as a tool in helping to determine whether we have a practical translation. For if we can show that the theories in question are not empirically equivalent, then we have shown that we do not have a practical translation, and thus we do not a correct translation. Now, in many cases it might seem obvious that certain theories are empirically equivalent. It generally seems to be agreed that nihilism and universalism are empirically equivalent, as are Platonism and fictionalism. So too many would argue, are three and four dimensionalism. But this latter is a source of debate. There are those who hold that the empirical discoveries of special relativity show that three dimensionalism is incoherent,[20] or at least very implausible, while more recently some proponents of gauge theory maintain that the most

19 For a discussion of this distinction see Hoefer and Rosenberg. (1994). Notice also that we might be tempted to say that theories are empirically equivalent just if any nomologically possible prediction of one is a prediction of the other. For we might think that what happens in some radically different logically possible but nomologically impossible world is irrelevant. But notice that there is no nomologically possible piece of evidence that could render two different laws of nature differentially true or false, yet we probably do not want to say that different laws of nature are empirically equivalent, and certainly do not want to say that they are equivalent. So we need to countenance the nomologically impossible worlds. For what is relevant is that there are nomologically impossible states of affairs which render one but not the other theory false.

20 See Smart (1968); Quine (1960); Hales and Johnson (2003).

basic elements of the universe are temporally extended and not wholly present at a moment in time. So empirical equivalence does at least provide one place to begin in attempting to determine whether two theories can be correctly inter-translated. Or, I should say, showing that theories fail to be empirically equivalent entails that they are not practically translatable and hence not correctly inter-translatable and thus not metaphysically equivalent.

2.2 Diagnosing a Practical Translation: the Principle of Charity

As we have seen, if we can show that two theories are not empirically equivalent, then we have shown that there is no assertibility mapping, and thus no practical translation. Even if we have an assertibility mapping, however, this does not entail that we have a practical translation. For the assertibility mapping might fail to be truth preserving. It is here that we discover a problem. After all, one who holds that two theories x and y are not equivalent, will surely maintain that of any pair of co-assertible sentences of those theories, at most only one of those sentences is ever true. Thus since nihilists deny that there exist any dogs, they will maintain that 'there is a dog' is just false, and thus that the assertibility mapping of nihilism and universalism is not truth preserving. And how is the proponent of equivalence is this case to argue that this is not so, beyond appealing to the equivalence of the theories in question?

Suppose we are considering the sentence 'there exists something composed of my dog and your shoe'—call it a Doe. The universalist affirms, while the nihilist and most non-universalists deny the truth of this claim. So, it seems, the two theories are not practically inter-translatable, and thus not correctly inter-translatable. But on what basis does the universalist conclude that the non-universalist's sentence 'there exists no Doe' is in fact false? Clearly 'there exists no Doe' is false when uttered by the universalist, given what she means by that sentence. But the universalist can only conclude that the non-universalist's claim is false if she has some theory about what the nihilist means when *she* utters

that sentence. Now suppose that Hirsch is right at least insofar as there are two *possible* meanings of the existential quantifier and thus two possible languages corresponding to these two different meanings, whether or not these two languages are in fact the languages of the universalist and the non-universalist.

Given the way Hirsch defines the two meanings of the quantifier, it follows that the two languages are indeed practically inter-translatable. For it is the case that we have an assertibility mapping, and that this mapping is truth preserving. For any universalist sentence there is some non-universalist sentence that has the same truth conditions in the sense that relative to any context of utterance, both sentences hold true in all and only the same possible situations. But why think that either of those languages are the languages of the universalist or non-universalist? One good reason is surely that if I am universalist attempting to translate the words of the non-universalist, I should expect most of the sentences that she asserts to be true. Some sort of Davidsonian charity tells me that.[21] Or at least, I should expect most of the sentences she asserts not to be inexplicably false. The principle of humanity tells me that.

But if as a universalist I interpret my non-universalist friend as meaning the same as I do by all of her statements of the form 'there exists an x', then it turns out that I must conclude that a great deal of what she says is just false, as for instance, with her claim that 'there exists no Doe'. Yet that my friend consistently utters falsehoods seems to be quite inexplicable: she is not suffering hallucinations, she has not been tricked by evil demons or anything of that nature. There are no facts that, once made clear to my friend, she would instantly recognise that some of her utterances were indeed false by her own lights. So there seems no good reason to suppose that most of my friend's utterances are false. A more plausible interpretation is that she simply means something different by some of the terms she uses—something alone the lines of those defined by Hirsch. And this can hardly be surprising given

21 See Davidson (1973).

that theoretical terms are in part defined by the roles they play in the theory in which they are embedded.[22]

Similarly, as I argue elsewhere, I think there is good reason to suppose that the parthood relation is construed differently by the three and four dimensionalist. I have argued that the parthood relation is only partially defined by the axioms of mereology, since mereology does not determine whether or not objects are ever composed of non-present parts. But once we see that the concept of parthood is defined by the theories in which it is embedded, and thus that it has a different meaning for the three and four dimensionalist, we can see how it could be that the two theories might be inter-translatable.

None of this is to imply anything about the relation between ordinary English and either of the sub-languages of the theories in question. One way to make sense of what is going on in the case of universalism and non-universalism is to hold that the English sense of 'there exists' is sufficiently semantically vague that both of the meanings of the quantifier as explicated by the universalist and non-universalist, are precisifications of the English sense. Hirsch thinks not, holding that only the non-universalist sense of the quantifier captures anything of the ordinary English meaning of 'there exists'. For Hirsch, the universalist is speaking a language all of her own, and her claim that 'there exists a Doe' is just false in English. But these considerations, while interesting, are not relevant to the issue of whether or not the theories in question are in fact practically or correctly inter-translatable.

So if Hirsch is right and the existential quantifier does have two meanings, then so far I think we have good reason to think that universalism, non-universalism and nihilism are practically inter-

22 Of course, not everyone agrees. Merricks, for instance, holds that the typical non-universalist means by her sentences of the form 'there is an x' just what the folk mean by their same claims. Thus for both the folk and non-univeralist, 'there exists a statue' does not mean just what the nihilst or near-nihilist means by 'there exist simples arranged statue-wise.' Thus he thinks that the sentences of the folk and the non-universalist are straightforwardly false (although they are 'nearly' true). Presumably Merricks would not be moved by such Davidsonian considerations. See Merricks (2001 chapter 1).

translatable. Of course, Hirsch may be wrong. There may be a number of reasons why quantifier variance is false. Perhaps non-universalism, universalism and nihilism are not languages that interpret the quantifier differently, but rather, place restrictions on a quantifier with a univocal meaning, and perhaps various considerations pertaining to the alleged incoherence of the vagueness of existence tell us that.[23] But that is not a matter to be discussed here. What matters is that considerations of charity and humanity mean that if we have an assertibility mapping, we have reason to think that that mapping is truth preserving where failing to do so would result in inexplicably interpreting a large proportion of the sentences of one of the theories as false.

2.3 Diagnosing a Practical Translation: Explanatory Power

We might hope, however that there are some additional tools that will help us to discern whether on an occasion we have a practical translation. For considerations of empirical equivalence (and co-assertibility) clearly will not help us when we are considering theories such as Platonism and fictionalism. So might considerations pertaining to the theoretical virtues of the theories in question aid us in determining whether some assertibility mapping is a practical translation? Let us consider first Alan Sidelle's striking claim that a number of 'ontological packages' are, in some sense, equivalent. It is not clear that Sidelle means to claim that the packages are metaphysically equivalent in the sense I have described, since this would require that each of the 'elements' of each package be inter-translatable, and this is a tall order when packages include components about persistence—three dimensionalism versus four dimensionalism—about time—presentism versus eternalism—and about ontology—nihilism versus universalism. While it is possible that all of these components are inter-translatable, and thus that the packages are metaphysically equivalent, prima facie

23 See for instance Lewis (1986: 212-213; and 1991: 80-81) and Sider (2001: 220-240; and 2003) and Hirsch (2002) for a response.

this seems unlikely. Now, Sidelle seems to acknowledge that these packages preserve different folk intuitions and have different explanatory virtues. So what is interesting here, is whether if the packages were metaphysically equivalent, they would also be explanatorily equivalent. That is, is the fact that they do not appear to be explanatorily equivalent, a reason to conclude that there exists no practical translation? Or is it a reason to conclude that although there exists a practical translation, there exists no correct translation? If either were the case, then consideration of the explanatory virtues of putatively equivalent theories could help us to determine whether those theories are in fact equivalent.

To address this question, let us consider the following case. Consider the theories of classical thermodynamics and statistical mechanics, and the theory of meteorology, and some physical theory that describes weather conditions at the microphysical level. These are not the sorts of theories with respect to which we usually tend to worry about metaphysical equivalence, because they are not, as we might say, theories that occur 'at the same level'. By this I mean, at this stage, only to capture some intuitive sense of theoretical levels, rather than relying on any particular account of levels. So, for instance, I take it that intuitively we all agree that microphysics, chemistry, biology, and psychology are all theories at different levels.

Now, it is debatable whether there is any correct translation that maps sentences of thermodynamics onto sentences of statistical mechanics, and similarly in the case of meterology and physical theory. In the former case it might be objected that in statistical mechanics the second law of thermodynamics is false, and in the latter case that on some occasions the generalisations of high-level meterological theory are literally false. In the case of thermodynamics and statistical mechanics one could argue that properly understood the second law is no law at all, but a mere empirical generalisation to the effect that entropy will increase from a level of low entropy, and this is perfectly consistent with what statistical mechanics tells us (it merely also tells us that in times of

high entropy, entropy may decrease). Regardless, whether these are cases where inter-translatation is possible, if we suppose that there *are* theories such as these that are inter-translatable, it seems plausible that they may not be equally explanatorily powerful.

There are two senses in which theories could fail to be equally explanatory, depending on how one understands the notion of explanatory power. At one end of the continuum we have an understanding of explanatory power that is in large part a psychological one, according to which the explanatory power of a theory is understood in terms of the extent to which it creates understanding in the mind of some relevant group of humans. Thus two theories would be equally explanatory just if they created the same degree of understanding in that same group of persons. This psychological notion of explanation is of course subjective, and entails that any infinite theory such as the physical theory in question, is less explanatorily powerful since no finite human mind could grasp it. I think that if the notion of explanation is to play any role in helping us to decide whether theories are metaphysically equivalent, then the notion of explanatory power at play is clearly not this highly subjective psychological one. Rather, the understanding we want is a more objective one, perhaps some sort of D-N/IS account of explanation, or in some cases a causal account.[24]

In the case of, say meterological and physical theories, if these theories were inter-translatable then it seems that they would still turn out to be explanatorily equivalent. For the infinite physical theory fails to tell us what it is in common between the infinite number of physical

24 These scientific notions of explanation are a little problematic when we are talking of metaphysical theories since there is no nomological or causal component. Of course, if two theories are equivalent, then they are also equally D-N explanatory, though of course the reverse need not hold. It seems that to capture some metaphysical sense of explanation that involves more than merely having two theories entail all of the same sentences, we might need to think of metaphysical explanation as some trade-off between power and simplicity or some such. It is not my task here to provide an account of metaphysical explantion however, whatever the best account might be can be plugged into this account of equivalence.

states mentioned by any particular disjunction, such that they all count as realising the same meteorological state. Arguably, it is an objective feature of the theories that the meteorological theory has greater explanatory power with respect to this particular aspect. So I think it is at least plausible that where we have theories at different levels, we should not expect those theories to be equally explanatory even if they are correctly inter-translatable and thus metaphysically equivalent.

But what of the more usual cases we have been describing, where we are considering theories that are at the same level? If we rule out a psychological notion of explanation, the question is whether two theories at the same level that are metaphysically equivalent could differ in explanatory power. A detailed answer to this question would require a considered examination of an objective account of explanation. As I see it though, metaphysically equivalent theories at the same level ought to be explanatorily equivalent in *some* objective sense. No doubt the truth of this claim also depends in part on which account of levels one accepts. It requires, for instance, that levels be genuinely objective, that is, that relative to different interests, one and the same theory cannot exist at multiple levels. It does not require, however, that there be a single hierarchy, or a non-branching hierarchy of levels. Nor does it require that theories at higher levels be irreducible to theories are lower levels, or be in some way 'emergent'. Nor does it require that in all cases there is some fact of the matter as to whether two theories are at the same level. Perhaps there is no fact of the matter as to whether a theory of flower pollination and a theory of economics are on the same theoretical level. All that is required here is that in some cases we are fairly confident in holding that theories are at the same level. And plausibly, frequently where we are considering putative cases of metaphysical equivalence these are precisely cases where we have such confidence. In part this is presumably because nihilism and universalism, three and four dimensionalism and so forth are, if not equivalent, then competitor theories aiming to fill a single 'theoretical slot': if such theories are not equivalent, then at best one can be true. Where such theories are at the

same level then, they are attempting to fill the same explanatory niche. Thus plausibly if they metaphysically equivalent, then we should expect them to be equally explanatory.

This suggests that where we have theories at the same level, if we can show that those theories are not equally explanatory, then we can conclude that we do not have a correct translation between them. Of course, the reverse is not the case: showing that two same-level theories are equally explanatory would not entail that an assertibility mapping is a correct translation.

So, for instance, Yuri Balashov argues that in the context of truths about special relativity and the structure of Minkowski spacetime, four dimensionalism has explanatory resources that three dimensionalism lacks.[25] Only four dimensionalism, Balashov argues, can explain why it is that the various three dimensional objects that exist, fit together to form nice unified four dimensional volumes. Indeed, for fictionalists and Platonists, the debate about equivalence rests almost exclusively on consideration of explanatory power. Thus fictionalists may complain that Platonism is explanatorily lacking, since given that mathematical objects are causally inert, it fails to explain how one could ever have knowledge of mathematical truths, while Platonists may complain that fictionalism is explanatorily lacking since it fails to explain the nature of our ordinary semantics in mathematical discourse.[26] If such arguments are compelling, then we have reason to suppose that if these theories are same-level theories, then they are not metaphysically equivalent.

All this, however, leaves us with a question. If same-level theories that fail to be equally explanatory thereby fail to correctly inter-translatable, do they also fail to be practically translatable? That is, is discovering that same-level theories for which there is an assertibility mapping are not equally explanatory, just discovering that a practical translation is not a correct translation, or is it discovering that the assertibility mapping is

25 Balashov (2000a); Balashov (2000b); Balashov (2000c).
26 See for instance Benacerraf (1973).

not truth preserving and thus that we have no practical translation and therefore no correct translation? Of course, either way we are able to rule out the existence of a correct translation. But it is an important question, since if we think that the former is the case, then we have a way of ruling out that some practical translation is a correct translation.

In fact I think that considerations of explanatory power are important both in determining whether we have a practical translation, and if so, whether we have a correct translation. In most cases, differences of explanatory power such as those described above, will surely signal that an assertibility mapping is not truth preserving. If four dimensionalism can better explain why there exist nice four dimensional volumes, or if fictionalism can better explain how we have mathematical knowledge, or if nihilism can better explain our intuitions about persistence and convention, then I think these substantive explanatory differences suggest that at least some of the assertibility mapping is not truth preserving. It entails, therefore, that we do not have a practical translation in these cases. In fact it might be possible to use the details of the difference of explanatory power to locate where the assertibility mapping fails to preserve truth, and then to give some explanation of why the proponent of one theory utters frequent falsehoods thereby meeting the principle of humanity.

2.4 Diagnosing a Correct Translation: Explanatory Idle Elements

But what of using explanatory considerations in determining whether a practical translation constitutes a correct translation. This brings us back to the issue we considered earlier, where what looks like a practical translation fails to count as a correct translation in virtue of the addition of necessary truths as conjuncts of the sentences of one theory. Now of course, in such cases we only have a practical translation if all parties agree that the 'additional' bits of the theory in question—the 'God exists' or the 'p or not p'—are in fact necessary truths: otherwise any assertibility mapping will fail to be truth preserving.

One way we might be tempted to try and deal with this problem is via

some theoretical constraint of simplicity. We might be tempted to argue that simplicity is a guide to whether a translation is correct or not. That is, we might argue that a practical translation is only a correct translation if the theories are equally simple. What tells us that the sorts of cases we have been discussing are not cases where we have a correct translation, is that one theory is obviously less simple than the other.

The problem with this move is twofold. First, it is not clear why we should think that only equally simple theories are ever metaphysically equivalent. This sounds a little more plausible if we hold that theories at the same level will, if equivalent, be equally simple. Even this claim though, would need to be convincingly argued for before we could use simplicity as a guide to whether a practical translation is correct or not. Second, even if such an argument were forthcoming, there would be momentous practical difficulties in using simplicity in this manner, given that it is notoriously difficult, if not impossible, to provide any formal account of simplicity.[27] Indeed, it may be that there is no objective marker of simplicity, but rather, that claims about the relative degrees of simplicity of theories are all culturally and inter-personally subjective. Those who think this latter will of course think that there is no reason to suppose that any of our subjective judgements about the relative degree of simplicity of theories is any guide to their equivalence or lack thereof. Even those who think that there is some objective notion of simplicity, concede that formulating judgements about simplicity is difficult because it is always possible to make a theory *appear* simpler by burying the complexity in the atomic predicates of the language.[28]

I suspect that if there is some objective measure of simplicity, then we should expect that the simplest version of theories at the same theoretical level will be equally simple if those theories are correctly inter translatable. Even if that is true though, it is arguably the case that determining whether the simplest versions of any two theories are

27 Cf. Sober (1979)
28 DeVito (1997).

equally simple or not is more difficult that determining whether or not they are correctly translatable. So whatever one's take on the simplicity issue, it is difficult to see how considerations of simplicity will be of much help in determining whether or not theories are correctly translatable.

There is, however, a way in which we can rule out some practical translations as being correct. In the previous section I spoke of the 'substantive' explanatory differences between theories, that one might argue exist in arguing that an assertibility mapping is not truth preserving. I use this term to distinguish those sorts of cases of differential explanatory power, from cases where we conclude that two theories differ insofar as one theory has additional explanatorily redundant elements. This latter kind of case covers the sorts of cases we have been considering where there are additional necessary truths. Since the addition of 'p or not p' to every sentence you assert makes no difference to the truth value of that sentence, and adds nothing explanatorily to the theory in which it is embedded, we can conclude that the theory with these additions is less non-substantively explanatory[29] than its rival. Of course, this is not the sort of substantive explanatory difference that we discussed earlier. It is sufficient, however, for us to conclude that where we have a practical translation between theories, if one of those theories is less explanatory in this non-substantive sense, then we fail to have a correct translation. Thus the use of such explanatory considerations rules out as equivalent theories of this sort which are practically translatable but which are not translatable in virtue of all of the same truth makers. Moreover, where we have theories that are practically translatable but fail to be correctly translatable in virtue of failing to be non-substantively explanatorily equivalent, there is

29 Of course, we might think that the theories are just equally explanatorily powerful (that there is no notion of non-substantive explanatory power) but that there is some additional constraint at play—a redundancy or idleness constraint. The difference here is merely terminological, we can talk of theories being equally explanatorily powerful but differing with respect to some idleness constraint, or we can talk of them differing in a non-substantive explanatory manner. The point is just that if they differ in that manner—whatever you call it—then they are not equivalent.

good reason to suppose that there is some 'pared down' version of one of the theories such that that theory and its putative rival are explanatorily equivalent and therefore possibly correctly inter-translatable.

3 Why is this Metaphysical Equivalence?

So far I have argued that theories are metaphysically equivalent just if they are correctly inter-translatable. Moreover, I have provided a number of diagnostic criteria to aid us in determining whether or not such a correct translation exists. But in effect, these diagnostic criteria provide necessary but insufficient conditions for the obtaining of the relation of metaphysical equivalence. Thus they provide a way of falsifying some claim about the equivalence of two theories: if it can be shown that some diagnostic criterion is not met, then we can know that the assertibility mapping does not amount to a practical translation, or that the practical translation does not amount to a correct translation. Suppose though, we find that in some cases all of the diagnostic criteria are met by two theories. Are we then in a position to claim that we have a correct translation, and thus a case of metaphysical equivalence? Well certainly such a conclusion is not entailed. For we have a correct translation only if both theories have the same truth makers, and meeting the diagnostic criteria does not entail that this is so. So what would license such a move?

One reason we might have to conclude that if the diagnostic criteria are fulfilled then we have a genuine case of metaphysical equivalence, is if we are committed to what Hirsch calls a shallow approach to ontology, and what Sidelle refers to as the semantic approach.[30] The idea here is that in the sorts of cases we have been considering, once we know all of the facts we see that deciding which of two theories is true is not a deep metaphysical matter, but rather is just a matter of deciding the best way to describe those facts in our language. Or, to put it another way, whether a Doe exists or not is not something that could be made true by

30 Hirsch (2002) pg 67 and Sidelle (2002) pg 137.

some mysterious metaphysical fact. Once we know how the various simples are arranged, we know all of the facts, and it is merely a semantic question whether or not, given what we mean in English, a certain description of those facts is a proper one. So in deciding which theory is true, we are merely deciding which fits best with our ordinary talk.

Given this semantic approach, it is easy to see why we would conclude that theories that meet our diagnostic criteria are in fact equivalent: for if theories meet the diagnostic criteria, then they could only fail to be metaphysically equivalent if there were some extra metaphysical facts in virtue of which one theory were true and the other false. That is, they could only fail to be equivalent if there were some unobservable truth makers that are explanatorily redundant. Since the semantic approach rejects the existence of such facts, it follows that proponents of this approach will have every reason to think that theories which meet these diagnostic criteria are in fact metaphysically equivalent.

Reaching this conclusion does not require, however, that one adopt this semantic approach to ontology. That is, it is not necessary that in general one deny the existence of such extra metaphysical facts. Rather, the question can become whether on any particular occasion, there is good reason to suppose that there are such extra facts. For consider, there are many unobservable facts. There are unobservable facts about the big bang. There are unobservable facts about the existence or not of other worlds that are causally inaccessible to this world. If modal realism is true, then it is so in virtue of facts that are unobservable to us. Although these facts are unobservable, however, we understand what sort of facts they are. Scientific theories about the big bang, or the inside of a black hole, tell us not only that there are certain facts and that some of these facts are unobservable, but they explain why it is that the facts that are unobservable are so. So too although it is impossible for us to observe worlds that are causally isolated from us, we at least have some understanding of what such worlds would be like, and why it is that they are unobservable. In each case proponents of these theories can point

out what facts would need to obtain for the theory to be true. Further, broad features of the theory itself explain why it is that some of those facts are unobservable. These sorts of unobservable facts are posited by a theory in an integrated, systematic way that provides additional explanatory power to the theory, and where this is an explanation for the unobservable status of the facts.

Suppose though, that we are considering theories that meet the various diagnostic criteria we have outlined. And suppose that someone insists that despite meeting those criteria, they are not metaphysically equivalent: there is some further fact that entails that at most only one theory is true. Is this a fact that is posited by either of the theories in question in an integrated manner? Well suppose the three dimensionalist insists that three dimensionalism is true because there is some further metaphysical fact such that objects are strictly identical across time. The existence of such a fact is no part of the theory of three dimensionalism. Three dimensionalists do not posits such facts as part of their general strategy of explaining change and persistence. Nor does the three dimensionalist theory explain in an integrated way, why we should expect there to exist such extra facts. Rather, here we seem to have an *ad hoc* measure of introducing some unobservable fact that provides no additional explanatory power to the theory, purely in order to maintain that the relation of metaphysical equivalence fails to hold. This seems no more than a desperate attempt to hold that there *must be* some fact that determines that one theory is true and the other false.

If we find, on some occasion, that there is a plausible non *ad hoc* explanation for why some relevant truth maker is unobservable and explanatory redundant, then we might feel justified in holding that there is such an extra fact on that occasion. So under certain circumstances we might feel justified in concluding that two theories that meet all of the diagnostic criteria are not metaphysically equivalent in virtue of the existence of some unobservable fact. In general though, the sorts of cases we have been considering do not appear to be cases where such additional metaphysical facts are posited by the competing theories

themselves. Rather, the positing of such facts appears to fall into the category of an *ad hoc* measure: these facts add no explanatory power, and are not part of the theory itself. Given this, there seems little reason to suppose that such facts exist. Once we see this though, something like inference to the best explanation should tell us that theories that meet the diagnostic criteria are metaphysically equivalent. What explains the fact that the theories meet all of these criteria? Well, that they are actually correctly inter-translatable—they have the same truth makers: they are metaphysically equivalent.

4 How does this account help?

It might seem though, that some of these criteria are unnecessary. We could, it might be argued, first determine whether there are any observable facts in virtue of which one theory is true and the other false. If we determine that the theories are empirically equivalent, then there is a further question about unobservable facts: are there any unobservable facts in virtue of which one theory is true and the other false? To determine whether there are such unobservable facts we turn to the idea of explanatory *ad hocness*, and ask whether or not the positing of such facts is part of the apparatus of the theory, or merely an *ad hoc* measure. If it is purely *ad hoc*, then we can conclude that the two theories are metaphysically equivalent, if it is not *ad hoc*, then we can conclude that they are not metaphysically equivalent. We then need no recourse to the criteria of translatability or relative degree of theoretical virtues.

We might think that such a proposal is a good one. After all, there is considerable controversy about inter-theoretic translation, and more still about the notion of explanatory power. But there are two reasons why rejecting these additional criteria is not a good idea. First, consider how we are to ascertain whether the positing of the relevant unobservable facts is principled or *ad hoc*. Suppose I am considering three and four dimensionalism as theories. Suppose everyone agrees that the theories are empirically equivalent. Is there some non observable fact in virtue of which one is true and the other false? Well if four dimensionalism is true,

then it is because objects perdure: they are composed of temporal parts. That fact is unobservable, but there is clearly a sense in which it is not *ad hoc*: after all, it is precisely part of the apparatus of the theory. Does this resolve whether or not the theories are metaphysically equivalent or not? No. For we have assumed that the theoretical terms as they are used by the four dimensionalist, are the same as the terms as they are used by the three dimensionalist. But if these terms are in part defined by their role in the theoretical apparatus in which they are embedded,[31] then there is no reason to suppose that when the three dimensionalist denies that there exist temporal parts and thus that objects perdure, that she is denying what the four dimensionalist is affirming. The claim that three and four dimensionalism are metaphysically equivalent and thus inter-translatable, is the claim that they mean something different by terms such as 'part' such that although they appear to be making contradictory claims, in fact they are not. Whether or not there is some principled unobservable fact in virtue of which four dimensionalism is true and three dimensionalism is false, is impossible to determine independent of knowing whether the two theories are inter-translatable.

More important that this though, the various diagnostic criteria are invaluable in determining precisely where parties disagree about whether theories are equivalent or not. At present it is often unclear why it is that proponents of the view that certain theories are equivalent think that they are so, and equally unclear why those who disagree do so. Just where does the disagreement lie? We can now see that there are a number of junctures at which parties might disagree. Parties might disagree from the outset about whether or not there is an assertibility mapping between two theories: they might disagree about whether the theories are not empirically equivalent or not. Or they might agree that such a mapping exists, but disagree about whether it is truth preserving: one party might contend that we fail to have a practical translation, and

31 Where their role in the theory can include causal facts about various theorists and so forth.

this is so because there are substantive differences of explanatory power. This disagreement might rest on differences the parties have in how they understand explanation. If so, then it is important to be aware of this. On the other hand, parties might agree that two theories are not equally explanatory, but disagree about whether they are at the same level. Thus one party may conclude that the lack of explanatory equivalence does not entail that there is no practical translation since the theories are not at the same level. Thus the disagreement is a product of a different account of theoretical levels. Or parties might agree that there exists a practical translation, but disagree about whether there is a non-substantive difference in explanatory power, thus disagreeing about whether the translation could be correct. Finally, parties might agree that all of the diagnostic criteria are met, but one might thump the table and maintain that there is some unobservable, explanatorily redundant truth maker in virtue of which one theory is true and the other false.

Seeing precisely where the nature of the dispute lies, and upon what it is based is all important in resolving such disputes, particularly given that they may sometimes rest on additional, sometimes controversial commitments such as to different theories of explanation or to different accounts of theoretical levels. So although this account is by no means a panacea, it does the apparatus with which to clarify the nature of disputes, and to provide a framework within which to argue that theories that we might have thought were radically different, are in fact equivalent.[32]

University of Queensland

References

Balaguer, M. (1998). *Platonism and Anti-Platonism in Mathematics*. New York: Oxford University Press.

32 With thanks to David Braddon-Mitchell, Mark Colyvan and Dominic Hyde for helpful discussion of these issues.

Balashov, Y. (2000a). 'Relativity and Persistence.' *Philosophy of Science*. 67(3): 549-562.

Balashov, Y. (2000b). 'Enduring and Perduring Objects in Minkowski Space-Time.' *Philosophical Studies* 99: 129-166.

Balashov, Y. (2000c). 'Persistence and Space-time: Philosophical Lessons of the Pole and Barn.' *The Monist*. 83(3): 321-240.

Benacerraf, P. (1973). 'Mathematical Truth' *Journal of Philosophy* 70: 661-80.

Carnap, R. (1936) 'The elimination of metaphysics through logical analysis of language.' in *A.J. Ayer Language, truth and logi*c 1946 Harmondsworth: Penguin.

Davidson, D. (1973). 'Radical Interpretation' in *Inquiries into Truth and Interpretation* Oxford: OUP.

DeVito, S. (1997). 'A gruesome problem for the curve-fitting solution.' *British Journal for the Philosophy of Science* 48(3): 391-396.

Hales, S.D and Johnson, T.A. (2003). 'Endurantism, Perdurantism and Special Relativity' *The Philosophical Quarterly* 53: 524-539.

Hirsch, E. (2002). 'Quantifier Variance and Realism' *Philosophical Isssues, 12 Realism and Relativism* 51-73.

Hoefer, C. and A. Rosenberg. (1994). 'Empirical Equivalence, Underdetermination and Systems of the World.' *Philosophy of Science*. 61: 592-607.

Lewis. D. (1986). *On the Plurality of Worlds.* New York Blackwell Press. pp 212-213.

Lewis, D. (1991). *Parts of Classes.* Oxford: Blackwell p 80-81.

McCall, S. (1994). *A Model of the Universe: Space-Time, Probability, and Decision*. Oxford: Clarendon Press.

McCall, S. and E.J. Lowe (2003). '3D/4D equivalence, the twins paradox and absolute time.' *Analysis* 63(2) 114-123.

Merricks, T. (2000). 'No Statues.' *Australasian Journal of Philosophy,* 78(8): 47-52.

Merricks, T. (2001). *Objects and Persons.* Oxford University Press.

Putnam, H. (1987). 'Truth and Convention: On Davidson's Refutation of

Conceptual Relativism' *Dialectica* 41: 69-77.

Putnam, H. (1988). 'After Metaphysics, What?' in *Metaphysics: The Big Questions*. eds D. Zimmerman and P. van Inwagen. Oxford: Basil Blackwell Press. pp 388-392.

Quine, W.V.O (1960). *Word and Object*. Cambridge: MIT Press.

Quine, W.V.O. (1975). 'On Empirically Equivalent Systems in the World.' *Erkenntnis* 9: 313-328.

Sidelle, A. (2002). 'Is there a True Metaphysics of Material Objects?' *Philosophical Issues 12 Realism and Relativism* 118-145.

Sider, T (1999). 'Presentism and Ontological Commitment' *Journal of Philosophy* 94(7) 325-347.

Sider, T. (2001). *Four-dimensionalism: an ontology of persistence and time*. Oxford University Press.

Sider, T (2003). 'Against Vague Existence.' *Philosophical Studies*, 114: 135-146.

Smart, J.J.C (1968). *Between Science and philosophy: an introduction to philosophy of science*. New York: Random House;

Sober, E. (1979). *Simplicity*. Oxford: Clarendon Press.

Unger, P. (1979). 'I Do Not Exist'. in *Perception and Identity* ed G. F. Macdonald. London Macmillan Press.

van Inwagen, P. (1990). *Material Beings*. Cornell University Press.

van Inwagen, P. (2002). 'The Number of Things' *Philosophical Issues 12 Realism and Relativism* 176-196.

Zalta, E.and Colyvan, M. (1999). 'Mathematics: Truth and Fiction?' *Philosophia Mathematica* 7(3): 336-349.