A response to Humberstone's review of *Barriers to Entailment*[*]

Gillian K. Russell

(draft of a paper written for a Festschrift for Lloyd Humberstone)

1 Prelude

In 2024 Lloyd Humberstone wrote a review of my book, *Barriers to Entailment*. As online commentators noted with some amusement, the review is 73 pages long and so it was fortunate that it was for a journal, *The Australasian Journal of Logic*, which has celebrated its capacity to publish longer pieces.

Reading public reviews of one's work—perhaps by very senior members of the profession, or by people one knows and admires—is a singular experience. More pointedly: wading through strawmen, lazy misreadings, and faint praise in order to figure out what, if anything, is worth responding to, either in person, on the internet, or in print, is a personal and professional minefield that can stir up rather negative feelings. *Where*, one thinks, *is the complimentary comparison with the Tractatus that I dreamt about?*[1]

Humberstone is himself a senior logician and philosopher, one I have known and admired for a long time, and my predominate feeling on starting to read was, well, *terror*. But though his review is critical, once I had worked my way through it, it became a source of much more positive feelings.

For one, it is strangely exhilarating to read an accurate exposition of one's own work—even as a prelude to a critical response. One has a sense that for once one has not been ignored or misinterpreted; rather, someone has *heard* and understood. And it turns out that successful academic communication can be an absolute joy.

For another, a good critic's objections and questions can take your understanding of the subject further, so that you finish reading their review with a

[*]On page 222 of his review Humberstone notes that *Barriers* often speaks of 'responses' to or 'responding to' an argument and he (Humberstone) thinks it helpful to distinguish a *reaction* to an argument from a *reply* to it: a reaction concedes the argument's force, a reply queries it. This seems like a helpful regimentation which I am happy to adopt (reaction!) and I thus deliberately choose 'response' here in the title to signify that I plan to do a little of each.

[1]My heartfelt thanks to Christopher John Searle in *Philosophy Now* for finally realising that particular dream. (Searle, 2024)

sharper appreciation of the issues than before. So there is the pleasure that comes with new insights.

In this case of this particular review, I also delight in many of the asides and footnotes, which reference everything from 1960s girl groups and Latin primers to Kaplanian *terratophobia*.[2] Two stand-outs are note 16, which elucidates *Barriers*' *breakable/makeable* terminology with reference to "The Best Part of Breaking Up" by the Ronettes, and page 251, which complains (quite fairly) about my own inconsistency in referring to a theorem as the *Limited General Barrier Theorem* or (mixing up the word order) the *General Limited Barrier Theorem*:

> "It is certainly hard to remember which order the words come in, and the reviewer resorted to the following mnemonic: the correct order is given by the first four letters (which were once all there was of it) in the ever-expanding initialism LGBTQIA+ (or the common alternative version, LGBTIQA+)." (Humberstone, 2024, 251)

OK. That works for me.

Anyway, while I can hardly recommend that we all write more 70-page book reviews[3]—either considering the interests of the prospective writers or those of their readers—I hope this prelude to my response makes clear that this one made me feel seen. I am glad to have the opportunity to engage with some of the issues it raises (and add a few additional asides and footnotes of my own) in this Festschrift in Humberstone's honour.

2 Helpful Background

This response will be getting into the weeds of some pretty gnarly issues; it is not intended to motivate my view or offer grand vistas.[4] Still, a brief presentation of three issues will help to make the present reply more self-contained.

First, some orientation: the part of *Barriers* that Humberstone's review focuses on is engaged in the project of using logic—formal languages with models and entailment relations—for precisifying, in a unified way, five different barriers to entailment that are assumed or discussed in philosophy.

1. The Universal Barrier
   You can't get universal conclusions from particular premises.

2. The Future Barrier
   You can't get conclusions about the future from premises about the past.

[2] Fear of giving birth to a monster.

[3] Especially on the present topic; I learned with concern from a talk by Professor Jennifer Saul recently that the words *Barrier* and *Barriers* are included on lists of words "banned or flagged" by US Federal Agencies under Trump. *Caveat lector*, then.

[4] For that I recommend *Barriers* itself (which is also good preparation for understanding Humberstone's review.)

3. The Modal Barrier
   You can't get necessity-conclusions from premises about the actual world.

4. The Indexical Barrier
   You can't get indexical conclusions from non-indexical premises.

5. The Normative Barrier
   You can't get normative conclusions from descriptive premises.

5. is also known as *Hume's Law* and it is deeply controversial. But *Barriers* aims to show that it is really of a piece with the other, more platitudinous, barriers, and indeed, an instance of a more general Barrier *Theorem*, from which one can extract the different barriers by fixing parameters in different ways.

Second, a key move in the logical literature on this topic is from Prior (1960). Prior gives his argument in natural language and focuses specifically on the normative barrier, but once we are considering all the barriers together it is clear that a generalisation of Prior's argument works against them all. Here is a quick introduction to a generalised version of Prior's argument.

Suppose we partition the sentences of a language into two sets (e.g. descriptive and normative, or past and future.) Call these $\mathcal{X}$ and $\mathcal{Y}$. Humberstone (1996) calls the study of such classifications *philosophical taxonomy*:

> Examples include the classification of statements as material object statements or sense-datum statements, as evaluative or non-evaluative, psychological or non-psychological, a priori or a posteriori, and so on. Often the interest in question takes the form of defending or contesting an 'inferential barrier' thesis to the effect that from premisses in one class there are no arguments of some favoured type (as it might be, no valid argument, or instead, no sound arguments) to conclusions not in that class. (Humberstone, 1996, 121)

That is indeed my aim, but Prior's argument is an obstacle for any such barrier thesis. Let the X be a sentence of kind $\mathcal{X}$ and Y a sentence member of kind $\mathcal{Y}$. Then consider:

$$X \vDash X \vee Y \qquad \text{(Prior 1)}$$

Here the argument is an instance of disjunction-introduction and hence valid,[5] the premise is certainly of kind $\mathcal{X}$ and Prior suggests that the disjunctive conclusion should be considered of kind $\mathcal{Y}$ (this is in fact a good fit with what many writers say about the normative barrier, since they assume a sentence is

[5]Fans of weaker logics might like me to qualify this with "classically", though nearly all of *Barriers* presupposes a classical framework (or at least, an extended "neo-classical" framework that allows for modal, deontic, and indexical logics) but the same techniques could also be explored in alternative model theories. Those will interests in both relevant logics and this approach to barriers might check out (Weiss, 2025)

normative if it contains a normative expression and all the expressions in sentence Y will be in the disjunction $X \vee Y$.) Still, Prior knows that the most likely challenge to this counterexample to the $\mathcal{X}/\mathcal{Y}$ barrier is to say $X \vee Y$ is not really of kind $\mathcal{Y}$. To such challenge he offers a second argument:

$$X \vee Y, \neg X \vDash Y \qquad \text{(Prior 2)}$$

This time the argument is an instance of disjunctive syllogism and hence valid, the conclusion is certainly of kind $\mathcal{Y}$, and Prior's opponent has given him that the first premise is not of kind $\mathcal{Y}$. Given that the two sets partition the set of sentences, the first premise is in $\mathcal{X}$ and given one more assumption: that the negation of X is in $\mathcal{X}$, (plausible in the Hume's Law case since descriptive sentences like *Snow is white* do not become normative just by being negated) Prior 2 is a counterexample if Prior 1 is not. So whether $X \vee Y$ is of kind $\mathcal{X}$ or kind $\mathcal{Y}$, there is a counterexample.

The final background issue to be explained is the overall model-theoretic approach the book takes to formulating barriers in order to avoid counterexamples like Prior's. A key idea is that formulating precise $\mathcal{X}/\mathcal{Y}$ barrier theses will require definitions which tell us what it is for a sentence to be of kind $\mathcal{X}$ or kind $\mathcal{Y}$. This will dissolve some counterexamples by clarifying that the premises are not really of kind $\mathcal{X}$, or that the conclusions are not really of kind $\mathcal{Y}$, so that the argument in question is not really a counterexample.

The approach taken in *Barriers* is explicitly in contrast with a simple syntactic one often presupposed—sometimes explicitly avowed—in the literature on which (in the normative case) a sentence is normative if and only if it contains at least one normative expression. (In our more general setting, this condition would amount to a sentence being a member of $\mathcal{Y}$ iff it contains a subexpression which is a member of a kind $\mathcal{Y}^e$, where $\mathcal{Y}^e$ is the set of $\mathcal{Y}$-*expressions*.)) In the normative case, writers often assume that the normative expressions must be identified independently by "listing" them, e.g. *ought*, *good*, *fitting*, *magnificent* ...etc. and so this is sometimes called the *list and containment* view. Anyway, the approach in *Barriers* explicitly disavows the list and containment view for one on which sentences are classified by paying attention to model theory: whether or not a sentence is classified a certain way will depend on how its truth-value responds to changes to models.

For example, a sentence is *Universal* if extending the model's domain can take it from true to false.[6] A sentence is *Indexical* if changing the context can take it from true to false and a sentence is *Future* if changing the truth-values of atomic sentences at future times can take it from true to false. A sentence is *Modally-Universal* if adding new accessible possible worlds to a model can take it from true to false, and a sentence is *Normative* if changing the normative standards can do that.

[6] I continue the practice from *Barriers* of capitalising these categories to mark the fact that they are terms of art, and to reduce the risk of confusion with the intuitive natural language versions. The capitalised expressions are intended to be regimentations, or explications, of the intuitive notions that bring out an aspect of their meaning which is useful for formulating the barriers.

Of course, the hard work is in the details and in each case we need to say what models we are talking about, define the change to models (less dynamically, a intermodel relation) i.e. what it is for one model to be an domain-extension/context-shift/future-switch/modal-extension/norm(ative standard)-shift of another, and in the normative case especially, what I have said here so far leaves it obscure what normative standards in a model *are* and what it would be to change them. But the key idea is that the conclusion-type sentences in barriers are characterisable as *breakable* with respect to a distinctive kind of relation, R on the set of appropriate models, where:

Definition 1 (R-breakable)**.** *Where U is the set of models for a language L, a sentence ϕ in L is R-breakable just in case there is a pair of models $M, N \in U$ such that MRN, M makes ϕ true, and N makes ϕ false.*

In the the case of the Particular/Universal taxonomy, models are pairs of non-empty domains with an interpretation function, ($\langle D, I\rangle$) and the R-relation is *domain-extension*, where for models M and N, N is a domain extension of M if i) N's domain is a superset of M's and ii) N's interpretation function agrees with that of M on the interpretation of individual constants, and, when it comes to n-place predicates, on any n-tuples all of whose members were in the old domain.[7] On these definitions $\forall xFx$ is Universal (domain-extension breakable) and Fa is not. But also $\neg\exists xFx$ is Universal—despite not containing a universal quantifier—and logical truths like $\forall x(Fx \rightarrow Fx)$ are not Universal, despite beginning with a $\forall$. This is in clear contrast with the list and containment approach.

Moving now to the normative case, let models be non-empty sets of worlds W, with an actual world, @, and additional non-empty subset of W, S—"S" for the worlds that are "superb".[8] Figure 1 gives two diagramatic representations of such models, with the grey shading representing S. Since these two models differ at most in the members of set S, each is an *s-shift* of the other. Normative sentences are defined as those which are s-shift breakable and since Op is true in the model on the left and false in the model on the right, this pair of models shows that Op is classified as Normative on these criteria. Similar pairs of models can show that the Normative sentences include Pp and $\neg Pp$, but not, for example, $\Box p$ or the logical truth $Op \rightarrow Pp$.

[7] So if $\langle o_1, o_2\rangle \in I_M(P)$ then $\langle o_1, o_2\rangle \in I_N(P)$ as well. And if $\langle o_1, o_2\rangle \notin I_M(P)$ then similarly $\langle o_1, o_2\rangle \notin I_N(P)$. But if there is a new element in N's domain, o_3, one not in M's domain, then N can assign it arbitrarily to the extensions of predicates, and put it in relations with new and old elements as it sees fit.

[8] Or, Humberstone notes wryly, as in Restall and Russell (2010), "for the rather less enthusiastic "satisfactory"." (Humberstone, 2024, 215) Enthusiastic or otherwise, the intuitive idea in *Barriers* is that the worlds in the subset S are the ones that meet the model's normative standards and normative sentences are the ones that are breakable by changes to that set alone.

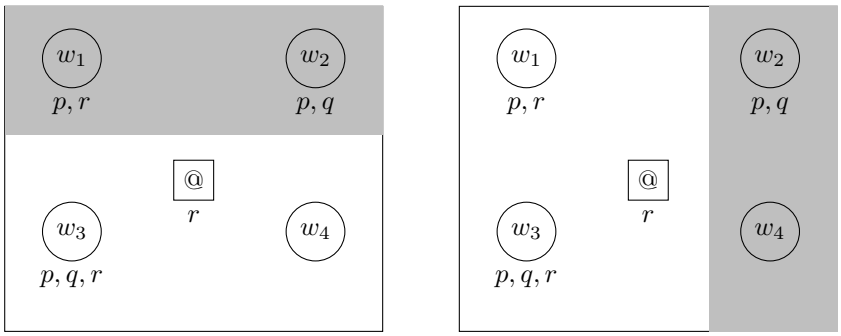

Figure 1: S-shifting: the models are the same except that on the left $S = \{w_1, w_2\}$, whereas on the right $S = \{w_2, w_4\}$.

The barriers are then formulated in terms of R-breakability, with the final upshot being that R-breakable conclusions do not follow from sets of premises that are not R-breakable—except when special conditions are met. These special conditions put the *Limited* in the Limited General Barrier Theorem, but since this limit will not be key to any of my points today, I refer the reader to *Barriers* for more details.

In the next section I turn to a question raised in section 1 of Humberstone's review.

3 Emergence and Relativism

There is a tradition that reacts to Prior's counterexamples by conceding that there are valid arguments from the descriptive to the normative, but defends the thesis that there are no *sound* such arguments, i.e. valid arguments all of whose premises are true.[9] Equivalently: whenever we have a valid argument from descriptive premises to a normative conclusion, at least one of the premises is false.

In response to Prior's pair of arguments, this approach notes that exactly one of X and $\neg X$ is true. If X is true, then $\neg X$ is false and Prior 2 is not sound. And while Prior 1 is valid and sound under this supposition—the view continues—under this supposition $X \vee Y$ is not a sentence of kind $\mathcal{Y}$. So Prior 1 is not a counterexample either. Alternatively, if $\neg X$ is true then X is false and Prior 1 is not sound. Prior 2 is valid and sound but under these circumstances $X \vee Y$ is of kind $\mathcal{Y}$, so the premises are not purely of kind $\mathcal{X}$ and Prior 2 is not a counterexample either.

It will be noted that getting this solution to work in the normative case requires i) the addition the soundness condition and ii) the idea that certain sentences—most obviously disjunctions of descriptive and normative sentences but also conditionals like $p \to Oq$—do not get classified as normative *simpliciter*

[9] See e.g. Shorter (1961).

but only relative to circumstances. If p is true, $p \vee Oq$ is not normative, if p is false, it is.

Following Karmo and Humberstone, we might think of this as a proposal to treat normativity as capable of being *world*-relative.[10] The status of the Prior-disjunction is normative in worlds where the descriptive disjunct is false, non-normative in worlds where it is true. In informal conversations with Charles Pigden[11] I have also heard the suggestion that we might say that normativity is "context-sensitive" i.e. a sentence might be normative relative to some contexts, not relative to others.

I am not fond of this suggestion about normativity, in any of these incarnations, and in the book I made some attempt to say why. Humberstone quotes that section, but here I want to have another stab at the apple. As I have sometimes said, on my view, whether a sentence is normative or not is a matter of its *truth-conditions*, rather than its syntax. This is a very loose way of drawing attention to the model-theoretic definition of normativity given above (i.e. s-shift–breakability). What we have to ask is what kinds of models make the sentence true, and what changes to them (if any) would make it false. Some sentences have truth-conditions that make reference to set S in the model—that is, to which worlds meet the normative standards. If they do this in such a way that changes to S can change the truth-value of the sentence, then the sentence is normative. But if the sentence is absolutely impervious to changes to the moral standards—e.g. the atomic sentence p, $\Box p$ in the language and model theory sketched in the previous section—then it is not normative. This is a conception of normativity on in which it is not context-sensitive, world-relative, or even model-relative. Given a set of models and a definition of denotation-and-truth in a model for the language, either there is a pair of models where one is an s-shift of the other and they each give the sentence different truth-values, or there is not. That is: either the sentence is s-shift–breakable or it is not.

When I say that this is a matter of the sentence's truth-conditions, this isn't inaccurate, but it is loose, because "truth-conditions" can mean a lot of different things. Perhaps the most common precisification in philosophy is to think of the truth-conditions of a sentence as a function from worlds to truth-values (on some ways of thinking of propositions this identifies truth-conditions of a sentence with the proposition it expresses.) That cannot be the kind of truth-conditions I mean here because I also say that a sentence is Indexical in virtue of its truth-conditions: that is, shifting the context of use can make an indexical sentence false. But this isn't a claim about the proposition the sentence expresses, but rather the antecedent fact that indexical sentences express different propositions in different contexts of use. So here we must mean something more by truth-conditions, i.e. a function that includes contexts of use in its arguments (possibly worlds as well) and yields a truth-value. There's also a sense in which sentences like $a = a$ and $a = b$ if true, while not differing in the functions they determine from contexts of use to truth-value, nonetheless determine different functions on

[10] Karmo (1988)(Humberstone, 2024, 220)

[11] The author of an alternative precisification of Hume's Law (Pigden, 1989) as well as an alternative review of *Barriers*, (Pigden, 2025).

models; the former the constant function to true, the later a more interesting one that depends upon the meaning selected by the model's interpretation function for the constants. For these reasons it makes sense to think of the kind of truth-conditions I have in mind as a function from models to truth-values. On my view the classification of sentences as *Normative*, *Universal*, *Indexical*, *Modally Universal* and *Future* depends on these functions, and is thus neither world, nor context, nor model-relative.[12]

But one of the things Humberstone does in this section of his review, is stress the similarity between the view I dislike, and a feature of the view I actually endorse and spend some time elucidating. On that view, just as we can ask whether a sentence is R-breakable (for some R), so we can ask whether a *set* of sentences is. That is, the properties of being Universal, Normative and so on are applicable to sets as well as to sentences. The mechanics of this are straightforward. First, the extend the property of truth (or truth in a model) to sets by saying that a set of sentences is true just in case every sentence in the set is true.

$$V_M(\Gamma) = 1 \quad \text{if and only if} \quad \text{for all sentences } \gamma \in \Gamma, \quad V_M(\gamma) = 1.$$

Then we say that a set Γ is R-breakable just in case there is model M in which Γ is true, and a R-related model N in which Γ is false:

Definition 2 (R-breakability (sets))**.** *A set of sentences Γ is R-breakable if and only if there is a model M such that $V_M(\Gamma) = 1$, and a model N such that MRN and $V_N(\Gamma) = 0$.*

For various reasons, Barriers formulates its theorems in terms of the status of the premise *set* rather than in terms of the status of the individual sentences: that is, what matters is that the premise *set* is non-normative, not that each of the premises is.[13]

Sometimes it is the case that a set has a model-theoretic property even though none of its members have that property. For example, unsatisfiability.

$$\{p, \neg p\}$$

The set above is unsatisfiable but each of its members is a satisfiable sentence. Rather, the property of unsatisfiability *emerges* from the way the member sentences interact. Sometimes it seems clear that a set of sentences inherits a property like s-shift–breakability from its members, e.g.:

$$\{p, Oq\}$$

[12] In the book I wrote "Whether a sentence is normative or universal is a matter of what the sentence says—its content—and in non-indexical sentences this does not vary with context." I now think that is a little sloppy, or at least, that it doesn't generalise well to the indexical case. I should really have said that a sentence's classification depends on its meaning—roughly, it's truth-conditions in the sense of a function from models to truth-vallues.

[13] Among those reasons: the quote from Foot discussed in Humberstone's review (223), informal 2-premise arguments that use disquotation.

Here Op is s-shift–breakable, as we have already seen, and so is the set above that contains it. In fact, the set is really R-breakable because Op is. But this is not always the case. The set

$$\{Oq, \neg Oq\}$$

is not s-shift–breakable, even though both its members are, because it is unsatisfiable—there is no model that makes it true, (a fortiori no such model with an s-shift that makes it false.) Here are two more interesting cases:

$$\{\neg p, p \vee Oq\}$$

$$\{p, p \vee Oq\}$$

First, the individual sentences: neither p nor $\neg p$ is s-shift–breakable but $p \vee Oq$ is.[14] So p and $\neg p$ are non-Normative, and $p \vee Oq$ is Normative.

What about the sets as a whole? $\{\neg p, p \vee Oq\}$ is s-shift–breakable, and so Normative. That is the heart of the *Barriers* response to Prior 2; the argument is not a counterexample because the premise set turns out to be Normative.[15]

By contrast, $\{p, p \vee Oq\}$ is not s-shift–breakable, even though it contains an s-shift–breakable sentence. Suppose we have a model that makes the set true. That model makes p and $p \vee Oq$ true. Can it be s-shifted to one that makes the disjunction false? No. Because p is preserved over s-shifts (it is not s-shift–breakable) and the truth of p is sufficient for the truth of the disjunction. So even though the set contains a Normative premise, it is not Normative.

One way to think about the situation intuitively is that on the *Barriers* picture normativity is a matter of truth-conditions, and when we combine truth-conditions they can interact in interesting ways to create a different overall condition, with no guarantee that Normativity—or satisfiability—will be preserved. So, $p \vee Oq$ is normative, but does its presence in a premise set make the set normative? Well, that depends.

Section 1 of Humberstone's review discusses these two strategies—the world-relativity/context-sensitivity vs. my approach on which whether or not a Normative sentence makes the set it is a member of Normative depends on its fellow set-members. I feel some anxiety that someone might finish reading this section of the review thinking that *Barriers* had rejected the former only to try to sneak it back in to solve the problems it was originally intended to solve. My first thought about *that* verdict is that it is only plausible on so loose a characterisation of the issues that one can hardly tell what is going on at all.[16] But also,

[14] Consider a model in which the disjunction is true, p is false and q is false. Oq is true in such a model, but the model can be s-shifted to one that makes Oq false. (q is false at the actual world, so you could just make that the only world in S.)

[15] Prior 1 has to fall under the *unless*-clause, since the premise is non-Normative and the conclusion, I concede, is Normative.

[16] That is, someone might think, well one side thinks that the normativity of a sentence is world-relative, and Russell thinks that whether or not a set containing two sentences, one of which is $p \to Oq$, is normative depends on *what the other sentence is*. So they both think that

Humberstone does not exactly say this himself, rather he is suggesting that the final view is one on which the significance of the classifications of individual sentences in terms of R-breakability is lost. He writes:

> "Now, *Barriers* was quoted above as [...] urging us to "resist views on which the classification of an unambiguous sentence can vary." [...] So it is worth noting that in the transition from a barrier thesis in which the classification of each premise in a multi-premise argument reflected its own individual content, as on the distributive understanding of the premises being descriptive/non-ethical, to a barrier thesis attending to the premises only collectively, the significance of that individual content is largely lost." (Humberstone, 2024, 224)

Well, sort of. The individual sentences $p \to Oq$ and Oq are classified as Normative (since s-shift–breakable). p, q, $\neg p$, $\neg q$ are classified as non-Normative. The sets $\{p, Oq\}$, $\{\neg p, Oq\}$, $\{p, p \to Oq\}$ are classified as Normative and the sets $\{p, q\}$, $\{\neg p, p \to Oq\}$, $\{Op, \neg Op\}$ as non-Normative, all on the basis of s-shift–breakability or lack of it. Here's a table that might make the situation and its patterns easier to visualise:

non-normative	normative
p	Oq
$\neg p$	$p \to Oq$
	$\{p, Oq\}$
	$\{\neg p, Oq\}$
	$\{p, p \to Oq\}$
$\{\neg p, p \to Oq\}$	
$\{Op, \neg Op\}$	

The key thing to note is that both $p \to Oq$ and Oq are normative sentences, but each is a member of a set that is Normative, as well as a member of a set that is not. If by "the significance of that individual content is largely lost" we mean only that Normative sentences can be members of non-Normative sets, so that the Normativity of the sentence is not always inherited by a set containing it, then yes, that's right. If we mean instead that the status of individual sentences makes no difference to the status of the set then that is a bit misleading. It is reasonable to say in the case of $\{p, Oq\}$ that the set is Normative in part *because* of the Normativity of one of the premises. Just because a different set—$\{Op, \neg Op\}$, say—is not Normative does not mean that, on the big picture, "the significance of that individual content is largely lost."

But Humberstone goes on to refine his point with an illustration:

---

the attribution of normativity depends on something else. They're both kinda like context-sensitivity. (At this point my brain just gives up. Attribution of normativity to the set or to $p \to Oq$? I think that getting clear on questions like this dissolves the hazy appearance of similarity.)

> "For example P_1, out of premises P_1, P_2, P_3, does not have its content attended to in any way differing from the case of an alternative to it, P_1' , say, for which $P_1' \wedge P_2 \wedge P_3$ is equivalent to $P_1 \wedge P_2 \wedge P_3$, makes the same contribution to the content of the premises collectively construed." (Humberstone, 2024, 225)

Here's one thing that is straightforwardly true about my model-theoretic approach to taxonomy: logically equivalent sentences (sentences which get the same truth-values in the same models) are classified the same way. And so for example, $\neg\exists xFx$ (logically equivalent to $\forall xFx$) is classified as a Universal sentence. And $\neg P\neg q$ (logically equivalent to Oq) is Normative. Moreover, in either case, substituting one for the other in a set of sentences will not change the status (Universal or not-Universal) of the set, since it will not change the model-theoretic truth-conditions of the set (which supervene on the model-theoretic truth-conditions of its members.) But that's not *exactly* Humberstone's point. He doesn't say that P_1 and P_1' are equivalent, but rather that the conjunction $P_1 \wedge P_2 \wedge P_3$ is equivalent to the conjunction $P_1' \wedge P_2 \wedge P_3$.

How could the conjunctions be equivalent without P_1 and P_1' being equivalent? Well, I suppose if $P_2 \wedge P_3$ were unsatisfiable, or more generally, if P_1 and P_1' were both logical consequences of $P_2 \wedge P_3$, then the two conjunctions would be equivalent without P_1 and P_1' having to be equivalent. We can devise an example of this where P_1 is normative and P_1' is not:

$$Op \wedge q \wedge \neg q$$

$$p \wedge q \wedge \neg q$$

Here the first conjuncts (Op and p) are not equivalent, but the two conjunctions are. Humberstone continues:

> Indeed, the taxonomic status of P_1 as it would be if there were no further premises, and which may differ from P_1''s corresponding solo status, plays no role in the bearing of the collective barrier thesis on the case of an inference from P_1, P_2, P_3. P_1's own specific content is washed away, to the extent that it differs from the content of any such P_1', as it is thrown into the mix with its co-premises. (Humberstone, 2024, 225)

This is correct about our example with p and Op, but it hardly seems an objection so much as an advertised feature of the view: since the normativity of a sentence is a matter of its truth-conditions, whether it makes a non-trivial contribution to the overall truth-conditions of a conjunction containing it depends on the truth-conditions of the other conjuncts. The same is true of any sentence: if we add p to a conjunction, it will be a non-trivial addition if the

only other other conjunct is q, or even $\neg p$ but not if it is p.[17] Humberstone continues:

> One might even choose to put this by saying in this setting that relative to $\{P_2, P_3\}$, P_1 has such-and-such-status—ethical or non-ethical—recording the status of $P_1 \wedge P_2 \wedge P_3$ (and coinciding in this respect with P_1''s similarly relativized status) while relative to, say P_2, P_4, P_5 (or indeed relative to $\emptyset$) it has such-and-such possibly different status on similar grounds (relative to which P_1''s status need not coincide).

This, I think, is Humberstone's real complaint. To put it perhaps a bit more strongly than he does (aiding clarity, but perhaps not charity): one might think that my view can be restated or *paraphrased*, with no changes that really matter, as one which explicitly relativises the status of a sentence to a set. So where above I described my view as being one on which $p \to Oq$ is Normative, the set $\{p \to Oq, p\}$ is Normative and the set $\{p \to Oq, \neg p\}$ is non-Normative (all on the basis of s-shift–breakability) Humberstone suggests we say instead: the status of $p \to Oq$ is a relative matter. Relative to the empty set, and to $\{p\}$, $p \to Oq$ is normative, but relative to $\{\neg p\}$ it is non-normative.

Now, Humberstone acknowledges, "it is indeed true that the kind of relativization we have just been considering does not involve as a relatum a particular world (or pointed model) to allow for an appeal to truth at which (resp. in which) to play a potentially decisive role in the classification of sentences." (Humberstone, 2024, 225) He concludes this section of the review:

> "One might tentatively conjecture then, that the only viable options for salvaging something from a Humean barrier thesis as traditionally conceived is to introduce a kind of relativity of the classification, whether to the co-premise company it keeps, as in Barriers, or to what is taken as true (in a given world, or, eliding these two for simplicity here, relative to a given model), as in Karmo-style accounts." (Humberstone, 2024, 225)

So Humberstone takes care to distinguish between relativising to a set of sentences, and the Karrmo-style accounts that relativise to a world, but I can certainly imagine some casual readers thinking that there is not so big a gap here: couldn't we just identify the set of sentences with the worlds in which it is true?[18]

[17] I suppose some readers might object to me talking about a conjunction with only one conjunct, but if so, feel free to add an additional conjunct to make the example work. I refrain here to avoid unnecessary clutter.

[18] This exact proposal would be anathema to the spirit of my project, which takes care to distinguish sentences from propositions and with that the different kinds of things that make them true. Roughly: models for sentences, worlds for propositions, and woe-betide the reader who confuses the two: but one obviously can't expect readers of a book review to have already internalised the lessons of the book it is about.

But even taking care to distinguish these two things, I want to ask whether it is really the case that the view can be paraphrased as making the normativity of a sentence relative to a set.

The flatfooted answer is *no*. The Normativity of an individual sentence, like $p \rightarrow Oq$ is a matter of its being s-shift–breakable on the set of models for the language, and this is something we can establish absolutely, and independently of any set of sentences.

But to this Humberstone might respond by pointing to the part where he says that this kind of normativity gets "washed away" (225), allowing a set of sentences like $\{\neg p, p \rightarrow Oq\}$ to be non-Normative. If we ignore s-shift–breakability as it is defined for sentences altogether, and consider only s-shift–breakability as it is defined on sets of sentences, then we might note that the conditional is sometimes the member of a Normative set and sometimes a member of a non-Normative set. Squint a bit and you might be happy to describe this as the Normativity of the sentence being "set-relative." Though of course it only works by ignoring Normativity as it is defined on sentences.

But in order to assess whether such a construal is sensible, fruitful, and worth continuing with, I think it worthwhile to consider what we might say about more familiar properties defined over models. Take *satisfiability*, where we say that a sentence is satisfiable iff there is at least one model that makes it true, and that a set of sentences is satisfiable iff there is at least one model that makes it true.[19] For definiteness consider the sentence of first-order logic, Fa. Now Fa is satisfiable, as the singleton set $\{Fa\}$. And some other sets containing Fa are satisfiable too, such as $\{Fa, Fb\}$ and $\{Fa, \exists x \neg Fx\}$. But some are not, including $\{Fa, \neg Fa\}$ and $\{Fa, \forall x \neg Fx\}$. Now, should we say that this means that the satisfiability of Fa in first-order logic is best thought of as a set-relative matter?

My own view is that the answer is clearly: *no*. The satisfiability of any one sentence is a matter of its "truth-conditions" on the disambiguation of that phrase where we think of these as a function from models to truth-values. If the function is a constant one with *false* as the only value, then the sentence is unsatisfiable, and if the function yields *true* for any model, the sentence is satisfiable. The truth-conditions of a set of sentences (or a conjunction of them) are determined by the truth-conditions (strictly: truth-and-denotation conditions) of its parts, and sometimes these can interact in interesting ways. The models that make Fa true might all make some other sentence (e.g. $\forall x \neg Fx$) false. But this does not mean that the contribution of Fa is "washed away" in considering its role in the status of the set $\{Fa, \forall x \neg Fx\}$. The truth-conditions of Fa are a critical part of what makes the set as a whole unsatisfiable. And

[19] This is a nice property for comparison because it requires the existence of a certain model, where s-shift–breakability requires the existence of two models (which are s-shifts of each other.) Perhaps an even closer analogue would be the property of being *logically synthetic*, i.e. true in at least one model and false in at least one model (equivalently, neither a logical truth nor unsatisfiable.) But this terminology is not as entrenched in logical practice, in part I suspect because of the hangover of logical positivism and its tendency to treat expressions like *synthetic*, *contingent* and *factual* as interchangable.

nor do we feel the need to say that the satisfiability of Fa is really something that is relative to a set of sentences: say, satisfiable relative to $\{Fa, Fb\}$, but not relative to $\{Fa, \forall x \neg Fx\}$.

I think the s-shift–breakability—Normativity—case is analogous, and I do not see any need to relativise to sets of sentences there either.

4 Final Thoughts

When I started work on this paper I had intended to respond to additional points—especially those on the characterisation of Universality in Humberstone's section II—but I see I have already spent more than 10 pages responding to about 5 pages of the review. Extrapolating: if I responded to everything at the same rate, that could result in a nearly 150 page response to a 73 page review. So I shall stop here before things get (even more) out of hand.[20,21]

[20] Humberstone concludes his review with some typos/errata he noted in his edition of the book, and these were extremely useful to me as I put together a list of corrections for the paperback. So while I am sure that *Daily Nous* readers could debate for days about whether this a suitable thing to put in a book review, in practice this is how the sausage gets made. It can be quite difficult to spot errors in one's own work, and also difficult to get critical, detail-oriented readers. (Or any readers!) This section improved the paperback edition of *Barriers* and, in the end, what more could an author want from a critic than to help them improve their work?

[21] Perhaps one last point: Humberstone complains that the relationship between some of my pieces of published work about barriers is unclear:

> "The 'further publications' alluded to in the opening paragraph above are [Russell (2011)], [Russell (2012)], and [Russell (2022)]. The first two of these deal with the already mentioned application to indexical sentences of the apparatus in play (taken up several times in *Barriers*, and the main focus of its Chapter 7), but unlike those, there is no reference made in *Barriers* to the third of them, which concentrates on the Hume's Law theme—the prominence of reference to which in the subtitle of *Barriers* would seem to afford it a special prominence. This is not an entirely satisfactory situation, since any reader following the development of Russell's thinking about these matters will wonder about how what is said there—especially since the content was delivered in several seminar and conference presentations around the world—is related to what is said in *Barriers*: is the content of [Russell (2022)] being endorsed here, being superseded, a bit of both, or what?"

These remarks caused Dave Chalmers to comment on a social media site: "wow! so is the content of [Russell (2022)] being endorsed here, being superseded, a bit of both, or what?"

The reason Russell (2022) is not cited in *Barriers* is that it was written after *Barriers* had been submitted to the press. Once I had finished the book it occurred to me that there might be readers who would be interested a shorter article that would explain the key insights, and then, further, that such readers might divide into two groups: those who would want the technical material in as concise and dense a format as possible and those who would want the informal view of Hume's Law without a lot of technical preamble. I thus planned to write two articles that would each serve as a kind of "advert" for *Barriers*: one focusing on formal material for a logic journal and one focusing on an informal presentation for a more generalist journal. The former became Russell (2022)—it is thus intended as a compressed version of the formal lessons from *Barriers*. The latter I am still working on.

References

Humberstone, I. L. (1996). A study in philosophical taxonomy. *Philosophical Studies*, 83:121–169.

Humberstone, L. (2024). A farewell to fragility: Critical notice of Gillian K. Russell, *Barriers to Entailment*. *The Australasian Journal of Logic*, 21(5).

Karmo, T. (1988). Some valid (but no sound) arguments trivially span the 'is'-'ought' gap. *Mind*, 97(386):252–257.

Pigden, C. (2025). Barriers to entailment: Hume's law and other limits to logical consequence. *Notre Dame Philosophical Reviews*.

Pigden, C. R. (1989). Logic and the autonomy of ethics. *The Australasian Journal of Philosophy*, 67:127–151.

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

Restall, G. and Russell, G. K. (2010). Barriers to Implication. In Pigden, C., editor, *Hume, Is and Ought: New Essays*, pages 243–259. Palgrave MacMillan.

Russell, G. K. (2011). Indexicals, context-sensitivity, and the failure of implication. *Synthese*, 183:143–160.

Russell, G. K. (2012). Lessons from the logic of demonstratives: what indexicality teaches us about logic and vice versa. In Restall, G. and Russell, G., editors, *New Waves in Philosophical Logic*. Palgrave Macmillan, Basingstoke.

Russell, G. K. (2022). How to prove Hume's law. *Journal of Philosophical Logic*, 51:603–632.

Russell, G. K. (2023). *Barriers to Entailment: Hume's Law and other limits on logical consequence*. Oxford University Press, Oxford.

Searle, C. J. (2024). Barriers to entailment by Gillian Russell. *Philosophy Now*, pages 52–53.

Shorter, J. M. (1961). Professor Prior on the autonomy of ethics. *Australasian Journal of Philosophy*, 39:286–287.

Weiss, Y. (2025). A relevant framework for barriers to entailment. *Journal of Applied Logics*, 12(5):1319–1347.