Hume's Law and other Barriers to Entailment

Gillian Russell

October 7, 2013

This paper consists of five sections from a book that I'm working on, called *Barriers to Entailment.* The sections are not contiguous in the book, but I hope they can be read together as a coherent whole anyway. The first section explains what a barrier thesis is, and offers examples and putative counterexamples. The second outlines a strategy for proving such theses in a simple case—the particular/universal barrier. The third shows how the strategy can be applied in other cases and proves a more general theorem. Section four focuses on some lingering (but difficult) problems. (This section could be skipped by a reader who just wanted the gist of the paper.) The final section introduces the possibility of similar techniques being effective in contexts where we have no agreed upon formal logic. You'll note that I don't end up getting as far as a section on Hume's Law (and the paper is already a bit long), but I do want to apply this work in that case and I would welcome discussion, questions, and objections on that topic.

1 Barriers to Entailment

This book is about barriers to entailment. A barrier to entailment is a thesis which says that no set of premises containing only sentences *of a certain kind X* logically entails a conclusion *of some other kind Y*, or, more colloquially, that you can't 'get' an X from a Y. There are lots of these in philosophy but the most famous is Hume's Law, which says, roughly, that no set containing only descriptive sentences entails a normative sentence, or in slogan form, that *you can't get an 'ought' from an 'is'*.[1] This barrier thesis is famous in part because it is very controversial.[2] And one reason it is controversial is that many have regarded it as important and so *worth* arguing about.[3] But there there are other, more platitudinous, barriers to implication which have received less attention, such as:

[1] There are other interpretations of Hume's Law and I don't mean to suggest that it can only be interpreted as a claim about entailment. The project here is just to explore one interesting interpretation of it.

[2] Versions of it have been endorsed by R.M. Hare, Karl Popper and Frank Jackson and attempted refutations have been published by, for example, Max Black, A.N. Prior, and John Searle. (Popper, 1948; Hare, 1952; Jackson, 1971; Prior, 1960; Searle, 1964; Black, 1964)

[3] Popper, for example, said that it was "perhaps the simplest and the most important point about ethics."

- You can't get a universal claim from a particular one
- You can't get claims about the future from claims about the past
- You can't get claims about how the world *must* be from claims that merely state how it *is*
- You can't get indexical claims from non-indexical ones.

These claims can seem so uncontroversial as to be part of the background philosophical atmosphere within which many of us work—certainly they have rarely been formulated precisely enough to earn attempted refutations in philosophical papers, as Hume's law has.[4]

One aim of this book is to show that versions of these barrier theses can be proved—in a straightforward, metatheoretical and mathematical sense of 'proof'—for standard logics for dealing with the various domains. For example, in *classical first-order logic*, we can prove a theorem which says, more or less, that you can't get a universal claim from a particular one. In *LD* we can prove that—speaking roughly for now—no non-indexical sentence entails an indexical one. The method of proof to be used will be explained and illustrated in the first chapter and subsequent chapters will tackle different barriers to implication: those concerning time, modality, context-sensitivity and normativity respectively. These barriers each receive chapters of their own because in many ways, with the proofs, the philosophical work is just beginning. There are questions about the robustness and significance of the proofs, the worth of the logic for which they have been proved, the ability of the theorems to endorse and explain philosophical theses that until now have rested on intuition, what the significance of the proof is for historically controversial positions in the area, and, of course, what *doesn't* follow from the proofs.

Much has been written that is relevant to the *individual* barriers to implication—the literatures on the problem of induction, the essentiality of the indexical, modal epistemology, natural laws and of course, Hume's Law itself are all relevant—but these barriers have not normally been studied together. It is worth studying them together because once we recognise their similarities there is much more pressure for a uniform approach, and conclusions that are sometimes drawn in one area look a lot more specious when the same move is attempted elsewhere. Sometimes philosophers who have endorsed Hume's law, for example, have suggested that it presents such a huge obstacle in moral epistemology that we should be driven either to Moorean intuitionism or to moral skepticism. But this approach would have much less plausibility if it were taken in discussions of the epistemology of *future matters of fact.* Suppose that claims about the past and the present never entail claims about the future (what we might call "Hume's 2nd Law.") Is it reasonable to conclude that we are forced to accept that we have intuitive access to future truths, or instead that we can

[4]But that is perhaps itself telling, because such papers could certainly have been written. Since (Kripke, 1980) many philosophers would have felt the pull of $a = b \vDash \Box(a = b)$ as a counterexample to the thesis that one can't get claims about how things are necessarily from claims about how they actually are, and those familiar with the logic *LD* (Kaplan, 1989) would have suspected that $\phi \vDash \textit{Actually}\ \phi$ was a counterexample to the claim that no indexical sentence follows from a set of purely non-indexical premises.

never have justified beliefs with respect to the future? No, in this case it seems much more likely that we have some non-deductive justification.

1.1 Putative Counterexamples

[The book contains a much broader survey of counterexamples. I focus on only a few here, for reasons of length.]

An initial attempt at clarifying the barrier theses might analyse them as taking the following form:

(1) No set containing only X sentences entails a Y sentence.

For example,

(2) No set containing only particular sentences entails a universal sentence.

(3) No set containing only sentences about the past entails a sentence about the future.

(4) No set containing only sentences about the actual world entails a sentence about all possible worlds.

(5) No set containing only non-indexical sentences entails an indexical sentence.

(6) No set containing only sentences about how things are entails a sentence about how they ought to be.

If this is the form that the barrier theses take, then a counterexample to a barrier is a valid argument from a set containing only X sentences to a Y sentence. Here are some counterexamples from (or adapted from) the literature.

Prior's Dilemma Schurz describes the counterexample proposed by A.N. Prior in his 1960 paper "The Automony of Ethics" as "initially shocking"[5] and Prior himself cites it as a reason for changing his own mind about Hume's Law:

> It has often been said—in fact, I have said it quite emphatically myself—that it is impossible to deduce ethical conclusions from non-ethical premises. This now seems to me to be a mistake. (Prior, 1960:199)

The original dilemma goes like this. Here is a counterexample to Hume' Law:

Tea-drinking is common in England.

Tea-drinking is common in England or all New Zealanders ought to be shot.

[5] (Schurz, 1997:11)

This can be formalised as a truth-functionally valid argument, and Prior proposes that it has descriptive premises, but a normative conclusion. He wonders, however, whether some might be tempted to defend the law by denying that the conclusion is normative. For such an objector he suggests the following counter-example instead:

Tea-drinking is common in England or all New Zealanders ought to be shot.
Tea-drinking is *not* common in England.

All New Zealanders ought to be shot.

The real force of Prior's argument against Hume's Law comes from the pressure exerted by the above arguments as a pair. If we say that disjunctions of descriptive and normative claims are normative, then the first argument is a counterexample. But if we say that they are not, then the second argument is.

The structure of the dilemma is easily adapted to provide an analogous argument for each of the barriers mentioned above.[6] For example, is $Fa \vee \forall x Gx$ a universal sentence? If so, then the first argument below is a counterexample. If not, then the second is.

$$\frac{Fa}{Fa \vee \forall x Gx} \qquad \frac{\begin{array}{l} Fa \vee \forall x Gx \\ \neg Fa \end{array}}{\forall x Gx}$$

Counterexamples via Contrapostion Many barrier theses are supposed to be uni-directional. The particular-universal thesis denies that particulars entail universals, but we don't generally think that the reverse is true; $\forall x Fx \vDash Fa$ is classically valid. Similarly, though the actual/necessity barrier thesis denies that necessity-style claims follow from actuality-style ones, claims about what is actually the case can follow from claims about what has to be the case.[7] But now we seem to be able to generate counterexamples to the barrier theses via the following principle of contraposition (CP):

(CP) If $\phi \vDash \psi$ then $\neg\psi \vDash \neg\phi$

Applying this to the entailments in the paragraph above gives us the following:

$$\frac{\neg Fa}{\neg \forall x Fx} \qquad \frac{\neg p}{\neg \Box p}$$

These look like potential counterexamples to the relevant barrier theses; or at least, the premises don't contain any of the relevant vocabulary ($\forall$, $\Box$ respectively) while the conclusions do. An important argument of this type is the 'ought implies can' problem for Hume's Law. The basic idea is that—at least

[6] (Russell, 2010)
[7] Some barriers do appear to be symmetric, however, for example, the past/future barrier.

on some understanding of 'ought'—to say that someone ought to do an action ϕ entails that they can do ϕ, so that the following argument would be valid:

$$\frac{\text{Alice ought to donate US\$10000 to Oxfam.}}{\text{Alice can donate US\$10000 to Oxfam.}}$$

But if this is valid then by contraposition so is this one:

$$\frac{\text{It is not the case that Alice can donate US\$10000 to Oxfam.}}{\text{It is not the case that Alice ought to donate US\$10000 to Oxfam.}}$$

This looks like an argument with non-normative premises and a normative conclusion.

Non-formal Counterexamples I gave the above argument in natural language in part because there is some controversy about the best way (or even if there is any way) to translate normative claims into formulas. There are other informal counterexamples, some of which are not easily formalised. In his own defence of Hume's Law, Frank Jackson discusses the following argument, which appears to be a non-formal example of the contraposition strategy:

$$\frac{\text{Joe did not kill Peter.}}{\text{Joe did not murder Peter.}^{8}}$$

As putative counterexamples to the particular-universal barrier we might consider things like:

$$\frac{\text{Alice is the only winner.}}{\text{Every winner is identical to Alice.}}$$

For the past/future barrier we could use:

$$\frac{\text{Alice is mortal.}}{\text{In the future Alice dies.}}$$

Another class of hard to formalise counterexamples uses metalinguistic constructions and proposition attitudes, as in:

$$\frac{\begin{array}{l}\text{Bob uttered the English sentence “everyone is happy.”}\\ \text{The sentence Barry uttered is true.}\end{array}}{\text{Everyone is happy.}}$$

Here the metalinguistic constructions are being used to "smuggle in" commitment to a universal claim. We might say that the first premise identifies a universal claim without committing to the truth of it, and the second commits to the truth of the claim without without identifying it as a universal one. That

way neither is transparently universal, though the conclusion is. A similar case for the past-future barrier is:

The utterance 'there will be a sea-battle tomorrow' is true now.

There will be a sea-battle tomorrow.

The locution involved doesn't, strictly speaking, need to be metalinguistic, as propositional attitude constructions can be used as well:

Everything that Callum believes is true.
Callum believes that he will grow up to be a pirate.

Callum will grow up to be a pirate.

2 The Proof Strategy

With these counterexamples in mind, we now return to the simplest barrier thesis, which says that you can't get a universal sentence from particular ones. There are at least three ways one could refine such a thesis. We could add a lot of extra "exception" clauses to it so that it no longer makes claims about the problematic cases. This brute force method might work—at least until we stumbled across a new exception—but the result would be ugly and ad hoc and that would deplete its explanatory power. We could instead weaken what we mean by "entails". Perhaps we should mean "relevantly entails" or entailment relative to some special weaker logic that we could invent. But the methodological principle from chapter **??** advises against this. Finally we could define the notions of "universality" and "particularity" more carefully, with the aim of making it the clear none of the potential counterexamples have premises which are *genuinely* particular and a conclusion that is *genuinely* universal. This last is the strategy that I will in fact pursue.

Here is an intuitive thought about the difference between particular and universal sentences. Particular sentences are in some sense *local*. They refer to an object and say something about it, and so, in order for the sentence to be true, the object has to be a certain way. But they don't generally tell you how the whole world has to look—just one particular part of it. Because of this, changing other parts of the world, and especially adding objects with various properties won't change the truth-value of a particular sentence. Universal sentences, by contrast, impose restrictions on *everything*. Even restricted universal claims are committed to the claim that everything is either outside the restriction, or a certain way. As a result, adding new objects, with various properties, can make them false.

One way to make these ideas more precise is using the usual model-theory for first-order logic.

[The book introduces a specific first-order-language and defines truth in a model, first-order validity and first-order logical consequence at this point.

Omitted here for brevity.]

Fa seems like a paradigm particular sentence. It has the feature that if it is true in a model M, then it is true in all extensions of M—models that we get from M by adding new objects to the domain, and extending, if we like, the extensions of our non-logical predicates to include them. (See figure 1)

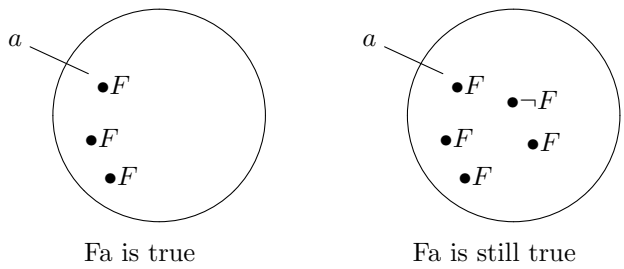


Figure 1: True particular claims stay true when the model is extended

$\forall x Fx$, by contrast, seems like a paradigm universal claim. It is such that if its true in a model—like the one on the left below—it can be made false by extending the model, in this case by adding an element which is not 'F' (see figure 2.)

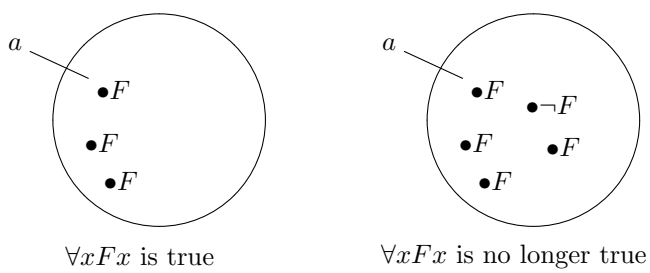


Figure 2: True universal claims may become false when the model is extended

We can use idea to define particular and universal sentences. First we will need to say what we mean by the *extension* of a model:

Definition 1 (Extension ($\supseteq$))**.** *A model M' is an extension of a model M ($M' \supseteq M$) just in case M' can be obtained from M by adding more objects to the domain and extending the interpretation of the predicates to cover the cases of the new objects. (If F is an n-place predicate and g an assignment of variables to values in the domain of M (avoiding the extra objects in M') then if $\Pi x_1, \ldots x_n$ is true in M, g, then it is true in M', g as well.)*

Then we use the relation to define two classes of sentence:

Definition 2 (Particular Sentence)**.** *A sentence S is particular if and only if, whenever it is true in a First-Order model M, it is also true relative to all extensions of M.*

Sentences that count as particular under this definition include Fa, $Fa \wedge Ga$, $\neg Fa$ and other satisfiable truth-functional compounds of particular sentences. Also classified as particular is $\neg \forall x Fx$, telling us that universality is not preserved by negation.

Definition 3 (Universal Sentence)**.** *A sentence S is universal if and only if, if it is true in a model M, there is at least one extension of M in which it is false.*

Sentences which count as universal include $\forall x Fx$, $\forall x(Fx \rightarrow Gx)$, and $\forall x Fx \vee \forall x Gx)$,.
Using the definitions we can the formulate a barrier thesis, and this, as it turns out, is provable:

Theorem 4 (Particular-Universal Barrier Theorem (version 1))**.** *If Γ is a satisfiable set of particular sentences and δ is a universal sentence, then $\Gamma \nvDash \delta$.*

Proof. Suppose Γ is a satisfiable set of particular sentences and δ is universal. Let M be a model which satisfies Γ. Either δ is true in M or it isn't. If it isn't, then M is a counterexample showing that $\Gamma \nvDash \delta$. But if δ is true in M, then since δ is universal there is some M such that $M \subset M'$ and δ is not true in M'. Since each member of Γ was particular, each member of Γ is also true in M'. Hence M' is our counterexample, and $\Gamma \nvDash \delta$. □

Some features of these definitions are worth remarking on. First, if two sentences are logically equivalent, then they are true in all the same models and so the definitions will classify them the same way. On a simple *syntactic* approach, Fa would be classified as particular (it doesn't contain a quantifier) and $\forall x(x \neq a \vee Fx)$ as universal (it does contain a universal quantifier.) But our new model-theoretic definitions reveal the second sentence to have been a particular sentence "in disguise"; it will be true if and only if Fa is and if it is true in a model, it is true in all extensions of that model.

Second, by the Łoś-Tarski theorem (Hodges, 1997:143–146) the set of particular sentences will be identical with the set of $\exists_1$ sentences, which are sentences which in prenex normal form consist of a string of existential quantifiers followed by a quantifier-free formula. Hence the fact that our sentences have been characterised semantically doesn't actually exclude a syntactic characterisation.

Third, all logical truths are particular. This is surprising, and perhaps unwelcome, given that some of them look disturbingly universal, for example $\forall x(Fx \rightarrow Fx)$ and $\forall x \exists y(x = y)$. It is perhaps obvious that any sensible barrier theorem will have to make a special case of the logical truths. They follow from anything and so any universal ones will follow from any particular sentence you like. But it would have been one thing to classify them as "neither particular

nor universal"—do we really want to say that they are *particular*? We *could* solve this problem by complicating the definitions now, but in fact there will be a more motivated and uniform solution later on, and so I will leave it as it is for now, with a promise to return to it in section 4.

Fourth, the two sets of sentences are not exhaustive of the sentences in the language. Some sentences are particular, some are universal, but some are *neither*. For example, $Fa \vee \forall xGx$ is neither particular nor universal because there are there are two different ways in which this sentence can be true—Fa could be true, or $\forall xGx$ could be true without Fa being true. If the former, then there will be no extensions of the model that make it false and so the sentence is not universal. If the latter, then there will be extensions of the model that make it false and so the sentence is not particular. Hence it is neither.

Fifth and finally, I note that the two sets of sentences are not exclusive either. There will be some sentences—i.e. the unsatisfiable ones—which count trivially as both particular and universal.

2.1 Some initial Treatment of Counterexamples

In section 1.1 I laid out some putative counterexamples to barrier theses. What should we say about these given our present approach? Here I'll say something about contraposition and Prior's dilemma, and one way of dealing with some informal examples.

Contraposition One thing we should note is that universality is not preserved under negation; though $\forall xFa$ is universal, $\neg\forall xFx$ is not. Hence the contraposition strategy is not guaranteed to be successful. In the particular case that we looked at, the premise is particular but the conclusion is not universal, and hence we do not have an argument from particular premises to a universal conclusion.

$$\neg Fa \vDash \neg\forall xFx$$

Prior's Dilemma A second kind of counterexample was modelled on Prior's dilemma. The key to dissolving this is the fact that we classify the disjunction $Ga \vee \forall xFx$ as *neither* universal nor particular. This means that the argument on the left is not an argument with a universal conclusion, and the argument on the right does not have only particular premises:

$$\frac{Fa}{Fa \vee \forall xGx} \qquad\qquad \frac{\begin{array}{l} Fa \vee \forall xGx \\ \neg Fa \end{array}}{\forall xGx}$$

Non-formal arguments (1) One strategy for dealing with arguments in natural language (admittedly not always available) is to translate the argument into an appropriate formal language. This is a natural strategy to take with this proposed counterexample:

$$\frac{\text{Alice is the only winner.}}{\text{Every winner is identical to Alice.}} \qquad \frac{\forall x(W(x) \to x = a.)}{\forall x(W(x) \to x = a.)}$$

The translation on the right is valid, but the premise is not really particular, and hence it is not counterexample to the particular/universal barrier thesis. In discussions of barrier theses there is often a temptation for one side to maintain that a particular sentence contains some disguised universality, or normativity etc.—something that is not visable in the surface structure of the sentence. Such claims can be hard to evaluate from a neutral standpoint. The model theoretic strategy for characterising these properties can make the stategy seem less ad hoc.

3 A second case, and the general case

> "All inferences from experience suppose, as their foundation, that the future will resemble the past...if there be any suspicion that the course of nature may change, and that the past may be no rule for the future, all experience becomes useless, and can give rise to no inference or conclusion. It is impossible, therefore, that any argument from experience can prove this resemblance of the past to the future; since all these arguments are founded on the supposition of that resemblance." (Hume, 1748:4.21/37)

Hume's work also suggests a temporal barrier thesis, which we might naively attempt to formulate as:

(7) No set of premises about the past entails a sentence about the future.

Or perhaps even the slightly stronger:

(8) No set of premises about the past and the present entails a sentence about the future.

But once we move to thinking about time, some new putative counterexamples present themselves, such as:

$$Pp \vDash FPp$$

And if we are thinking about

$$a = b \vDash G(a = b)$$

What we need to apply the strategy of the last section to a new case is:

i) a logic suitable for formalising arguments in the relevant area (in this case, time)—that is, we'll need a formal language and some model theory sufficient for defining logical consequence on that language

ii) a relation that plays the role that model-extension played in the particular/universal case

iii) definitions of two classes of sentence (here, *past* and *future*). We'll then use those to formulate the barrier theorem, and prove it.

3.1 Tense Logic

[The section introducing tense logic is omitted here for brevity. Our tense-logic specific expressions are F, P, G and H and models are quadrupels $\langle T, <, n, I\rangle$ in which T is a set of times, $<$ the "earlier than" relation on T, n is a member of T (thought of as the present moment and represented by a square box in the pictures below) and I is the interpretation function which assigns 0 or 1 to atomic sentences relative to a time in T. Truth in the model is truth at n. Later on we'll also need to add a domain of quantification to models so that we can think about first-order tense logic. For much of the discussion it is safe to assume that the $<$ relation is transitive, dense, both left- and right-extendible and total. This generates a fairly strong tense logic.]

3.2 Future-Switching

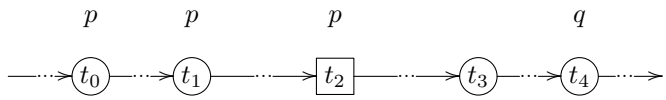
With our tense logic in hand, the next step is to identify a relation which can play analogous role to extension in the particular-universal case. It should be a binary relation on the set of tense logic models and it should be one which preserves the truth of sentences concerning the past but not those concerning the future. A natural idea is that we might employ the relation of *future-switching*: we keep the model the same up to and including the present time—but allow it to change after that. If future-switching can change the truth-value of a sentence, then the sentence is surely at least in part about the future.

A simple way to implement the idea would be to say that one model is a future-switch of another just in case the models are identical (up to isomorphism) except that for any time t later than the present, and any sentence letter ϕ, the values of $I(\phi, t)$ may differ. That is, we could define the relation as follows:

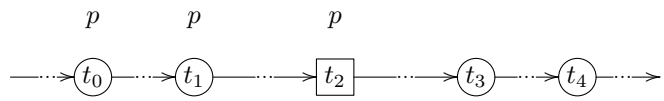
Definition 5 (Basic Future-Switching)**.** *Let M, N be tense logic models, where $M = (T_M, <_M, n_M, I_M)$ and $N = (T_N, <_N, n_N, I_N)$. We say that N is a basic future-switch of M ($M \prec N$) just in case there is a 1-1 function $f : T_M \to T_N$ such that:*

1. *$f(n_M) = n_N$.*
2. *$u <_M v$ if and only if $f(u) <_N f(v)$ (for all $u, v \in T_M$)*
3. *$I_M(\phi, t) = I_n(\phi, f(t))$ for all $t \leq n_M$, and all sentence letters ϕ*

The basic future-switch relation is reflexive, symmetric and transitive. Here are pictures of two models that stand in the basic future-switch relation:



Model M



Model N

Let $f(t_n) = t_n$, for all n. This function is 1–1, and, as you can see from the pictures, $t_2 = n$ in the first model and $f(t_2) = t_2 = n$ in the second. Moreover, the ordering relation $<$ is preserved, and, crucially, for all moments prior to the present, each model gives the *same* value to the sentence letters p and q. But the two models are not identical. They differ in which values they assign to sentence letters relative to future times. In particular, $I_M(q, t_4) = 1$ but $I_N(q, (f(t_4)) = I_N(q, t_4) = 0$

Some sentences that are about the future will be fragile with respect to basic future-switching. For example, $I_M(Fq, n) = 1$ but $I_N(Fq, n) = 0$. So basic future-switching can make Fq "go false". Some sentences cannot be made false by basic future-switching. For example, Pp: if that is true in a model, the definition of basic future-switching guarantees that it is true in all future-switches of that model.[9]

3.3 Past and Future Sentences

I've been using the expressions 'preserved' and 'fragile' in an informal sense. Now I will define them more carefully, so that we can use them in talking about all the barrier theses:

Definition 6 (Preservation)**.** *A sentence ϕ is preserved with respect to a relation R (such as extension, or future-switching) on a class of models U if and only if, for all models $M \in U$, if $V_M(\phi) = 1$ then for all models $N \in U$ such that $R(M, N)$, $V_N(\phi) = 1$ as well.*

Definition 7 (Fragility)**.** *A sentence ϕ is fragile with respect to R iff, for all models $M \in U$, if $V_M(\phi) = 1$ then there is at least one model $N \in U$ such that $R(M, N)$ and $V_N(\phi) = 0$.*

Note that particular sentences are extension-preserved (relative to models for first-order logic) and universal sentences are extension-fragile. We can also define Past and Future sentences:

[9]The full version of this section includes a more sophisticated future-switch relation that also allows changes in the structure of models after the present moment.

Definition 8 (Past sentences (simple version))**.** *A sentence ϕ is Past iff it is future-switch preserved.*[10]

The intuitive idea is that sentences which are genuinely just about the past will have their truth-values *preserved* over future-switches. Since they are not making any claims about the future, changing the future won't make them false. An example would be Pp. Suppose that is true in some model:

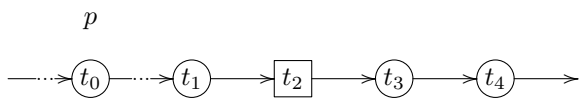


Figure 3: A model in which Pp is true will have no future-switch in which Pp is false, so we classify the sentence as *Past*.

Sentences which are genuinely about the future, by contrast, are future-switch fragile; they make claims about the future, and so changing the future can make them go false:

Definition 9 (Future sentences (simple version))**.** *A sentence ϕ is Future iff it is future-switch fragile.*

Fp is an example of a Future sentence. Suppose Fp is true in a model. Then there is some time t later than n where $I(p, t) = 1$. We can construct a future-switch of that model by changing the value of p at all times later than n to 0. (See Figure 4.)

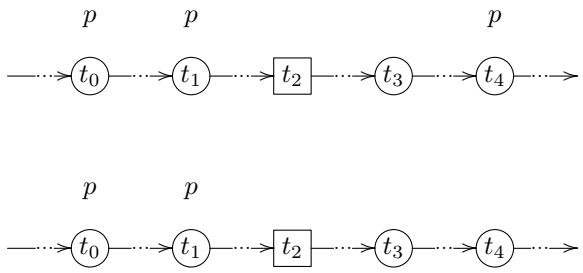


Figure 4: A model (above) in which Fp is true and a future-switch of that model (below) in which it is not.

[10] As we've defined them, Past Sentences might more accurately be called 'past or present sentences', since sentences that are intuitively just about the present don't change their truth-values if we are only varying what goes on at later times. But 'Past or Present Sentence' is a pain to say, read, and think, so I'll just call them 'Past Sentences.'

3.4 The General Barrier Theorem

We could proof the past/future barrier as we proved the particular/general one, but instead we'll prove a more general theorem, of which all our barriers will be consequences. Suppose we have a formal language L, a set of models U, and a binary relation on those models R. Then we define our various properties of sentences and sets as follows:

Definition 10 (R-preserved)**.** *A sentence ϕ is R-preserved iff, for all models $M \in U$, if $V_M(\phi) = 1$ then for all models $N \in U$ such that $R(M, N)$, $V_N(\phi) = 1$.*

Definition 11 (R-fragile)**.** *A sentence ϕ is R-fragile iff, for all models $M \in U$, if $V_M(\phi) = 1$ then there is at least one model $N \in U$ such that $R(M, N)$ and $V_N(\phi) = 1$.*

On our definitions, particularity is a matter of being extension-preserved, and generality a matter of being extension-fragile. Pastness is a matter of being future-switch-preserved and futurity a matter of being future-switch fragile. Now we can formulate a General Barrier Theorem:

Theorem 12 (General Barrier Theorem (version 1))**.** *No set of satisfiable sentences Γ, each of which is R-preserved, entails a sentence δ which is R-fragile.*

The proof has the same simple structure as our proof of the particular-general barrier theorem:

> *Proof.* Suppose Γ is a satisfiable set of R-preserved sentences and δ is R-fragile. Let M be a model which satisfies Γ. Either δ is true in M or it isn't. If it isn't, then M is a counterexample showing that $\Gamma \nvDash \delta$. But if δ is true in M, then since δ is R-fragile there is some M' such that $R(M, M')$ and δ is not true in M'. Since each member of Γ was particular, each member of Γ is also true in M'. Hence M' is our counterexample, and $\Gamma \nvDash \delta$. □

Both the particular-general and the past-future barrier theorems follows as special cases.

3.5 More on counterexamples

$Pp \vDash FPp$ An interesting feature of FPp is that, like $Fa \vee \forall x Gx$, there is "more than one way to make it true", and this leads to its being classified as neither future nor particular. For if it is true because at some point in the past p is true, then it will be preserved with respect to future-switching. But if p has never been true and will not be true until some point in the future, (after which Pp will be true, and hence at the present moment FPp is true) then it will be fragile with respect to future-switching instead, as figures 5 and 6 illustrate:

FPp is thus classified as neither Past nor Future. This classification is promising, because, like disjunctions of past and future claims, such sentences can be

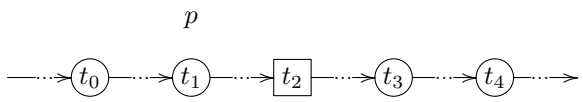


Figure 5: FPp is true in this model and no future-switch makes it false.

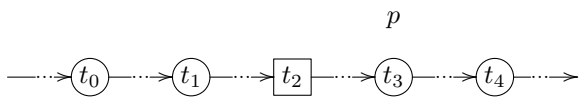


Figure 6: FPp is true in this model and some future-switches make it false

used in Prior-like arguments against the past-future barrier. Look, our Prior-like interlocutor might say, here's an argument from past-premises to future-conclusion:

$$\frac{Pp}{FPp}$$

Then if we deny that the conclusion here is really 'Future', they can respond that this will serve as a counterexample instead:

$$\frac{\begin{array}{l} FPp \\ \neg p \wedge \neg Pp \end{array}}{Fp}$$

The similarity to Prior's dilemma is striking, and of course the solution is the same: FPp is *neither* Past nor Future, and so neither argument above is from a set containing only past sentences to a future one.

4 Two Problems that Motivate Revisions

Problem 1: Vranas' Objection Vranas' objection concerns the sentences which our definitions classify as "neither". When Greg Restall and I wrote (Restall and Russell, 2010), we thought that the fact that mixed, disjunctive statements such as $Fa \vee \forall x Fx$ were classified as "neither" was a nice consequence of the definitions which allowed us to respond to Prior's dilemma. Since then, some helpful (slightly too helpful!) criticism from Peter Vranas has persuaded me that this isn't quite right.[11]

[11] Professor Vranas' criticisms were first presented to me in his comments at the Formal Epistemology Workshop in Austin in 2005, and they are published as (Vranas, 2010), along with (Russell, 2010) and (Restall and Russell, 2010) in the collection (Pigden, 2010).

Here is the problem. Take the conditional $Fa \to \forall x Gx$. Material conditionals are equivalent to disjunctions (in the presence of negation), so it isn't too surprising that a "mixed conditional" like this is also classified as "neither".[12] Suppose we were using our general strategy to prove Hume's Law in a deontic logic.[13] Then we would instead be interested in conditionals like

(9) $\phi \to O\psi$

Suppose that ϕ is descriptive, but that $O\psi$ is normative, making this a mixed conditional. The conditional would, then, be classified as *neither* descriptive nor normative. Our barrier theorem has nothing to say about sentences which are neither and so would say nothing about (9). But, says Vranas, intuitively Hume's Law *does* say something about conditionals like (9). That conditional might translate a natural language sentence like:

(10) If Alice is a first-year, then she ought to hang her coat on one of the blue hooks.

Intuitively, this is just the kind of thing that Hume's Law says is not entailed by any set of purely descriptive sentences. If you think you can't get an *ought* from an *is*, then you think you can't get (10) from an *is*. And so if the barrier theorem says nothing about it, then the barrier theorem is too weak to be Hume's Law—it doesn't say enough. That is part one of Vranas' objection.

Part two makes everything worse. Suppose we were to respond to Vranas by strengthening the theorem. We tinker with the definitions and make it say that $\phi \to O\psi$ doesn't follow from a set of purely descriptive sentences. Such a revised thesis would be *false*, because the conditional follows from $\neg\phi$:

(11) $\neg\phi \vDash \phi \to O\psi$

So this is the problem: leave the barrier as it is, and it is too weak to be Hume's Law. Strengthen it in the relevant ways, and it would be too strong to be true.

A Response to Vranas' Objection

> "as plants are contained in their seeds, not as beams are contained in a house." (Frege, 1884)

We have been dealing with a number of model-theoretic properties, including truth-in-a-model, satisfiability, and of course now, universality and particularity, and being Past and being Future. We have defined the new properties for individual sentences, but they are very naturally extended to *sets* of sentences as well. For example, with truth-in-a-model we will say:

[12] If it is true in a model because $\neg Fa$ is true, then then it is true in all extensions of that model, but if it is true in a model because $\forall x Gx$ is true (though $\neg Fa$ is not) then there will be extensions which make it false.

[13] I switch to talking about Hume's Law here (as Vranas does) because I think the intuition that Vranas is right here is strongest for this case.

Definition 13 (Truth in a model (sets)). *A set of sentences Γ is true in a model M if for all sentences $\gamma \in \Gamma$, γ is true in M.*

Sometimes a set can have a model-theoretic property P even if none of the individual members has P. For example, we call a sentence is *unsatifiable* if there is no model which makes it true. Reasonably enough, we can extend this to sets as follows:

Definition 14 (Unsatisfiability (sets)). *A set of sentences Γ is unsatifiable if there is no model which makes every sentence $\gamma \in \Gamma$ true.*

There are cases where $\Gamma = \{\gamma_1, \gamma_2\}$, where γ_1 and γ_2 are both satisfiable sentences, but still Γ is an unsatisfiable set, for example when $\gamma_1 = Fa$ and $\gamma_2 = \neg Fa$. We will define a particular *set* of sentences as follows:

Definition 15 (Particularity (for sets)). *A set of sentences Γ is particular if and only if, if it is true in a model M, it is also true in all extensions of M.*

Then we will use this to reformulate the particular-universal barrier theorem as follows:

Theorem 16 (Particular Universal Barrier Thesis (version 2)). *No (satisfiable) particular set of sentences entails a universal sentence.*

Proof: Assume that Γ is satisfiable and particular, and that δ is universal. Since Γ is satisfiable, there is some model M which makes every sentence in Γ true. If δ is false in M, then M is a counterexample to the entailment. If not, there is still some extension of M, (call it N) in which δ is false (from the definition of universality.) But Γ is true in N (from the definition of particularity.) Hence N is a counterexample to the entailment and so either way, $\Gamma \nvDash \delta$.

Let's look at how the new barrier theorem treats Prior's two arguments:

$$\frac{Fa}{Fa \vee \forall x Gx} \qquad \frac{\begin{array}{l} Fa \vee \forall x Gx \\ \neg Fa \end{array}}{\forall x Gx}$$

The set of premises in the first argument is particular, but the conclusion is not universal, so it is not a counterexample. The set of premises in the second argument is *universal*. And so neither argument is (per impossible) a counterexample to the theorem.

One central feature of our new treatment of the second argument is that it contains two premises, neither of which is universal, but the premise *set* itself is nonetheless universal. Our treatment of mixed sentences might be thought of like this: they contain the seeds of universality—so they are not strictly particular when considered on their own—but whether or not they actually contribute any universality to the premises depends on the soil that they find themselves in—that is, it depends on what *other* premises are in the set. If our

mixed premise is $Fa \to \forall x Fx$ then its universality will be "activated" in a set that also contains Fa but not in a set whose only other member is $\neg Fa$.

This is a satisfactory solution to Prior's dilemma, but then we had one of those before, it was just that it ran into a further problem. What can we say now about Vranas' objection, namely that we want to be able to say (something close to this anyway) that mixed sentences like $Fa \to \forall x Gx$ do not follow from descriptive premises? Well, no-one should say *that* quite generally, since it is false because of examples like (11). Because of this, we cannot make this claim: no particular set of premises entails a mixed-or-universal sentence.

But there is something notable about the two arguments in which we get mixed conclusions from descriptive premises: they are both cases in which the truth of the premises is sufficient to rule out any models of the mixed sentence which have extensions that make that sentence false. Perhaps all such arguments from particular premises to mixed conclusions are like this, i.e.

Definition 17 (Premise-relative particularity)**.** *A sentence δ is particular relative to a set of premises Γ if $\Gamma \cup \{\delta\}$ is a particular set.*

Now we might suspect that:

(12) If Γ is a particular set and δ a mixed conclusion, then $\Gamma \vDash \delta$ only if $\Gamma \cup \{\delta\}$ is particular.

And in fact we can prove this:

Proof: Suppose Γ is particular and δ mixed, and $\Gamma \vDash \delta$. Either Γ is satisfiable or it isn't. If it isn't, then $\Gamma \cup \{\delta\}$ isn't satisfiable either, so it is trivially particular. If there are models which satisfy Γ, then since Γ is particular, it is true in all extensions of any model in which it is true. And since $\Gamma \vDash \delta$, δ is true in those extensions too, and hence $\Gamma \cup \{\delta\}$ is particular.

Putting (12) together with version 2 of the barrier theorem (which we proved already) and noting that every sentence is particular relative to an unsatisfiable set, we get:

Theorem 18 (Particular-Universal Barrier Theorem for First Order Logic (Version 3))**.** *No particular set of sentences Γ entails a sentence δ unless δ is particular relative to Γ. A fortiori, no such Γ entails a universal sentence, or even a sentence which is universal relative to Γ.*

This new version of the barrier theorem does say something about mixed conditionals. And what it says is true.

Problem 2: Identities between Names

Consider identity statements between names, such as $a = a$ or $a = b$. Just as such sentences express necessary propositions (if true), so they also express propositions that (if true) are true across all times. Part of the reason for this is that just as a is a modally–rigid designator, designating the same object with respect to every possible world, so it is also a temporally–rigid designator,

designating the same object with respect to all times. As a claim about our formal system, this is merely a technical fact, but the same also seems to be true about names in English: for the sentence "Aristotle is alive" to be true at a time t, it has to be the case that *Aristotle*—that guy, the guy the name refers to now—is alive at t.

Given this, if $a = b$ is true in a model, then it is true at all times in the model. Future-switches of models are only allowed to change the extension of predicates at times after the present, so any future-switch of a model in which $I(a = b, n) = 1$ is also one in which $I(a = b, n) = 1$. But again, identity statements, if true in a model, are true at all times in the model. So there is *no* future-switch of a model in which $a = b$ is true that makes $a = b$ false. The sentence is preserved with respect to future-switching. And our definitions classify a sentence that behaves like that as Past.

This doesn't seem right. On the surface, $a = b$ looks like a claim about the present; it certainly doesn't contain any tense operators. Moreover, if one began to suspect that it might contain disguised temporal content, much as "only" has disguised universal content, then one is likely to suspect that it has disguised *Future* content: after all, the following is a valid argument:

$$\frac{a = b}{G(a = b)}$$

And even though

$$\frac{a = b}{H(a = b)}$$

is valid too, this would at most show that $a = b$—like $Fp \wedge Pq$—has a little of both, and so should be classified as Future, as that conjunction is.

So the classification of $a = b$ as Past looks like a mistake. We are going to fix it and happily, the fix will take care of the slightly odd status of logical truths that we noted in section 2 at the same time.

How to deal with identities between names and logical truths, all at once

Suppose we were trying to define a conception of 'Future sentence' for use in *both* the past-future barrier thesis, and also its reverse: the future-past barrier thesis. Given our strategy, sentences belonging to the premise class for a barrier must be preserved with respect to some relation, and sentences belonging to the conclusion class must be fragile with respect to it. So if our conception of a Future Sentence is to give us the premise class in the case of one barrier, and the conclusion class in the case of another, we'll have to define Future sentences as fragile with respect to one relation, but *preserved* with respect to another. We've already said that Future sentences are fragile with respect to future-switching. What relation should we say they are preserved with respect to? How about this: past-switching.

Past-switching is just like Future-Switching, except this time you have to keep the future (and the present) fixed, and you can change the value I gives for any times before n. As before, there is a basic version, and a more sophisticated version that allows for changes to the structure of time in the past.

Definition 19 (Basic Past-Switching)**.** *Let M, N be tense logic models, where $M = (T_M, <_M, n_M, I_M)$ and $N = (T_N, <_N, n_N, I_N)$. We say that N is a basic past-switch of M ($M \prec N$) just in case there is a 1-1 function $f : T_M \to T_N$ such that:*

1. *$f(n_M) = n_N$.*
2. *$u <_M v$ if and only if $f(u) <_N f(v)$ (for all $u, v \in T_M$)*
3. *$I_M(\phi, t) = I_n(\phi, f(t))$ for all $t \geq n_M$, and all sentence letters ϕ*

Now we use both Future-switching and Past-Switching to define our classes of sentences.

Definition 20 (Just Past)**.** *A sentence ϕ is Just Past iff it is preserved with respect to Past-switching and fragile with respect to Future-switching.*

Definition 21 (Just Future)**.** *A sentence ϕ is Just Future iff it is preserved with respect to Past-switching and fragile with respect to future switching.*

These new definitions are more demanding than the previous ones. We retain e.g. the old condition on being a future sentence (being future-switch fragile) and add another (being preserved with respect to past-sentences.) As a result, some of the sentences which were previously classified as Past or Future are now neither. In particular:

- logical truths like $p \vee \neg p$ do not count as Just Past because they are not fragile with respect to past-switching.
- identity statements between distinct names like $a = b$ do not count as Just Past because they are not fragile with respect to past-switching either.
- tense operator–free sentences which are neither unsatisfiable nor logical truths, like $p \wedge q$, are neither Just Present nor Just Future, since they are fragile with respect to neither past- nor future-switching. (These are claims that one might think of as Present, as opposed to Past or Future.)
- conjunctions of Just Past and Just Future sentences like $Pp \wedge Fq$ are neither Just Past nor Just Future, since they are neither Future-switching nor Past-switching preserved.

Here is a table summarising some of the results:

Just Past	Hp, Pp $P\neg p$, $\neg Pp$ $Hp \wedge Pq$ $Hp \rightarrow Pq$
Just Future	Fp, Gq $F\neg p$, $\neg Fp$ $Fp \wedge Gq$ $Fp \rightarrow Gq$
Neither Just Past nor Just Future (class 1) (Fragile with respect to *both* $\prec$ and $\succ$)	$Fp \wedge Pp$
Neither Just Past nor Just Future (class 2) (neither fragile nor preserved with respect to at least one of$\prec$ and $\succ$)	$p \vee Fq$, $p \rightarrow Fq$ $Pp \vee Fq$, $Pp \rightarrow Fq$ PFp
Neither Just Past nor Just Future (class 3) (preserved with respect to both $\prec$ and $\succ$)	p, $p \wedge q$ $p \vee \neg p$ $Fp \vee \neg Fp$ $Hp \vee \neg Hp$ $a = a$ $a = b$
Unsatisfiable sentences	$p \wedge \neg p$

Definition 22 (Just Past (for sets))**.** *A set of sentences* Γ *is Just Past if and only if,* Γ *is future-switch preserved and past-switch fragile.*

For example, the set $\{Hp, Pq, Hp \rightarrow Pr\}$ is Just Past, but $\{Hp, p\}$ is not, since p is not fragile with respect to Past-Switching.

Definition 23 (Just Future (for sets))**.** *A set of sentences* Γ *is Just Future if and only if* Γ *is past-switch preserved and future-switch fragile.*

For example, the set $\{Fp, Fq, Fq \rightarrow Gr\}$ is Just Future, but $\{Gp \rightarrow q\}$ is not, since if q is true in a model, no future-switch will make the set false.

Definition 24 (Just Past relative to a set Γ)**.** *A sentence* ϕ *is Just Past relative to a set of sentences* Γ *if and only if,* $\Gamma \cup \{\phi\}$ *is future-switch preserved and past-switch fragile.*

For example, FPp is Just Past relative to the set $\{Pp\}$. If $\{Pp\} \cup \{FPp\}$ is true in a model, Pp will be true in all future-switches of it, and that will make FPp true too. Hence the set is future-switch preserved. It is also past-switch fragile—a past-switch can always make Pp false. Since $\{Pp\} \cup \{FPp\}$ is future-switch preserved and past-switch fragile, it is Just Past, and so we say that FPp is just Just Past relative to $\{Pp\}$.

However, FPp is not Just Past on its own, since it is not future-switch preserved (if it is true in a model like the one in figure 6, no *past*-switch will make it false.)

The new definitions solve some of our classification problems. Now we need to formulate the barrier theorem in terms of them. A first pass might be:

(13) No Just Past set of sentences Γ entails a sentence which is Just Future.

But the set of sentences which is Just Future excludes sentences like $Fp \wedge Hp$, and always excludes FPp. The intuitive barrier thesis is stronger than this; it says that no set of sentences *just* about the past entails a sentence that is *even in part* about the future. $Fp \wedge Hp$ is in part about the future, so this should be included in the conclusion class. FPp is one of the sentences that changes its status depending on the other premises; so we would like to be able to include it in the conclusion class in the right contexts. So the kinds of sentences that we would like to be able to say don't follow from sets of sentences which are Just Past include:

- Fp (things just about the future)
- $Fp \wedge Pp$ (things about the future but also about the past)
- $PFp/Fp \vee Pp$—when these are not Just Past relative to the premise set

A couple of propositions will help us here:

Proposition 25. *Any sentence which is Future-switch fragile is future-switch fragile relative to every set of sentences.*

Proof. Suppose ϕ is future-switch fragile and suppose (for reductio) that there is some set of sentences Γ such that $\Gamma \cup \{\phi\}$ is not future-switch fragile. Then there is a model (call it 'M') of $\Gamma \cup \{\phi\}$ that has no future-switches that make $\Gamma \cup \{\phi\}$ false. But M is a model of ϕ and since ϕ is future-switch fragile, all its models have future-switches that make ϕ false. So M has a future-switch (call it 'N') that makes ϕ false. So $\Gamma \cup \{\phi\}$ is false at N too. Contradiction. $\square$

This means we can identify the conclusion class as the class of sentences which are future-switch fragile relative to the premises, and this will cover the kinds of sentences we want, and give us the stronger claim:

Theorem 26 (Past/Future Barrier Theorem)**.** *No Just Past set of sentences* Γ *entails a sentence which is future-switch fragile relative to* Γ.

Proof. Suppose that Γ is Just Past. Then Γ is $\prec$-preserved. Suppose also that ϕ is $\prec$-fragile relative to Γ. We show $\Gamma \nvDash \phi$.

- If Γ is unsatisfiable, then $\Gamma \cup \{\phi\}$ is unsatisifiable, and so ϕ is not $\prec$-fragile relative to Γ. Contradiction. So Γ is satisfiable.
- Since Γ is satisfiable there is a model M which makes it true.
  - If M does not make ϕ true then $\Gamma \nvDash \phi$.
  - If M does make ϕ true, there is a model N, MRN such that $\Gamma \cup \{\phi\}$ is false at N (this is what it means for ϕ to be $\prec$-fragile relative to Γ.) But Γ is $\prec$-preserved and hence true at N. Hence ϕ must be false at N. And again $\Gamma \nvDash \phi$.

□

This theorem—which makes use of both past- and future-switching—is strictly weaker than the previous one. We have excluded new things from the premise class, namely, theorems and sentences like $a = a$, and of course, it is easier to show that you can't "get stuff" from a smaller premise class. So perhaps from a strictly logical perspective, this theorem is not really progress. However, it does have a serious advantage if our aim is to explicate barrier theses, namely that the interpretations of 'Past' and 'Future' used in this version are much closer to the concepts that feature in the intuitive Past/Future barrier.

5 An Informal Approach

We saw that one way to deal with informal counterexamples is to translate them into a formal language, but not every informal argument is amenable to translation into the language of some familiar, uncontroversial logic. A obvious kind of example is the kind that exploits metalinguistic techniques and/or propositional attitudes:

Bob uttered the English sentence "everyone is happy."
The sentence Bob uttered is true.

Everyone is happy.

One problem is that enriching a logic so that it includes either metalinguistic predicates like "true" or propositional attitude verbs like "believes" is very difficult. I suspect that no-one knows how to do this well yet, and certainly, there is no approach that is so uncontroversial that I can exploit it here.

But we might try something else. When Kripke introduced the ideas of rigid and non-rigid designators in (Kripke, 1980), these were clearly inspired by their formal counterparts in the model theory for first-order modal logic. In such a theory, individual constants are assigned their referents independently of possible world, whereas the extension of a complex term can vary with the world-relative extension of some of its parts. Moving to natural language, Kripke tells us that a term is a *rigid designator* if it refers to the same object with respect to every possible world. By contrast it is a *non-rigid designator* if it does not refer to the same object with respect to every possible world. This is an example of ideas from model theory being brought over into the study of natural language and used to characterise natural language expressions.

Rigidity is a property of a term, but universality and particularity as we have defined them above, are properties of *sentences*. So if we are looking to define natural language counterparts for them, we will be defining them for sentences, rather than terms. Rigidity was a matter of what stays constant (i.e. referent) when we vary possible world. Universality is not a modal notion in this way, but it is clearly a matter of what stays constant (truth) when we change something else (increase the size of the domain.) So here is our definition of a particular sentence for natural languages:

Definition 27 (Particular (natural language sentences))**.** *A particular sentence is a sentence which, given situation which would make it true, will remain true if new objects are introduced to the situation.*

For example, the sentence *Aristotle is a philosopher* is particular. Given any situation in which it is true, simply adding new objects to the situation—Alexander, Aristotle's dog, Plato (non-philosopher, non-philosopher, philosopher respectively)—would not affect its truth-value.

Aristotle is the only philosopher however, is not particular. Suppose we were in a situation where it was true. Adding new objects to the situation can make it false. For example, if we were to add Plato, who is a philosopher. Rather, we will say that *Aristotle is the only philosopher* is universal.

Definition 28 (Universal Sentence (natural language))**.** *A universal sentence is one which, given any situation which would make it true, can be made false by adding new objects to the situation.*

For example, *everyone is philosopher* is universal because, given any situation in which it is true, we can always make it false by adding someone who is not a philosopher. *Aristotle is the only philosopher* is universal too—it becomes false once we introduce add Plato—only in this case its universality is perhaps more hidden. So again, one feature of our definitions is that they allow us to reveal universality in sentences that don't wear that universality on their sleeves, in that they don't contain any obvious 'every' or 'all' or 'each.'

You can think of situations in a very intuitive way—maybe you would like to think of them as incomplete possible worlds or something, and I don't really object to that if you prefer it (pending further specification of what an incomplete possible world is), but I have in mind something much more innocent. The same goes for "adding more objects" too, though perhaps, given my audience, I should note some explicit restrictions on that. You can add more objects by imagining that Plato shows up, or Plato's dog, or a person who isn't fond of dogs, or a black swan, or 10 white ravens. You can't add objects that are already in the situation, or that overlap things already in the situation—for example, the way I am thinking of it, you can't add a *second* Aristotle (though you might add someone who looks like him, or who also *claims* to be him, or someone else *called* "Aristotle.") You can't add objects by cutting objects which are already included in half. You can't add objects like "a person who is mortal but has the special secret superpower of making "all men are mortal" false". I don't currently have formal definitions of "situation" and "adding objects" that would rule these things out, but I think the informal notion is a fair start anyway.

We can also classify *sets* of natural language sentences as particular.[14]

[14] I'm sort of tempted to call a set of natural language sentences a "paragraph" but maybe that would be too cute. Anyway, there's nothing to stop us putting natural language sentences in sets, even if the language of set theory is not a part of natural language. When we describe natural language sentences as *particular* or *universal*, or when we talk of them entailing other sentences, we have already started to use a metalanguage. In that metalanguage we'll help ourselves to some technical expressions not in the object language, including the language of set theory.

Definition 29 (Particular Set (of natural language sentences))**.** *A set of natural language sentences is particular if any situation in which every member of it is true, will still be such that every member of it is true whenever we add objects to the situation.*

The set {Alice is a philosopher, Bob is not a philosopher} is particular, because no matter what objects we add to the situation in which both sentences are true—e.g. Carl, a philosopher, Doris, a non-philosopher etc.—the sentence will remain true. And we can say that sentences like *Carl is a philosopher or everyone is a frog* are neither universal nor particular simpliciter, but only relative to certain sets of sentences.

Definition 30 (Particular relative to a set Γ)**.** *A sentence δ is particular relative to a set Γ if $\Gamma \cup \{\delta\}$ is particular.*

Now we can formulate a natural language counterpart to the metalogical particular/universal barrier theorem:

Theorem 31 (Particular-Universal Barrier)**.** *If Γ is a set of natural language sentences which is particular, and δ is a natural language sentence which is not particular relative to Γ (i.e. it is universal or neither) then δ is not a logical consequence of Γ.*

So now what ought we to say about this proposed counterexample?

Bob uttered the English sentence "everyone is happy."
The sentence Bob uttered is true.

Everyone is happy.

The first premise is particular—adding new objects to the situation won't change the truth of *Bob uttered the English sentence "everyone is happy"*. The second is neither particular nor universal; consider situations that make *The sentence Bob uttered is true* true. One might be a situation in which Bob uttered *2+2=4*. Another might be one in which he uttered *all men are mortal.* In the former, adding new objects won't change the truth-value, because that sentence he uttered is true in all situations. In the latter, adding a non-mortal man would make *The sentence Bob uttered is true* false. Where Bob utters a universal sentence, such universality infects the metalinguistic sentence which attributes truth to that utterance. So the second premise is neither. (And of course the conclusion is universal.)

Now we should consider the set of both premises. Suppose we have a situation that makes both sentences true. Naturally that will be one in which everyone is happy. Now suppose we add a new object—an unhappy person. The set of sentences is no longer true. So premise *set* is not particular but universal, and so the argument is not a counterexample to the informal particular-universal barrier thesis.

References

Black, M. (1964). The gap between "is" and "should". *The Philosophical Review*, 73(2):pp. 165–181.

Frege, G. (1980/1884). *The Foundations of Arithmetic*. Northwestern University Press, Evanston, Illinois, second edition.

Hare, R. M. (1991/1952). *The Language of Morals*. Oxford University Press, New York.

Hodges, W. (1997). *A Shorter Model Theory*. Cambridge University Press, Cambridge.

Hume, D. (1975/1748). *Enquiries concerning Human Understanding and Concerning the principles of morals*. Oxford University Press, Oxford. edited by L. A. Selby-Bigge.

Jackson, F. (1971). Defining the autonomy of ethics. *The Philosophical Review*, 83:88–96.

Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives. In Almog, J., Perry, J., and Wettstein, H., editors, *Themes from Kaplan*. Oxford University Press, New York.

Kripke, S. A. (1980). *Naming and Necessity*. Blackwell, Oxford.

Pigden, C., editor (2010). *Hume, Is and Ought: New Essays*. Palgrave MacMillan, Houndsmill, Basingstoke, UK.

Popper, K. (1948). What can logic do for philosophy? *Aristotelian Society supplementary*, xxii.

Prior, A. N. (1960). The autonomy of ethics. *The Australasian Journal of Philosophy*, 38:199–206.

Restall, G. and Russell, G. (2010). Barriers to Implication. In Pigden, C., editor, *Hume, Is and Ought: New Essays*. Palgrave MacMillan.

Russell, G. (2010). In defence of Hume's law. In Pigden, C., editor, *Hume, Is and Ought: New Essays*. Palgrave MacMillan.

Schurz, G. (1997). *The Is/Ought Problem: An Investigation in Philosophical Logic*. Kluwer, Dortrecht.

Searle, J. R. (1964). How to derive 'ought' from 'is'. *The Philosophical Review*, 73:43–58.

Vranas, P. (2010). Comment on 'barriers to implication'. In Pigden, C., editor, *Hume, Is and Ought: New Essays*, pages 260–267. Palgrave MacMillan.