A Farewell to Fragility

Critical Notice of Gillian K. Russell, *Barriers to Entailment*

Lloyd Humberstone
Monash University

Or, to give the book its full title – *Barriers to Entailment: Hume's Law and other limits on logical consequence*, and publication details: Oxford University Press, Oxford 2023 (pp. *ix* + 303). This volume, to be referred to in what follows as *Barriers*, presents the current state of the author's comprehensive, fascinating, thoughtful and (no less importantly) thought-provoking exploration of inferential barrier principles such as Hume's *Is/Ought* Law. The line of investigation pursued was initiated thirteen years earlier in Restall and Russell [65] and Russell [71], and since developed in a succession of further publications. Numerous additional such principles are also investigated, including the articulation of barrier theses declaring the deductive underivability of universal conclusions from particular premises, of conclusions about the future from premises about the (present and) past, and of indexical conclusions from non-indexical premises, as well as – roughly speaking – modal conclusions from non-modal premises.[1]

The Hume's Law (or deontic) and (alethic) modal cases actually come in two versions apiece, depending, roughly speaking, on whether obligation and permissibility are lumped together or considered as relevantly different in the former case (in the main discussion of normativity, Chapter 8, and briefly as an isolated afterthought in §9.7, respectively), and likewise in the latter, with necessity and possibility (modal barrier 'version B' in §5.5 and 'version A' in §5.4, respectively). The discussions in which the operators involved are treated as significantly different are modelled after the discussion of another candidate barrier thesis – for the particular/universal barrier – discussed in Chapter 2, where, again roughly speaking, the purely existentially and universally quantified first-order sentences get a treatment on which the barrier separates the former as premises from having the latter as validly drawn conclusions.

Each of these various barrier issues has had its own distinctive flavour, and its own associations and repercussions in the history of philosophy. This makes *Barriers*, quite aside from its intrinsic interest and its many expository virtues, the perfect way of illustrating to a student whose *penchant* may be – not just for ethics, as in the *Is/Ought* case – but equally for metaphysics, epistemology or the philosophy of mind, how useful a formal approach to the conceptual fundamentals of their favoured territory can be. This makes it much more valuable, motivationally, than any dry, purely technical introduction to contemporary logical work could ever be. But students aside, any reader with logical interests will find *Barriers* to be bristling with imaginative curiosity, resourcefully pursued (which is not to say that the occasional query won't be raised on a point here or there as we proceed). The further

[1] Talk of barriers seems natural in this connection. The reviewer, in [37], made extensive use of what was there called the Barrier Lemma, without any particular claim to originality in that regard. (As the title of the paper suggests, only a preliminary exploration was on offer there; the title was also chosen as a nod in the direction of a well-loved Latin primer of yesteryear, Ritchie [67].)

candidate barrier theses just mentioned, we will come to later, sticking mostly, in Part I below, to the Humean application, though mentioning the particular–universal barrier in passing. Part II goes into more detail on the latter, in which the relevant case of what will be called an 'intermodel relation' is not symmetric. The case presents in this respect a characteristic taxonomic difference from that in focus for the Humean case, where the relevant intermodel relation is symmetric. Part III looks briefly at the modal/non-modal and constant/indexical barriers, largely in preparation for material postponed for discussion to Appendices A, B, and C. These address, respectively, one aspect of the finally favoured approach to barrier theses, the role of the aboutness relation in the book, and the availability of a streamlined modal-style treatment of the main results in the book. Some chapters of *Barriers* in the divisions of the book labelled *Getting Started* and *Getting Complex* – see the Contents listing below – get very scant attention, and those in the final part, *Getting Informal*, though highly original as well as extremely interesting, come in for no commentary here at all (to allow for a full tour of the 'formal' side of the story).

A listing of the various chapters may help with following the discussion; the book's own contents page is much more informative, including in addition the sections making up each chapter.

I

The 'further publications' alluded to in the opening paragraph above are Russell [72], [73], and [74]. 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 [74] being endorsed here, being superseded, a bit of both, or what? (We return to this briefly on several occasions below, especially in note 97.) The situation is different in respect of how the current proposals, whether in respect of Hume's Law or more generally, differ from those in Restall and Russell [65], about which *Barriers* is helpfully explicit.

To spell this difference out it is convenient to begin with the background supplied by Prior's argument in [61]; those familiar with this material are accordingly advised to skip to the paragraph after next. We have our hands, let us suppose, on a class of statements we think of as ethical (moral, or normative, etc.), of which E is one. (You could think of E as suggesting "evaluative" as much as "ethical", provided this is not taken to debar E from participating in logical relationships.) Suppose, further, that some other statement F is not one; as a shorthand "F" can be thought of as similarly suggesting Factual, but the important point here is just that factuality in this connection amounts to no more than non-ethicality. (In particular "factual" does not mean – as it might in everyday conversation be taken to mean – factually *correct*.) It is a twofold, or dichotomous, division that is under threat from arguments like Prior's and also at the heart of Hume's Law, not some three-block partition with a slot for 'mixed statements' to be parked in when they threaten the ethical/non-ethical barrier. The point is meant to be that you can't get deductively into the ethical realm from starting points all of of which lie outside of that realm – *anywhere* outside of it.

One might further think (very plausibly) that the negation of an arbitrary non-ethical statement is bound to be similarly non-ethical, though for Prior's argument we don't need such a strong assumption as this. It is enough that at least one non-ethical statement has a non-ethical negation, and that F has been chosen as such a statement, so that $\neg F$ is another such statement. Hume's Law can be taken as the claim that from no non-ethical premises can we validly infer an ethical conclusion. So, taking disjunctive syllogism as a valid form of inference, since E follows from $E \vee F$ and $\neg F$ by that inference form, and $\neg F$ is, *ex hypothesi*, non-ethical, Hume's Law could only be correct if $E \vee F$ counted as being ethical, as, otherwise, neither premise would be. But then that would itself be an even more immediate violation of Hume's Law, since by $\vee$-Introduction, this follows directly from the non-ethical F. So there is no way of classifying $E \vee F$ in terms of the binary division into *ethical* and *non-ethical* which is consistent with the correctness of Hume's Law, assuming that the two inferential steps – Disjunctive Syllogism and $\vee$-Introduction – involved are both accepted. Of course, each of these forms of inference has its detractors, and in each case considerations of what they are inclined to call *relevance* are brought to bear. In the $\vee$-Introduction case, there are various rather different motivations also in play, as with Parry-style "analytic implication" etc. – not only truth-conditional but also *conceptual* ampliativity taken as undermining validity – but for qualms on the score of relevance, see Weingartner and Schurz [88], and §2 of Schurz [79].[2] Since some robust notion of validity need not be taken as undermined on the latter approach, which can treat relevance as a logic-external filter on the inferences one is comfortable to endorse, whereas in the case of relevance in the Anderson–Belnap tradition, it is treated as internally constraining the logic(s), the Disjunctive Syllogism issue may encourage the reaction among adherents of this approach to think of this as a reply to reasoning like Prior's, apparently threatening Hume's Law, as further vindication of the approach. Relevantist celebrations on this score may be premature, however, as a variant of Prior's argument exploiting the full force of the idea that not only are *some* statement and its negation both non-ethical (the above F, $\neg F$) but that the appeal to the assumption that the class of non-ethical statements is

[2] Aspects of these considerations enter the discussion of Hume's Law in respect of the 'conclusion-irrelevant' appearance of ethical vocabulary in such inferences, in the discussion of what is called the General Hume thesis – or GH – in §3.3 of Schurz [78]. Russell has some remarks on similar 'analytic implication' or 'conceptivist' responses in note 18 on p. 39 of *Barriers*.

closed under negation enables us to sidestep the appeal to disjunctive syllogism.[3] Note that instead of formulating the stronger assumption about $\neg$ as the claim that the class of non-ethical statements is closed under negation, we can equivalently state it as the claim that the class of ethical statements is closed under negation.

Prior's argument above, as well as several involving further examples from Prior [61] and numerous other writers are presented as threats to Hume's Law in Chapter 1 of *Barriers* pp. 24–34, pp. 44–47, p. 49*f.*, and pp. 57–63, where similar putative counterexamples to the other barrier theses under consideration are also raised. Some serve to clarify the constraints of the plausible formulation of such theses, and others are raised only to be defused with relative ease. The discussion is interesting and useful, though quite a few of the examples – including several from Prior himself and Gideon Rosen's *flurg* case (*Barriers*, pp. 46*f.* and 278) involve conditionality, explicitly or otherwise, in a way that suggests a satisfying treatment would involve a discussion of the comparative merits of rendering "(All) Fs ought to G" as (i) $\mathcal{O}\forall x(Fx \to Gx)$, (ii) $\forall x(Fx \to \mathcal{O}Gx)$, or (iii) $\forall x\mathcal{O}(Gx/Fx)$, where I follow the font choice of *Barriers* with $\mathcal{O}$ for the deontic Box operator ("*it is Obligatory/it Ought to be the case* that") and $\mathcal{O}(\cdot\ /\ \cdot)$, sometimes written "$\mathcal{O}(\cdot \mid \cdot)$", for its dyadic companion (conditional obligation), not touched on in *Barriers*.[4] In any case, we jump now to Chapter 8 of *Barriers*, in which Russell addresses the Humean barrier thesis most explicitly, having to hand by then some model-theoretic apparatus crafted for addressing other such theses in the interim.

(Such jumping ahead is very much discouraged for a reader of *Barriers*, as a paragraph beginning on p. 9 makes plain, so Russell may well take issue with a reviewer doing precisely that. The paragraph in question begins:

> Some books can be dipped into in different places, but this one is designed to be read from beginning to end. A glance at the contents page shows there are chapters on individual barriers, but there is also an overarching narrative. If you jump ahead to the chapter on the *is/ought* barrier, you will miss some of the reasons that barrier takes the shape it does. (...) Similarly, if you were to stop after the chapter on the particular/universal barrier, the barrier won't yet be in its final form and several issues with it are unresolved. From the perspective of the narrative, that barrier can't be in its final form yet, since the reasons and perspective that shape it don't show up until we look at other barriers.

Russell continues with advice to skim through in order rather than read any single chapter in detail. The ellipsis indicated in the quotation above tells of the crucial role that seeing how to handle the indexical barrier – roughly: no context-dependent conclusions from context-independent premises – played in guiding her to a solution which can then be applied in the case of the other barriers, essentially, leading to an "unless-such-and-such" escape clause in formulating them. A second crucial development is that the concept of fragility,

[3] Suppose again that F and E are respectively non-ethical and ethical. Then $E \vee F$ is non-ethical, on pain of an immediate violation of Hume's Law, so by the closure condition under negation $\neg(E \vee F)$ is non-ethical, so if we are to avoid another violation of Hume here, since this has $\neg E$ as a consequence, $\neg E$ must be deemed non-ethical, and along with it, therefore, its own negation $\neg\neg E$, giving another, now inescapable violation of Hume, since this has E as a consequence. Moreover, all the inferential steps here are sanctioned by the system of first-degree entailment, the core of the Anderson–Belnap relevance tradition. (The present argument is a version of that depicted in Figure 3 of Humberstone [38], p. 135, where its connection to some considerations in Lewis [52] is noted, and further discussed in [43], pp. 1384–1388.) This is not to say that no role can be found for the deployment of relevant-logical resources in managing other aspects of philosophical taxonomy that may be found worrisome: see note 48.

[4] Famous earlier discussions include van Fraassen [84] and Lewis [51]; additional references are supplied in Humberstone [42], Example 4.4.4. Another issue complicating matters of logical form in deontic territory is the agent-relativity or agent-specificity of obligation and permissibility statements; for references, see [42], Example 4.4.14. Note, finally, that even in just the case of the first of cases (i)–(iii) above, there is an issue of the relative placement of the deontic operator and the universal quantifiers.

explained presently, is gently being retired from active duty in favour of something to be called breakability. This development is most explicitly foreshadowed in a section on p. 104 called "Beyond fragility" and marked with an asterisk to indicate its optionality for following "the thread of the central narrative," as it is put on p. 10 – though it can be quite helpful to know in advance where one is heading. Back, now, to 'jumping in.')

For simplicity, the Humean barrier thesis and proposed counterexamples and counterarguments to it are mainly treated in the setting of a propositional language with the usual truth-functional resources as well as deontic and (for good measure) alethic modal operators – $\mathcal{O}$ (with dual $\mathcal{P}$) and $\Box$ (with dual $\Diamond$), respectively, interpreted in models $\langle W, S, @, I\rangle$, where $W \supseteq S \neq \varnothing$, $@ \in W$, and I maps sentence letters to subsets of W. Where $M = \langle W, S, @, I\rangle$ and clarity is served by doing this, Russell refers to $W, S, @, I$ more explicitly as $W_M, S_M, @_M, I_M$, respectively, and writes $V_M(\phi, x) = 1$ [$= 0$] to indicate that ϕ is true [is false] at the element (or 'world') $x \in W$ in the model M – with $V_M(\phi, x)$ defined by induction on the construction of ϕ, beginning with ϕ a sentence letter, by setting $V_M(\phi) = I_M(\phi)$ and proceeding in the usual way for the various connectives available and uses $V_M(\phi)$ to denote $V_M(\phi, @_M)$. To reduce visual clutter, here we write $M, x \models \phi$ and $M \models \phi$, rather than $V_M(\phi, x) = 1$ and $V_M(\phi) = 1$, respectively. See the 'Aside' (on notation) below, for another notational deviation (in respect of "$\mathcal{R}$") from *Barriers*. In the "$\models$" notation just described, the inductive cases for $\mathcal{O}$ and $\Box$ would appear, for $M = \langle W, S, @, I\rangle$, as follows:[5]

$$M, x \models \mathcal{O}\phi \text{ iff for all } u \in S\colon\ M, u \models \phi; \qquad M, x \models \Box\phi \text{ iff for all } u \in W\colon\ M, u \models \phi.$$

In the terminology of Restall and Russell [65], given a relation $\mathcal{R}$ between models of the same type – for example, any two models of the type described above, one says that a sentence ϕ is $\mathcal{R}$-*preserved* when for any M if $M \models \phi$, then for all models N such that $M\mathcal{R}N$ we have $N \models \phi$, and that ϕ is $\mathcal{R}$-*fragile* when for any M if $M \models \phi$, then there exists a model N such that $M\mathcal{R}N$ and $N \models \phi$.

Aside on Notation and Variations. We denote the various binary *inter-model* relations collectively by using "$\mathcal{R}$" rather than "R" to set them off notationally from any binary intra-model relations in play, in, for example, §§5.2, 5.4; this means that while p. 118 of *Barriers* has "R" for the *inter*-model relations in play, four pages later it is playing the *intra*-model, picking out an accessibility relation. Since this is potentially quite confusing, it is best avoided with some such device as the $R/\mathcal{R}$ distinction. After this initial introduction of the terminology, the terms 'inter-modal' and 'intra-model' will from now on appear as single words. (In addition, for a suitable contrast we are using a slightly enlarged version of "$\mathcal{R}$" to increase readability when it occurs between what are already capital letters $M, N, \ldots$ for models.) The usual ('essentially binary') accessibility relations for the current deontic–alethic logic have been simplified down to 1-ary relations: in the deontic case holding between $x, y \in W_M$ just in case $y \in S_M$ and in the alethic case between any $x, y \in W_M$ whatever (whereas in Chapter 5 and the sections cited above, we are considering a greater range of monomodal logics than can be handled by means of such simplifications). Mnemonically, Russell encourages us to think of the elements of S_M as the (morally) Superb worlds of the model M, in which no obligations are violated. Very definitely to be put out of one's mind is the 'sanction' constant S of A. R. Anderson's reduction of deontic to alethic modal logic, true in exactly the worlds in which at least one obligation is violated.[6] Of course, in general, S_M can be any binary relation on W_M, typically assumed to be at least serial,

[5] For the cases of $\mathcal{P}$ and $\Diamond$, the "all" here becomes "some".

[6] See for example Anderson [1], and, for discussion and further references, Humberstone [42], pp. 253–258 and other pages cited in the index entry under 'Anderson, his reduction. . . ' there.

rather than the current 'monadically representable' simplification (see [42], p. 15 and elsewhere). In Restall and Russell [65] a capital S (mnemonically for the rather less enthusiastic "satisfactory") is used for such an accessibility relation – the pertinent intramodel relation – while the intermodel relation called s-shift in *Barriers* is notated as "$\between$," for which *Barriers* itself uses "$\circledcirc$" when the need is felt for a symbolic abbreviation. An interesting aside on p. 200 of *Barriers* remarks on the existence of evidence for the hypothesis suggested by remarks from Ladd(-Franklin) as well as Schröder – that the use of symmetrical symbols for symmetric relations, as in these choices in the present case (and all symmetric intermodal relations) reduces cognitive load. This is the practice followed in *Barriers*, with intermodel relations not guaranteed to be symmetric represented by symbols which are not left-right symmetrical.[7] The book also tells us where to find out more, providing a reference to Wege et al. [87].[8] In [65], p. 253, we read that the authors "assume that S is transitive, Euclidean, serial and secondarily reflexive" ("range-reflexive" in the terminology of [42]). The last property follows already from being Euclidean,[9] and collectively, the point-generated models possessing all these properties give us the same models as in the *Barriers* presentation, determining the same purely deontic (i.e., $\Box, \Diamond$-free) logic, namely KD45, sometimes called "deontic S5".

General notational convention: Whenever propositional (modal, etc.) logic is under discussion we take the sentence letters (or propositional variables) to be $p_1, p_2, \ldots$, occasionally writing the first three elements of this sequence as p, q, r. (This is a very slight variation on the notation in *Barriers*.) ***End of Aside.***

The particular $\mathcal{R}$ of current interest is the *s-shift* relation, holding between M and N when $W_M = W_N$, $@_M = @_N$, and $I_M = I_N$: in other words, the models M and N differ at most in respect of their S-components – which sets of worlds they deem to be "superb".[10]

[7] The LaTeX typesetting in *Barriers* is immaculate. (Yes, I absolutely refuse to write "LaTeX" at this point.) There is even, in the end-matter just before the bibliography, a list of the codes used for the special symbols used in the book, so that readers wanting to use them for themselves can easily do so. *Declaration of Interest* – or make that *Gratitude*: the reviewer was first introduced to LaTeX, and how to use it, by the author, some decades back. I would also like to acknowledge the solution provided by Shawn Standefer to a thorny typesetting problem that arose at one point in the course of writing this review.

[8] See note 2 on p. 200 of *Barriers* for the former and [87] for the latter. See also Pollard [59], pp. 21, 40, 54, 73, 97, and 110, for Schröder's interest in this iconic aspect of symbolic notation. As will be clear from [87], the hypothesis is naturally extended to apply not only to symmetric binary relations but also to commutative binary operations. In fact Wege and coauthors suggest that the visual symmetry of the conventional minus-sign may be implicated in the tendency of young learners to think of subtraction as commutative (as representing absolute difference, in fact), though the authors rather spoil their illustration of this apparently common classroom mistake in writing (p. 384): "These findings were consistent with Weaver's (1972), who noted that a belief in the commutativity of single-digit subtraction can lead children to assert, for example, that $78 - 45 = 33$ (since $7 - 4 = 3$ and $8 - 5 = 3$). The results from Study 1 suggest that the symmetry of the minus sign might contribute to these difficulties." Since $78 - 45$ *is indeed* 33, it may be hard to see what 'difficulties' Wege et al. are trying to explain and to help teachers forestall. Interested readers will find the answer, not so much in the 'Weaver (1972)' source they cite in the passage quoted, as in – another paper they cite – Mitchell [55], p. 136, top para. Note that several of those mentioned here are a bit more casual about observing the symmetric/symmetrical distinction, and even the symmetric/commutative distinction, than I have been here. Wege et al. [87] also go into the issue of horizontal vs. vertical symmetries.

[9] Or "euclidean", if one prefers that spelling; this is that property of binary relations R possessed when for any x, y, and z with Rxy and Rxz we have Ryz.

[10] Perhaps one might have preferred to see "S-shift" here, since it is S – or more explicitly S_M – that has been introduced as an official ingredient of the semantic metalanguage, and we can hardly regard "s" as abbreviating "S" after the manner of the other examples. It is true that in the indexical case, the context of a model is denoted by C and the corresponding $\mathcal{R}$-relation is called c-shift – but there, we can think of this as directly abbreviating "context shift". The policy in *Barriers* is to use lower case letters for all analogous cases – conveniently listed in Figure 9.1 on p. 201: d(omain)-extension, f(uture)-switching, m(odal)-extension, c(ontext)-shifting, s(uperb)-shifting. (Confusingly, although m-extension is the modal notion, d-extension, the converse of the usual substructure relation between first-order models, is sometimes – as in the title of §2.2 – called model-extension, but this can't be abbreviated to *m-extension* since that is the abbreviated form of *modal-extension*). The contrast here is with the more loosely analogous cases

The relevant result for Hume's Law in Restall and Russell ([65], p. 255) is given as Corollary 25, which appears below. Here I have modified their 2010 notation to match that of *Barriers*, with the earlier "A" replaced by "ϕ", "Σ" by "Γ", and "$\vdash$" by "$\models$", which, when it separates a set of sentences from a sentence, stands for the consequence relation determined by the class of models under consideration:[11]

CORO. 25 (2010): *If Γ is a satisfiable set of sentences, each of which is descriptive, and ϕ is normative, then $\Gamma \not\models \phi$.*

Here, $\models$ is just the consequence relation determined by the class of models under consideration, and *descriptive* and *normative* sentences have been defined as those which are, respectively, s-shift-preserved and s-shift-fragile.[12] As remarked in Humberstone ([43], p. 1423) the defining conditions for these properties – as for being $\mathcal{R}$-preserved and being $\mathcal{R}$-fragile generally – are respectively $\forall\forall$ and $\forall\exists$ conditions, so the distinction does not provide us with the dichotomous division needed for any suitable articulation of a Humean barrier thesis, as argued above: we need to know that no ethical conclusion follows from any (satisfiable, one may need to add, as above) set of non-ethical premises.[13] So any definitions offered for being normative and being descriptive need to make these notions each other's complements. Restall and Russell were aware of this in [65], with the crucial 'mixed' cases such as $E \vee F$ in the above presentation of Prior's argument falling into neither the descriptive nor the normative class, but they didn't seem to care. (The same goes for the variant $E \vee \neg F$, appearing as $D \to N$ in the quotation below, as well as for other Boolean compounds.) In *Barriers*, Russell has had a change of heart, for reasons consilient with those given above, writing on p. 6, in an overview of the book:[14]

> Vranas [86] pointed out that this sacrifices much of the intuitive strength of the barrier: surely Hume's Law is supposed to say that you cannot get a normative sentence (N) from a descriptive sentence (D) without some additional normative premise: and $D \to N$ is exactly the kind of additional normative premise which would fit the bill. An approach which classifies conditionals like this as neither normative nor descriptive has missed something that matters.

Thus, if we want $\mathcal{R}$-preservation and $\mathcal{R}$-fragility, at least for $\mathcal{R}$ as the s-shift relation with a Humean barrier thesis in mind, we should keep $\mathcal{R}$-preservation as it is (an $\forall\forall$ condition) and trade in $\mathcal{R}$-fragility for the corresponding $\exists\exists$ condition, or else stick with $\mathcal{R}$-fragility as it is and replace $\mathcal{R}$-preservation with the corresponding $\exists\forall$ condition. The latter has

arising in Part III of *Barriers* (Chapters 10 and 11), 'Getting Informal' in which barrier theses are discussed as they arise for natural language cases and a corresponding role is played by relations for which Russell reserves, instead, upper case letters in the abbreviated forms of their labels, DE (for 'domain extension', p. 250) and CS, FS, MDE and NS, explained on p. 255), which are relations between the informal versions of the models in the earlier discussion, called by Russell *Combinations* (with a capital "C").

[11] That is, $\Gamma \models \psi$ is defined to hold if and only if for every model M in the class, if $M \models \phi$ for each $\phi \in \Gamma$, then $M \models \psi$. This double use of "$\models$" is not to be found in *Barriers*, where, as already mentioned, Russell writes "$V_M(\phi) = 1$" rather than "$M \models \phi$" for truth in the model M.

[12] As note 10 mentioned, in [65] these properties go by the names "≬-preservation" and "≬-fragility"; "◎" replaces ≬ in *Barriers* for those occasions on which a concise symbolic rendering seems called for. Further, instead of *preservation*, *Barriers* speaks of *anti-fragility* and *unbreakability*, not quite realising, as we shall see in Part II, that the latter pair of terms pick out the same property as each other, for any given $\mathcal{R}$.

[13] We shall see later than one of the moves made in *Barriers* – and discussed below under the heading collective/distributive – to salvage a workable barrier thesis is to change this "set of non-ethical premises" into "non-ethical set of premises," with a collective rather than a distributive understanding of the latter phrase (though, as already noted, the actual terminology in *Barriers normative*/*descriptive* rather than *ethical*/*non-ethical*)

[14] I have adjusted the bibliographical reference to the style used in this review, and, correcting a presumed typo, interchanged the "(D)" and "(N)" as they actually appear in *Barriers*.

no plausibility as an explication of non-ethicality/descriptiveness, so Russell goes for the former option (as she did in fact in the bibliographically absent Russell [74]), calling the resulting ∃∃ condition $\mathcal{R}$-*breakability.* That's a nice terminological decision, since a fragile (e.g.) vase is far more prone to getting broken than a typical vase that's merely *capable of* being broken. Now we just need, for the sentence whose status is at issue, *some* verifying model from which we can get via $\mathcal{R}$ to a falsifying model: no need to be able to pass from *absolutely any* verifying model to a falsifying one. In another mnemonically helpful terminological choice pertinent for some choices of $\mathcal{R}$ other that the s-shift relation,[15] Russell considers the condition we get from breakability by interchanging "verifying" and "falsifying" here and calls a sentence meeting the resulting condition $\mathcal{R}$-*makeable*,[16] and of those which are either $\mathcal{R}$-breakable or $\mathcal{R}$-makeable as $\mathcal{R}$-*sensitive* – another well-chosen term. As *Barriers* notes, $\mathcal{R}$-breakability and $\mathcal{R}$-makeability coincide (with each other and therefore also with $\mathcal{R}$-sensitivity), when $\mathcal{R}$ is a symmetric relation, as of course the s-shift relation, being an equivalence relation, certainly is.[17] This repercussion of $\mathcal{R}$'s symmetry will receive further attention in Appendix C, in which a formal apparatus for displaying the meta-logical reasoning of many parts of *Barriers* is illustrated.

Let us pause to note some terminological decisions that have not been quite as felicitous as those reported so far. At the heart of these is the decision evident in *Barriers* to drop the 'preservation' terminology used in Restall and Russell [65], and indeed as recently as in Russell [74], where (p. 614) sentences are defined to be $\mathcal{R}$-preserved when they are not $\mathcal{R}$-breakable, and the analogue of Theorem 8.13 above – Theorem 2 on the page just cited – reads (essentially) like Theorem 8.13 except with "Γ is $\mathcal{R}$-preserved" in place of "Γ is not $\mathcal{R}$-breakable." One point in favour of retaining the talk of preservation is the connection it recalls with the *preservation theorems* which are the bread and butter (if not the life blood) of conventional mainstream model theory, with its various well-known preservation theorems tying $\mathcal{R}$-preservation for various relations $\mathcal{R}$ between first-order models to being equivalent to sentences of specified syntactic forms. In the following section, the Łos–Tarski Theorem, historically the first of such preservation results to have been proved, will be cited in this connection. The other advantage of retaining talk of being $\mathcal{R}$-preserved is that it might well have resulted in less terminological proliferation with the attendant danger of confusion, as will be also illustrated in Part II with the case in which it is indeed ordinary first-order models rather than the alethic–deontic models in play above. (The confusion in question arises not, as is more common, through conflation of categories, but rather through their over-differentiation, encouraged by the use of distinct but equivalent labels for them.)

For the moment, let us update what was headed "Coro. 25 (2010)" above with the *Barriers* version (p. 191), though this is worked up to progressively through the book in a heuristically (and didactically) useful intertwining of the discussion of the current barrier thesis with that of the others under discussion in earlier chapters – indexicality, universality etc. – rather than just being plucked out of the blue, as here, again arguably overruling the earlier-quoted remarks from *Barriers* about how the book is "designed to be read from beginning to end"; noting that now, instead of taking normativity as s-shift fragility (2010-style), it is being taken as s-shift breakability (and an *unless*-clause is added):

[15] See *inter alia* p. 11, p. 107ff., p. 176, and p. 204f. of *Barriers*.

[16] Think "make or break" (opportunity, decision, etc.), or perhaps the (chorus of the) Ronettes' 'The Best Part of Breaking Up'. *Barriers* also has the category of the $\mathcal{R}$-*constructible* to house those sentences for which, whenever false in a model are guaranteed to be true in some $\mathcal{R}$-related model; so these are sentences with $\mathcal{R}$-fragile negations, in the same way that $\mathcal{R}$-makeable sentences are those with $\mathcal{R}$-breakable negations.

[17] Thus an $\mathcal{R}$-sensitive sentence is one whose truth-value can change as we pass from some model to some $\mathcal{R}$-related model, and for the Hume's Law case, descriptive sentences are taken as those which are s-shift-*in*sensitive.

Thm. 8.14 (2023): *If* Γ *is descriptive, and* ϕ *normative, then* $\Gamma \not\models \phi$*, unless for all models* M *of* Γ*, every model* N *such that* N *is an s-shift of* M*, is such that* $N \models \phi$*.*

This result is a specialization to the Humean case of the 'Limited General Barrier Theorem', given on the preceding page of *Barriers* minimally adjusted to the present notational conventions, and retaining "not $\mathcal{R}$-breakable" rather than Russell's more commonly used "$\mathcal{R}$-unbreakable" for this (among other things – rather confusingly, it will be suggested in Part II):

(Limited General Barrier Theorem)

Thm. 8.13 (2023): *If* Γ *is not* $\mathcal{R}$*-breakable, and* ϕ *is* $\mathcal{R}$*-breakable, then* $\Gamma \not\models \phi$*, unless for all models* M of Γ*, all models* N *such that* $M\mathcal{R}N$ *are such that* $N \models \phi$*.*

The straightforward proof of Theorem 8.13 is not reproduced here.[18] Appendix C includes a handy formalism for representing such reasoning about preservation conditions, and will illustrate its interest in various ways, including a version of the above proof using the formalism involved (see Example 4 there), while Appendix A describes aspects of the treatment of this redundant element in *Barriers* itself.[19]

There are evidently a number of differences between the new Theorem 8.14 and the earlier Coro. 25, the most conspicuous being

(1) the replacement of fragility with breakability;

(2) the appearance of the *unless* clause (and associated disappearance of the satisfiability condition).

There is also a less immediately evident difference, explained more full after discussion of (1) and (2) below:

(3) the attribution of properties like breakability and fragility to sets of sentences is no longer construed *distributively*, but is now to be interpreted *collectively.*

Optional Aside on the General Barrier Theorem. Features (1) and (2) apply to the version of Coro. 25 with which Theorem 8.13 is actually contrasted in *Barriers* itself, namely Theorem 4.1 on p. 117, the *General Barrier Theorem*, which reads as follows, with the $\mathcal{R}$ notation above, and writing "preserved" rather than "anti-fragile":

If Γ *is satisfiable and* $\mathcal{R}$*-preserved but* $\Gamma \cup \{\phi\}$ *is* $\mathcal{R}$*-fragile, then* $\Gamma \not\models \phi$*.*

[18] About its special case, Theorem 8.14, *Barriers* includes remarks, plausibly enough, on p. 193: "Of course, the critic who calls Theorem 8.14 *trivial* might not have anything as specific as one of these three complaints in mind; they might just mean to say that it is too simple to be mathematically interesting to them. But here I want to suggest that the very accessibility of the theorem speaks in its favour as an interpretation of Hume's Law: Hume's Law was never an esoteric mathematical view understandable only by those with special training, but an intuitive informal principle graspable by and persuasive to those without advanced training in mathematics. It's not a criticism of its formulation in model-theoretic terms that it is relatively easy to grasp and prove once you understand those terms."

[19] Incidentally, what is called the 'Limited general barrier theorem' in §5.4 of Russell [75] is what in *Barriers* is called the 'Argument Set Formulation' of the Limited General Barrier Theorem (given as a Corollary to it on p. 207). As explained in the 'Optional Aside' below, I would prefer to avoid – or at least to downplay, as *Barriers* does by making this a Corollary rather than the main result – extending the taxonomic apparatus to argument-sets in the interests of fidelity to the original motivation for barrier theses. *Barriers* correctly points out that it is, however, easier to work with and to remember, so here it is: "Let ϕ be an $\mathcal{R}$-breakable sentence and Γ be an $\mathcal{R}$-unbreakable set of sentences. Then $\Gamma \not\models \phi$ unless $\Gamma \cup \{\phi\}$ is $\mathcal{R}$-unbreakable."

Barriers provides a proof of this, a somewhat formalized version of which appears as Example 3 in Appendix C below, but I would rather avoid any further discussion of it here because the collectivizing tendency mentioned under (3), which pertains to the attribution of the classifying properties descriptive/s-shift-preserved, etc., to the set Γ of premises takes us some distance from the original issue of barrier theses such as Hume's Law (as presented above), in which it is the sentences in Γ individually that have or lack these taxonomic properties. With this formulation we have gone even away from that starting point, by lifting their application to the whole union $\Gamma \cup \{\phi\}$ – *Barriers* calls this the *argument set*[20] – which throws the conclusion into the mix. In view of this, the next time this formulation will be encountered will be in Example 3 of Appendix C, as an illustration of the apparatus there in play. (There is, however, a closer approximation to the barrier thesis as originally conceived, the Corollary – labelled 'Weaker Barrier Theorem' – on p. 118: if Γ is satisfiable and $\mathcal{R}$-preserved and ϕ is $\mathcal{R}$-fragile, then $\Gamma \not\models \phi$.) ***End of Aside.***

(1) represents a general tendency as *Barriers* progresses for $\mathcal{R}$-breakability to edge out $\mathcal{R}$-fragility in characterizing what the book calls the *conclusion class* of a barrier principle: the class of sentences to which conclusions are barred from being drawn on the basis of certain sentences – those in the corresponding *premise class*. When, as here, the premise class is cashed out in terms of $\mathcal{R}$-unbreakability (alias $\mathcal{R}$-preservedness or $\mathcal{R}$-preservation, *inter alia*), replacing $\mathcal{R}$-fragility with $\mathcal{R}$-breakability makes the taxonomy a dichotomy, conveniently for the Hume's Law case and perhaps often useful, though sometimes making for exceptions which the *unless*-clause in (2) is there to handle.[21] Several aspects of the *unless* formulation are discussed in Part III below and in Appendix A. The demise of fragility in *Barriers* is a somewhat protracted affair, though, and Russell clearly retains a fondness for this (certainly interesting) notion, even after letting the reader in on its terminal diagnosis.[22]

The presence of the *unless*-clause, present also in the 'general case' Theorem 8.13, and occasioned by the move to the dichotomous approach, visible in these formulations with Thm. 8.14's descriptive vs. normative contrast now understood Thm. 8.13 style, as "not $\mathcal{R}$-breakable" vs. "$\mathcal{R}$-breakable" for the case in which $\mathcal{R}$ is the s-shift relation between deontic–alethic models as characterized before. Appendix A below goes into further aspects of (2) above. Here we turn our attention to (3), after noting the relation between the Descriptive/Normative taxonomy in its latest incarnation – as the s-shift unbreakable/s-shift unbreakable distinction – and a proposal alluded to in *Barriers* on p. 75, namely that of Karmo [50].

Karmo writes with the assumption that we are dealing with an interpreted language and a particular – perhaps *intended* – model which consists of a set of worlds thought of

[20] Anyone agreeing with the reason given in note 31 for putting a hyphen into "premise-set" may prefer to do so here also, for uniformity.

[21] This point about this leading to a dichotomy – a two-block partition – is stressed by Russell (apropos of several applications of the Limited General Barrier scheme) on p. 149 of *Barriers* at the end of the first new paragraph, and at a similar part of the page on p. 177, after the definition of *normative* on p. 186, and on p. 205*f.*.

[22] It is interesting to see that in something of an aside on p. 114, concerning temporally rigid terms such as (non-titular) proper names which "thus have a reference which is (as we might put it) anti-fragile over time (. . .) descriptions, meanwhile, often have a denotation that varies over time – they are fragile." The term *anti-fragile* turns out, as we shall see in Part II, to amount to the same as *not breakable*, and so picks out an $\forall\forall$ property so what would be needed for the analogue of the rigid/non-rigid distinction here would be an $\exists\exists$ rather than an $\forall\exists$ property. As it happens, in this particular case, if we think of $den(c, t)$ as a total function giving the denotation of the name c relative to time t, then the two conditions (i) $\exists t \exists t'\big(den(c, t) \neq den(c, t')\big)$, and (ii) $\forall t \exists t'\big(den(c, t) \neq den(c, t')\big)$ end up being equivalent. The point remains that the go-to terminological choice in *Barriers* at this point is *fragile* rather than *breakable*. (Note, incidentally, that if we wanted temporally directed versions of these concepts, for example replacing "$\exists t'$" by "$\exists t' > t$," then the resulting pair of conditions would not be similarly equivalent.) However, by the time we get to Chapter 10 – which in this review, we won't – the reaction is different and more appropriate, with rigidity likened instead, on p. 233, to unbreakability, leaving breakability as the analogue to rigidity.

as containing the actual world and other merely possible worlds, together and something called an ethical standard, which he suggests (note 6) can be thought of a subset of that set, comprising the morally perfect worlds by the lights of the ethical standard thereby represented. The intended model gives us not only a distinguished world (as actual) but also its own distinguished perfect subset (representing the *correct* ethical standard), though since reinterpreting non-logical vocabulary is not at issue, there is nothing corresponding to the V of the models in *Barriers*. Karmo writes ([50], p. 254):

> When people simply say, 'Sentence S is true', we take them to mean 'S is true in the actual world with respect to the correct ethical standard'. When people simply say, 'S is true in world w', we take them to mean 'S is true in w with respect to the correct ethical standard'.

We consider other models which do nothing but vary what that subset is to be. Ethicality is then defined in a world-relative way: a sentence is *non-ethical* in w iff changing the set of perfect worlds has no effect on whether S is true in w, otherwise it is *ethical* in w. In the more highly formalized setting of Chapter 8 of *Barriers*, models have a distinguished element @ built into their identity conditions and so we can think of this world-relativity as model-relativity, and Karmo's proposal becomes: a sentence is non-ethical (ethical) in M if it is s-shift-unbreakable (resp. s-shift breakable). And the reconsidered ('post-fragility') conception of the descriptive/normative distinction in *Barriers* is given by characterizing the descriptive sentences as those which are non-ethical in M for all models M, with the normative sentences being those for which this is not so. So the point of difference between Karmo and Russell is whether to precisify a Humean barrier principle at the model-relative level, or instead at the more global level in which all model-relativity is quantified away. At the stage at which Karmo is cited, *Barriers* is discussing a conditional $\neg D \to N$,[23] and we read (p. 75*f.*) that on the above account:

> the classification of mixed sentences is *world relative* – it depends on what other contingent facts obtain. If D is true, then $\neg D \to N$ be descriptive; if D is false, the same condition is normative. Independently of a world to fix the value of D, the conditional does not yet have a status. Relative to a world, the descriptive/normative taxonomy is a dichotomy.
>
> There is *something* important here, but I think we should resist views on which the classification of an unambiguous sentence can vary. 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. Or with the world; content determines the function from worlds to truth-values, making truth-values world relative – it isn't world-relative itself.

Fair enough. I have reformulated Karmo's approach in terms of model-relativity rather than world-relativity so as to relate it more clearly to that of *Barriers*, rather than to urge its all-round superiority.[24] By way of background, the world-relativity enters the story rather naturally on a reading of Shorter [80], who notes that Prior appealed, as represented in our initial formulation above, to the validity of the inference from F to $E \vee F$ and of that from $E \vee F$ and $\neg F$ to E, and Shorter characterizes such procedures variously as 'useless' and

[23] I would have preferred the simpler equivalent $D \vee N$ to avoid distractions about conditionals and to recall the $E \vee F$ used in presenting Prior's argument above (though with the order of the disjuncts the other way round).

[24] Perhaps the two should be thought of as competing accounts, for which reason the terminologies ethical/non-ethical and normative/descriptive were separated above. Humberstone [43] (e.g., p. 1423) concentrates on model-relative formulations. Note that since to obtain these, we instantiate the first $\forall$ to get to the model to which such formulations are relativized, the characteristic $\forall\exists$ and $\exists\exists$ difference between fragility and breakability is nullified; the former terminology – "fragile from M" – is used in Definition 3.5(ii) on the page of [43] just cited.

'futile', though we can take take this as suggestive of something a little less vague if we think of the purpose to which such inferences might naturally be thought of as aimed at doing: establishing their conclusions. An argument does not establish its conclusion simply in virtue of being valid. For that we need, at the very least, a *sound* argument.[25] But the two inferences just recalled can never both be sound, requiring F's truth for the soundness of the first and F's falsity for the soundness of the second.[26] Now, the soundness of an argument (with contingent premises, at least) is a highly world-relative matter. If we take Hume as protesting that there was no hope of establishing moral conclusions on the basis of non-moral premises, we immediately arrive at considering the soundness-based barrier principle in the title of Karmo [50], and with it the interest in world- (or model-)relative taxonomies.[27]

We turn now to the collective/distributive contrast mentioned under (3) above, and foreshadowed in note 13. On p. 70, the favoured explication (as of at that stage in the discussion) of Particular and Universal as $\Subset$-anti-fragility (alias $\Subset$-unbreakability) and $\Subset$-fragility respectively, where $\Subset$ is what would usually be called the *substructure* (or 'submodel') relation between models, a precise definition will be supplied in Part II.[28] The converse of this relation is the relation of *extension*, which matters only for the sake of grasping the general point illustrated by the following definition, which could equally well have been made with the deontic example above with s-shift anti-fragility s-shift fragility as marking off the Descriptive and the Normative, respectively. What the definitions do is *lift* such properties from applying to sentences to *sets of* sentences. This we may conveniently call the *collective* sense of such property attributions to the sets, as opposed to the *distributive* sense of these attributions, which just means that the property is attributed to each sentence in the set. The following definitions are taken verbatim (except for the omission of the word 'Definition') from p. 76 of *Barriers*, recalling that "anti-fragile" just amounts to "preserved". The particular/universal barrier is cited here only because of its use in *Barriers* to illustrate (what we are calling) the collective/distributive distinction; a more detailed discussion that barrier will appear in Part II below:

> (Particularity (sets)) A set of sentences Γ is *Particular* if, and only if, it is $\Subset$-anti-fragile, i.e., iff, if $V_M(\Gamma) = 1$, then for all extensions of M, N, $V_N(\Gamma) = 1$ too.
>
> (Universality (sets)) A set of sentences Γ is *Universal* if and only if it is $\Subset$-fragile, i.e., iff, for every model M such that $V_M(\Gamma) = 1$, there is an extension of M, N, $V_N(\Gamma) = 0$.

The notation we see in "$V_N(\Gamma) = 1$" has been explained as meaning that $V_N(\phi) = 1$ for each $\phi \in \Gamma$ (where "$V_N(\phi) = 1$" is what we have been writing instead as: $N \models \phi$). The notation "$V_N(\Gamma) = 0$", in the second definition, is not introduced explicitly in *Barriers*, leading one to wonder whether it is to be understood analogously, as meaning that $V_N(\phi) = 0$ for each $\phi \in \Gamma$, but the answer, which can be figured out from careful reading of examples

[25] And maybe even more, depending on how much epistemic *oomph* is read into "establish."

[26] Dialetheists will have to excuse what they would regard as an oversimplification here.

[27] Humberstone [37] was an earlier Shorter-inspired foray into this area, on which Karmo's note 7 ([50], p. 257) provides a useful comparative remark. "The present account, unlike Humberstone's, has the appealing feature that if it makes a sentence S ethical at a world w, then it makes the negation of S ethical at w also. On the other hand, Humberstone's account possesses, while the present one lacks, a different appealing feature: if a sentence S entails a sentence S′, and S′ is ethical in w, then so is S. (. . .)–Exercise: show that a theory having both features will make ethics nonautonomous, in the sense of admitting sound arguments from non-ethical premises to ethical conclusions." Diagnostically, the situation is reminiscent of the discussion in Lewis [52], which coincidentally appeared in the same year (1988), summed up on p. 8 there: "I claim that the Entailment and Compositional Principles are separately acceptable, but should not be mixed." The latter considerations do not, however, involve the world-relative feature of Karmo's account. They appear in Humberstone [38] in the discussion of kinds and co-kinds of statement.

[28] A frequent notation in mainstream model theory for this relation is just $\subseteq$ (with $\supseteq$ for its converse), but as *Barriers* observes (p. 200, n. 2), this is less than ideal.

(but which the reviewer confirmed, for good measure, by asking the author personally), is: *No* – instead "$V_N(\Gamma) = 0$" is intended to mean simply that $V_N(\Gamma) \neq 1$. The upshot of this is that the collective sense of these predications applied to (finite) sets of sentences amounts to the application of the original predicate of sentences to the *conjunction* of the sentences in the set concerned.

An illustration of this way of lifting sentence properties to set-of-sentences properties is immediately illustrated in *Barriers* with the example of a set of sentences, namely $\{Fa \rightarrow \forall xGx, Fa\}$ which is Universal even though neither of its elements is. Note then contrast between Restall and Russell's Coro. 25 (2010) above where the "each of which is descriptive" makes explicit the distributive interpretation of such phrases as "non-ethical set of premises" and Theorem 8.14 (2023), from *Barriers*, where the hypothesis that Γ is descriptive turns out to have instead this collective reading, which it retains in the eventually favoured treatment given by corresponding Limited Barrier Theorem (i.e., the Limited General Barrier Theorem (8.13) with $\mathcal{R}$ taken as the s-shift relation): even setting to one side the "Unless" escape clause, which comes in only with the eventual solution, things have taken a strange turn. Recall the 'disjunctive syllogism' part Prior's of argument (as summarized earlier) taking us from $E \vee F$, presumed non-ethical in order to save a Humean barrier from immediate failure since this follows from F, and the further premise $\neg F$, *ex hypothesi* also non-ethical, to the conclusion E dooming the barrier after all. We are now asked to attend to the status, not of the premises individually, but of the premise set $\{E \vee F, \neg F\}$ as a whole. Since this amounts, as noted, to considering the conjunction $(E \vee F) \wedge \neg F$, which is classically equivalent to $E \wedge \neg F$, and so by the later account with breakability replacing fragility is clearly ethical, as *Barriers* notes on p. 193, taking E, F as respectively Op, q, for definiteness and we have a conjunction whose which can go from true to false with a judiciously chosen s-shift, making it (in the terminology of *Barriers*) normative rather than descriptive.[29] In the 'All the Barriers' chapter (Ch. 9), the disjunctive syllogism step is represented more generically as having premise-set $\{D \vee N, \neg D\}$ – D(escriptive) vs. N(ormative) and qualifying as normative *qua* set, i.e., collectively, despite (its satisfiability and) the non-normative pedigree, at this stage in Prior's reasoning, of each of its elements.[30]

Barriers often speaks of 'responses' to or 'responding to' Prior's argument (such as that quoted in note 29), and it is worth asking on any such occasion whether the response being considered is some suggested *reaction* to the argument, conceding its success in achieving its goal, or, instead, a *reply* to the argument, querying its success. The goal in question to be that of showing that sentences cannot be partitioned into two sets, the ethical and the non-ethical, with some sentence and its negation lying in the latter set, in such a way that no consistent/satisfiable subset of the latter set has a member of the former ('ethical') set as a consequence. Querying the logic used (which permits *or*-introduction and disjunctive syllogism) is a response in the sense of a reply – though, as stressed above, querying the latter rule still permits a variant of the argument involving closure under negation (for each of the two of the hypothetical partition) to go through. But the move currently envisaged is more of a considered *reaction* to the argument: Yes, Prior was right about that – we can't

[29] For the earlier fragility-based treatment of these barriers, there is a comment in the Universality chapter of the book, on a universal/particular version of Prior's argument on p. 79. Beware of a typo appearing twice in this discussion: see the Errata list before the bibliography below. At that stage, it was possible to write: "We have thus responded to Prior's dilemma by classifying the mixed disjunction as Neither." That option is not available once breakability replaces fragility, of course.

[30] This means that when the premises are taken individually rather than collectively, the $\wedge$-introduction step from them to the conclusion is itself a violation of Hume's barrier thesis, as interpreted by Prior (and others). Blocking that way of construing the premises seems disconcertingly reminiscent of the suggestion that only *one*-premise inferences ever need to be considered in logic, since we can restrict attention to the case of a single premise conjoining what might originally have been several premises – to which the reply is (of course): Yes, but only because the now-invisible inference from the several premises to their conjunction is valid, when *not* taken to be of the form $\phi \therefore \phi$.

partition the set of sentences and have a barrier from the sets of non-normative premises to normative conclusions, but we might still save a Humean-style barrier if we partition instead the premise-sets in some other way that counting these as non-normative when all their members are, and perhaps suitable further adjustments elsewhere (the *unless*-clause, in the case of *Barriers*). Once this collective treatment of the premises is regarded as a generally available option, care may be needed against defining normative (or ethical) conclusions as those which are s-shift breakable and then, rather less informatively, premise-sets as descriptive (or factual) iff they have no such conclusions among their logical consequences.[31] (In drawing attention to this 'trivializing' move I am not saying that this is the move *Barriers* itself makes. If it were, there would be no need for the *unless* clause – though a somewhat similar qualm about the *unless* clause itself will be raised at the end of Part II below.)[32]

We have, however, been forewarned on p. 62, in Chapter 1 ('Survey of Counterexamples') of *Barriers*, that something like the collectivization strategy might be in the offing, with a well-chosen passage from Foot ([23], p. 507); I have rendered Foot's "f" as "F" for easier readability (though some might prefer "F-ness" here):

> We cannot possibly say that at least one of the premises must be evaluative if the conclusion is to be so; for there is nothing to tell us that whatever can truly be said of the conclusion of a deductive argument can truly be said of any one of the premises. It is not necessary that the evaluative element should "come in whole", so to speak. If F has to belong to the premises it can only be necessary that it should belong to the premises *together*, and it may be no easy matter to see whether a set of propositions has the property F.

Foot proceeds immediately to a discussion of what might be expected – wrongly, it turns out (see note 34) – to constitute an illustration of what she has in mind, involving what have since come to be called *thick* moral/evaluative concepts, and in particular, the concept *rude*.[33] Parts of *Barriers* not covered in the present review provide an extensive discussion of the vocabulary involved here, p. 44–50 and in Chapter 11, including Foot's *rudeness* example in the former selection, though not taking up this particular collective/distributive feature of the example.[34] The details of one's treatment of such thick moral concepts

[31] *Barriers* inserts no hyphen in "premise-set" which I have written that way simply to stress that this is *not* the concept of the "premise class" (for a candidate barrier thesis) used in *Barriers* in the manner explained above, as with "conclusion class."

[32] To repeat the preceding paragraph's considerations, for Prior's argument, in which the derivability of $E \vee F$ from F to settle its factual status, allows this disjunction to participate in the disjunctive syllogism inference with F's negation to the Hume-violating conclusion E, we do need to consider the disjunction's status *as an independent premise in the latter argument*. Forcing us to consider only the status of the premises collectively, and thus in effect of the single premise $E \wedge \neg F$ completely stymies the original reasoning, and changes the barrier thesis Prior was discussing.

[33] Foot's introduction of the issue ([23], p. 507) is quite carefully formulated, except that perhaps the final "condemnatory" is too strong given what has gone before, and is better replaced by "adversely evaluative": "I think it will be agreed that in the wide sense in which philosophers speak of evaluation, 'rude' is an evaluative word. At any rate it has the kind of characteristics upon which non-naturalists fasten: it expresses disapproval, is meant to be used when action is to be discouraged, implies that other things being equal the behaviour to which it is applied will be avoided by the speaker, and so on. For the purpose of this argument I shall ignore the cases in which it is admitted that there are reasons why something should be done in spite of, or even because of, the fact that it is rude. Clearly there are occasions when a little rudeness is in place, but this does not alter the fact that 'rude' is a condemnatory word." (There may be some confusion here, between the *reasons* one might give for regarding something as rude, and the *premises* in a deductively valid argument to the conclusion that the action in question is (or would be) rude in other than a purely descriptive sense – the former being, as is perhaps suggested by Foot's "other things being equal," better modelled under the rubric of nonmonotonic inference.)

[34] Indeed the example is presumably not, after all, intended to illustrate that feature: the reference in *Barriers* to it (p. 49) cites the one-premise argument with premise "Action X caused offence by indicating lack of respect," and conclusion "Action X was rude." The idea that someone, an ethical egoist, say, might not even regard the giving of offence as grounds for even a provisional negative evaluation, evidently seems

and the language embodying them are not really relevant to the collectivity issue, since Foot's claim that an argument from premises $P_1, \ldots, P_n$ to conclusion C can be valid and have each P_i non-normative while C is normative, entails that this is a possibility that obtains in the case in which C is the conjunction of the P_i, since the validity of arbitrary $P_1, \ldots, P_n \therefore C'$ marches – in the absence of more explicitly heterodox moves – in step with that of $P_1 \wedge \cdots \wedge P_n \therefore C'$, and the latter is a one-premise argument, for which the collective/distributive distinction marks no difference. So Foot's position would seem to require that whenever we have a case of normative C's following from (individually non-normative) $P_1, \ldots, P_n$ then the conjunction of the P_i is itself normative and follows from $P_1, \ldots, P_n$. She does not, however, give any concrete examples illustrating this possibility.[35] *Barriers*, on the other hand, does.

The kind of examples given at this point (p. 61*f.*) in *Barriers* are in the style, as is there noted along with similarly subsequently provided cases (see note 26 in the book), of the example labelled (1) in Karmo [50], p. 253, the idea behind which is there credited to a student of his:

> (1) 'Everything that Alfie says is true' looks like the sort of sentence one should call non-ethical; 'Alfie says that it ought to be the case that everyone is sincere' is non-ethical; but from these there follows the sentence 'It ought to be the case that everyone is sincere', which is ethical.

From the earlier discussion of Karmo, we already know the kind of thing he will say about this example, but let us hear him saying it:

> Is there any possible world in which the argument mentioned in example (1) is sound? Suppose that there is – suppose that there exists a world w such that, relative to what is in fact the correct ethical standard E, both of the premises in the argument from example (1) are true. E then is a standard which, whatever (if anything) it pronounces on other moral questions, at any rate prescribes universal sincerity. Take a standard E′ which does not prescribe universal sincerity. Relative to E′, the first premise of (1) ('Everything that Alfie says is true') is false in w. So if, in w, the premises of (1) are true, the first premise of (1) is ethical in w.

Now, *Barriers* was quoted above as expressing qualms about this kind of treatment, and urging us to "resist views on which the classification of an unambiguous sentence can vary. 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. Or with the world; content determines the function from worlds to truth-values, making truth-values world relative – it isn't world-relative itself." 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. (Indeed collectively, the resulting classification relative to the taxonomy in place of the premisses together need not match

as perverse to Foot as did the case (from her other famous paper of the same year) of a "moral eccentric," who "could argue to moral conclusions from quite idiosyncratic premises; he could say, for instance, that a man was a good man because he clasped and unclasped his hands, and never turned N.N.E. after turning S.S.W." ([24], p. 84). "Is it possible to extract from the meaning of words such as 'good' some element called 'evaluative meaning' which we can think of as externally related to its objects?", she asks incredulously (p. 85).

[35] If she had, she would have established Prior's conclusion in [61] two years before he did (see note 30 above, and the ambient discussion): that there is no two-block normative/non-normative partition of the set of sentences with no member of the former block following from any set each of which is drawn from the latter. (With $n = 2$, $P_1 = E \vee F$, $P_2 = \neg F$ and their conjunction conclusion C equivalent to $E \wedge \neg F$ as in note 32, perhaps.)

that of any of them taken individually.) 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. 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. 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 $\varnothing$) it has such-and-such possibly different status on similar grounds (relative to which P_1''s status need not coincide).[36]

The move from the distributive to the collective classification of the premises is made on the page of *Barriers* immediately after the page (p. 75) on which Karmo-style relativizations are put to one side, as not doing justice to the idea that classification should be on the basis of content alone, and 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[37]) to allow for an appeal to truth at which (resp. in which) to play a potentially decisive role in the classification of sentences. *Barriers* makes a similar point in connection with the (at that stage fragility-based) account, when considering (p. 75) the particular–universal taxonomy and something there called Vranas' problem – though this seems to differ mainly notationally from the problem raised by Prior's argument from our opening presentation of that argument (with F, E, and $E \vee F$ replaced respectively by $\neg Fa$, $\forall x Gx$ and $Fa \to \forall x Gx$):

> It almost seems tempting to say that the status of a mixed conditional varies with the company it keeps. When it hangs around with Fa it "acts" universal. But when it spends time with $\neg Fa$ it "acts" particular. Similarly for mixed disjunctions.

Here 'hanging around with' covers being in the company of other premises (since $Fa \to \forall x Gx$ with the additional Particular-rather-than-Universal premise Fa yields the conclusion $\forall x Gx$), as well as being a subsentence of conclusion (as in the case of the 'mixed conditional' $Fa \to \forall x Gx$ when taken as a conclusion following from (the equally Particular, with a capital 'P') $\neg Fa$. As we have seen, however, the first aspect of this 'varying with the company it keeps' idea is eventually endorsed, in effect, by the shift from individual to

[36] How should we think of content in the present setting, shared by Karmo's discussion and *Barriers*, with these models in which truth is determined by two independent features, the distinguished world of the model and the model's distinguished subset (of superb worlds, setting the 'ethical standard')? Since the latter is not being treated as a contextual parameter, at least at the hands of Karmo and Russell, the account that comes most readily to mind for the content of a disambiguated declarative sentence is as a set of pairs $\langle w, S\rangle$ (to use the *Barriers* notation), called "G-worlds" or "Gibbard-worlds" in Humberstone [38], p. 153, [43] p. 1439, for reasons made clear there; further discussion of the issue of content in this setting is provided in Bedke [5].

[37] A pointed model is a Kripke model with a distinguished element/world, as in Kripke's own original semantics for modal logics, though the phrase is not as useful as it once was, since the emergence of different roles the distinguished point might be playing. In Chapter 6 of *Barriers* we shall see (in Part III below) see models in which there is a distinguished world as part of the context-representing apparatus as well as a distinguished world as a 'circumstantial parameter' the latter shunted around in the course of unpacking the truth conditions every time a modal operator is encountered, the former remaining imperturbably fixed. Crossley and Humberstone [14] allowed something playing the former role to be shunted around but called the operators effecting this global operators rather than modal operators, because they were given what in Humberstone [40] called model-changing semantics, though Section 4 of [40] points out that a certain arbitrariness arises here over what exactly to build into the notion of a model. Relative to the way the notions are developed in Appendix C below, the operator $\blacksquare$ on stage there is very much a 'global' operator. See also note 100 below.

collective formulations of a barrier principle, the company in question being constituted by the relevant co-premises. 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.

II

The particular/universal contrast was encountered briefly in Part I for the sake of a quotation from p. 79 of *Barriers* introducing the favoured collective as opposed to distributive understanding of set of sentences w.r.t. these properties, with reference to simple first-order languages and the relation $\Subset$ between them. Unlike the s-shift relation, this is not an equivalence relation, in particular, not being symmetric, which gives the formal side of the situation a rather different flavour. One would normally read "$M \Subset N$" as either "M is a substructure of N"[38] or "N is an extension of M," though *Barriers* favours specifically a variation on the latter way of speaking: model-extension or d-extension. (See note 10.) But we need to begin by saying what the models under discussion are. They are introduced in Chapter 2 ('Universality'), p. 67, of *Barriers* as pairs $\langle D, I\rangle$ with D a non-empty set (the domain) and I, an interpretation function, mapping individual constants to elements of the domain, and n-place predicates to subsets of D^n. (For simplicity, n-ary function symbols for $n \geq 1$ are not present in the first order languages under consideration.) Then we also have, not considered as part of the model itself, assignments g of elements of D to the individual variables, such assignments playing their usual role in the definition of an assignment's satisfying ϕ relative to a model M, with truth in the model (writing here $M \models \phi$ for this, in place of the $V_M(\phi) = 1$ of *Barriers*), defined as being satisfied by all assignments relative to M.[39] As above, the domain and interpretation of a particular M will be referred to as D_M and I_M. The definition of model-extension, symbolized by "$\Subset$", is given on p. 69 by setting $M \Subset N$ if and only if:

(1) $D_M \subseteq D_N$;

(2) for all individual constants α, $I_N(\alpha) = I_M(\alpha)$;

(3) for all n-place predicate letters Π and all $d_i \in D_M$ (with $1 \leq i \leq n$):

$$\langle d_1, \ldots, d_n\rangle \in I_N(\Pi) \text{ iff } \langle d_1, \ldots, d_n\rangle \in I_M(\Pi).$$

In other words, you can extend one model to another by adding new individuals to the domain while not losing any of the individuals you started with or changing the interpretations of constants (1, 2), or, insofar as they concern relations among individuals already in the original model's domain, the interpretations of predicate letters (3). At this stage in

[38] Alternatively: "M can be isomorphically embedded in N"

[39] Since the consequence relation determined by a class of models is given by truth-preservation from left to right, with truth being truth in a model rather than truth in a model relative to assignment, this means that the consequence relation determined by such a class will typically lack the unrestricted 'Deduction Theorem' property – that $\Gamma, \phi \models \psi$ implies $\Gamma \models \phi \to \psi$. (For example with $\Gamma = \varnothing$, we will have $Fx \models \forall y(Fy)$ while $\not\models Fx \to \forall y(Fy)$.)

the book, fragility is still in play as well as breakability, and the $\Subset$-specific incarnations of these notions are defined on p. 70 in essentially the following ways:[40]

- A (first-order) sentence ϕ is $\Subset$-*fragile* iff for every model M, if $M \models \phi$, then $N \not\models \phi$ for some model N such that $M \Subset N$.
- A sentence ϕ is $\Subset$-*breakable* iff there are models M, N such that $M \Subset N$, $M \models \phi$ and $N \not\models \phi$.

$\Subset$-anti-fragility and $\Subset$-unbreakability are defined on the following page, essentially thus:

- A sentence ϕ is $\Subset$-*anti-fragile* iff every model M such that $M \models \phi$ is such that for all models N, if $M \Subset N$, then $N \models \phi$.
- A sentence ϕ is $\Subset$-*unbreakable* iff it is not $\Subset$-breakable, i.e., if there is no pair of models M, N, such that $M \models \phi$, $M \Subset N$, and $N \not\models \phi$.

This last pair of definitions present something of a problem for the reader, because of the implicature the use of different terms as *definienda* (on the left of the "iff") and differently formulated *definientia* (on the right) has: that distinct notions are under discussion here. That was certainly the case with the clearly non-equivalent notions of fragility and breakability on the preceding page. Here, though, the difference in formulation does not make for a difference in respect of the condition formulated: in either case, it is just a matter of the truth of ϕ being guaranteed to be preserved on passing from any model to any extension of that model (i.e., to any model to which the first model bears the relation $\Subset$). This makes the passage at the top of p. 72 of *Barriers* quite misleading; the reference to "the intuitive idea outlined above" is to the idea that (p. 70) "Universal sentences are sensitive to $\Subset$. Particular ones are not.":

> The question then, is how we should precisify the intuitive idea outlined above. Should we say that to be a Universal sentence is to be $\Subset$-fragile, while to be a Particular sentence is to be $\Subset$-anti-fragile? Or should we instead say that to be a Universal sentence is to be $\Subset$-breakable, while to be a Particular sentence is to be $\Subset$-unbreakable?

The misleading suggestion is that in addition to replacing one candidate explication of Universality with another, we have made a suitably corresponding replacement of the one correlative explication of Particularity by emending it to match – whereas in fact, there has been no change on the latter front (other than in respect of terminology).[41] What we have is, rather, a choice as to which notion of Universality – that in terms of fragility or that in terms of breakability – to set against Particularity construed as $\Subset$-unbreakability (alias $\Subset$-anti-fragility, alias $\Subset$-preserved status). The most conspicuous aspect of choosing the breakability option, of course, is that we end up with a dichotomous taxonomy for the barrier principle in question.

In fact, *Barriers* notes (p. 74 Remark 2.7) that what we have in fragility-based account is "not just a trichotomy but a tetrachotomy." This occurs in the course of fulfilling a promise made in the paragraph immediately following the last inset quotation above (about the two ways of explicating the Universal/Particular contrast):

[40] In using the "$M \models \phi$" notation in place of Russell's, except that in the definition immediately following I have transcribed the "$V_N(\phi) = 1$" appearing at this point as "$N \not\models \phi$," taking it to be a typo for Russell's "$V_N(\phi) = 0$." The typo is unlikely to confuse readers, because the informal gloss given after the definitions on p. 70 makes it clear what was intended.

[41] Similarly in the end-of-chapter summary, p. 83: "We saw that this informal idea [of sensitivity, on which terminology, see further note 47] can be precisified in two ways: as the $\Subset$-anti-fragile/fragile distinction, or as the $\Subset$-unbreakable/breakable distinction." And later still, with the book laid open at pp. 106 and 107, we have $\curlyvee$-anti-fragility on the left-hand page and $\curlyvee$-unbreakability on the right, about a fifth of the way down, in each case. ($\curlyvee$ is the *future switch* relation: see note 50).

> It would help to know more about the consequences of each option. Two kinds of consequence are of particular interest. First, what does the resulting taxonomy look like? Is it intuitive, sensible, or defensible? And second, does it permit us to prove a Particular/Universal barrier? §2.4 examines the first question for the anti-fragility/fragility taxonomy, while §2.5 answers the second.

What the interested reader may be wondering at this point is: yes, that's all very well, but these questions both concern the fragility-based option – what has happened to the other option? Any readers of *Barriers* already familiar with Russell [74] will already know that *fragility* is well and truly on the way out, while others will have to be patient and wonder how things will end. But perhaps it would have been good to mention at this point something said more than once elsewhere: that a resolution will in due course emerge (an application of the Limited General Barrier Theorem) from a suggestion in Chapter 7 (on context-sensitivity, originally in Russell's 2011 paper [72]), then applied in Chapter 8 (to the subject – Normativity – of that chapter, and then more generally including to the Universal/Particular case in Chapter 9 ("All the Barriers"). This would be better than leaving them wondering why one of two options distinguished has suddenly been left out of the discussion in favour of the other (fragility-based) option.

It is hard to believe that the resulting proliferation of equivalent terms would have occurred if what had been defined in either of these equivalent ways had been the obvious – though pointedly shunned – choice: "⋐-preserved". Using that terminology, the initial and provisional proposal of Chapter 2 of *Barriers* is that defended in Restall and Russell [65], where the talk was of preservation, as indeed it is in Russell [74]:[42] the particular sentences are those that are ⋐-preserved, while the universal sentences are those that are ⋐-fragile. As with the other barrier principles, all initial formulations are subject to renegotiation as we proceed through the different cases and allow points noticed along the way in one area to inform what we end up saying in others (as summarized in Chapter 9, "All the Barriers," – though that still leaves the last two chapters of the book, not covered here, to address the question of what to make of barrier theses in informal settings for which the model-theoretic apparatus is not obviously on hand to assist with formulations and proofs). But about this particular suggestion, *Barriers* notes immediately that we do not have a dichotomy (a two-block partition) on our hands here, not that this seems like the worry it is in the Humean case discussed above.

When it *does* turn into a dichotomy in the post-fragility Limited General Barrier Theorem applications in Chapter 9 (pp. 204, 206), the fact that it is only the *conclusion* class – where the Universal sentences lie – that has been changed in the process, with the replacement of ⋐-fragility by ⋐-breakability (along with the *unless*-clause), and not also the premise class – housing what were formerly the Particular sentences, and are now just those that are not Universal (not ⋐-breakable) – remains the same. The old characterization (line 2 of §2.4) of the premise class for this Particular-to-Universal barrier was defined as those that were ⋐-anti-fragile, a notion which had been given its separate definition, as we saw, from that provided for ⋐-unbreakability. So it could easily escape the casual reader's

[42] Perhaps Russell is now avoiding this to forestall any possibility of a confusion between of talk of truth-preservation, not in the transition from one model to others, but in the passage, relative to an arbitrary single model, from premises to conclusion (for an argument to be valid). (Incidentally, on p. 161 of *Barriers* she warns us, with examples such as "Actually, it is raining. ∴ It is raining" in mind that "valid arguments can have premises which are true in worlds where their conclusions are not—so that it is *possible* for the possible for the propositions expressed by the premises to be true while the one expressed by the conclusion is false," adding in note 18: "This gives us a further reason to be cautious about saying 'necessary truth preservation' when we are characterizing entailment." A sophisticated point – though I would have preferred a disambiguating hyphen here: 'necessary truth-preservation' since with "Actually" prefixing the premise rather than the conclusion, the example also, at least arguably, illustrates a failure to preserve *necessary truth* and so the gloss could be interpreted with the hyphen after "necessary" instead.)

attention – as it seems to have escaped the author's[43] – in this welter of terminology, that whether they're called "not $\Subset$-breakable", or "$\Subset$-unbreakable", or "$\Subset$-anti-fragile", this is the same premise class all along: the $\Subset$-preserved sentences.[44]

Let us turn our attention to the background prompting the provisional decision as to what universality and particularity are to consist in. On p. 2 of *Barriers*, Russell writes that, by contrast with the Humean barrier thesis, the remaining such theses to be discussed are "so uncontroversial as to be platitudes" and begins, as Restall and Russell [65] began, with the other Russell, in *The Philosophy of Logical Atomism* (Russell [70] and Slater [81] being the editions I consulted) with

> You can never arrive at a general proposition by inference from particular propositions alone. You will always have to have at least one general proposition in your premises.[45]

A more detailed examination of matters, skippable by the reader not feeling the need for one, follows.

Interlude: Further Historico-Philosophical Background. A few lines before this, Russell writes ([81], p. 206, [70], p. 69):

> Of course, it is clear that we have general *propositions*, in the same sense in which we have atomic propositions. For the moment I will include existence propositions with general propositions. We have such propositions as "All men are mortal" and "Some men are Greeks". But you have not only such *propositions*; you have also such *facts*, and that, of course, is where you get back to the inventory of the world: that, in addition to particular facts, which I have been talking about in previous lectures, there are also general facts and existence-facts, that is to say, there are not merely propositions of that sort but also facts of that sort. That is rather an important point to realize. You cannot ever arrive at a general fact by inference from particular facts, however numerous.

It is because of this first airing of the barrier thesis under consideration that when it is raised for a second time, the formulation of the remark quoted in [65] and in *Barriers*, as corrected in note 45, has what would otherwise seem to be a strange syntax – "You never can" rather than "you can never" – making better sense once we see that the issue has already been given a preliminary airing.

But the earlier passage is confusing, all the same. We have the announcement: "For the moment I will include existence propositions with general propositions." Just how long is this moment supposed to last? Doesn't the subsumption of the existential case apply to the case of propositions, and if it does, why the "and" in "there are also general facts and existence-facts", only a few lines after the announcement? The formulations in [65] and *Barriers* with "general" replaced by "universal" avoid this confusing subsumption, which is obviously fatal to the barrier thesis that (Bertrand) Russell is trying to endorse, since from the particular premise that a is F we can immediately derive the (is-it-or-isn't-it?) general

[43] This is, indeed, the only point in *Barriers* at which the reviewer felt that Russell was not in full control of the material.

[44] Here, to keep a clear focus on the present issue, this summary simplifies away the fact that the various labels apply to the sentences in the premise class collectively, i.e., as corresponding properties of the premise-sets.

[45] In *Barriers* this is sourced to p. 101 of an edition of the work in question. The same passage quotation (verbatim) in Restall and Russell [65] is sourced to p. 206 of a different edition. Armed with two (different again) digital versions of the work, I searched on the phrase "never arrive" to locate the passage and check out its surrounds, but found nothing. (One of these, Slater [81], also has the corresponding passage on its own p. 206, so I expect the text is also the same.) the point of mentioning this is to save anyone else from doing the same thing. In both of the versions I had access to, the relevant passage actually starts with "You never can arrive at" rather than "You can never arrive at" – the subtle stylistic/connotational difference between which will be touched on in a moment.

conclusion that there exists an F. The whole language of *particular* vs. *universal* makes this reviewer somewhat uncomfortable, threatening to collapse a threefold distinction the threefold contrast suggested by $\forall x(Fx)$—Fa—$\exists x(Fx)$, with F a predicate letter here (rather than a schematic letter for a possibly complex predicate), into a twofold distinction, lumping the existential case and the (shall we say?) *specific* case together under the word *particular*. To avoid lopsidedness, one should also consider the matching initial binarization we get by linking the specific with the universal and contrasting that with the existential.[46] (A hunch: the idea that Particular and Universal might make a usefully contrasting pair in a taxonomy of sentences may be a vestige of thinking of the ontological contrast between universals and particulars. But of course these two contrast have very little direct connection: *Some Fs are Gs* can be thought of as a claim about two universals, if one insists on speaking that way, to the effect that they are coinstantiated, but in the current terminology it is not a Universal, but a Particular – because existential – claim.) In the full interests of even-handedness, we can also contrast the specific with the combined existential and universal – which is what the passage from *The Philosophy of Logical Atomism* above seemed to do (rather half-heartedly) with the category of the *general*. With such potential qualms in mind, consider the following passage from p. 68 of *Barriers* recalling the suggested identification of particularity and universality with $\Subset$-preservation and $\Subset$-fragility:

> Here is the informal thought behind the formal definitions of Particular and Universal. Particular sentences are, in a sense, local. They refer to an object and say something about it, and in order for the sentence to be true, the object referred to needs to be a certain way. Particular sentences don't tell you how the whole world has to be, just a particular part of it. Because of this, changing other parts of the world – and especially adding new objects – won't change a particular sentence from true to false.

Fa (or Rab, $Scac$, etc.) involves reference, relative to a given model, to an object (or objects) which they say something about, and their truth is preserved to extensions of that model, and indeed, to the extent that elements of the domain are thought of as parts of the world (as represented by the model), are naturally thought of as "in a sense, local". But $\exists x(Fx)$ is also preserved from a model to any extension of it, since any witnesses it has will survive the addition of new elements. And no object is referred to in such a sentence, nor does any object have to be a certain way for it to be true – and when we say no *particular* object has to be a certain way, to emphasize the intended scope relations, it's not as though there are some other-than-particular objects lying around in the domain. This sentence precisely *does* tell you how the whole world, rather than any (proper) part of it, has to be for it to be true: the world has to contain, just somewhere or other, something that is F. ***End of Interlude.***

§2.3 begins at p. 70 with a summary of the gist of some examples. Figure 2.1, referred to in the following passage, depicts a model in which everything in the domain, including what a denotes according to that model is in the extension of F, together with a model extending it in which a new object is added which is not in the extension of F, the pair of models being captioned with the observation that true particular claims – here Fa – stay true when the model is extended. Figure 2.2 features the same pair of models with the caption that

[46] The phrase "initial binarization" here is intended to refer to the classification before it is extended to cover Boolean compounding of cases classified, rather than to pertain to what in Part I was called the dichotomous nature of the eventual taxonomy after such an extension. The issue being raised here over neglect of what I'm calling the *specific*, inferentially intermediate between the universal and the existential, is analogous to the neglect of the *present* in the discussion of tense-logical barriers in Chapter 3 of *Barriers*, where the principal taxa are Past and Future (past-oriented and future-oriented sentences) and the – this time, *chronologically* – intermediate case of the Present does not get to share top billing. This seemed particularly surprising in view of the presentism–actualism debate and the special attention devoted to the taxon of the Actual in Chapter 5 (on modality).

true universal claims – here $\forall xFx$ – can be made false by extending the model. Footnotes appended to the quotations offer commentary supplied here rather than reflecting notes in the source:

> The intuitive idea above was that we should use sensitivity to model-extension to define what it is for sentences to be *Particular* or *Universal*: Universal sentences are sensitive to $\Subset$, Particular ones are not. But when we try to make this idea more precise, we find that there are two promising ways to do it.[47]
>
> On one, a sentence is Universal just in case it is model-extension *breakable*, where this means that extending a model of the sentence can make the sentence go false. Fig. 2.2 showed that $\forall xFx$ was $\Subset$-breakable. Fa, by contrast is not $\Subset$-breakable, because every extension of every model of Fa makes Fa true.
>
> On the other precisification of *sensitive*, a sentence is Universal just in case it is *fragile* with respect to model-extension, where this means that every model of the sentence has an extension that makes it false. $\forall xFx$ is $\Subset$-fragile, in addition to being $\Subset$-breakable: whenever $\forall xFx$ is true in a model, you can always make it false by extending the model (e.g. by adding a single element which is "$\neg F$"). We saw before that Fa was not $\Subset$-breakable and it is not $\Subset$-fragile either. At least one model of Fa (e.g. the one in Fig. 2.1) cannot be extended to a model that makes Fa false.

Barriers, p. 70, goes on to mention a (satisfiable) sentence for which "breakability and fragility come apart. One example is the Prior-disjunction $Fa \vee \forall xGx$," where the allusion to Prior recalls the 'mixed disjunction' cases touched on in Part I above. The discussion continues on p. 73*f.*:

> Our final two remarks in this section highlight less encouraging consequences of the fragility-based definitions. We will tolerate them for now, though eventually a later solution to a different problem will take care of these worries too.
>
> **Remark 2.6.** *All logical truths are Particular.*
>
> Logical truths are Particular no matter how many universal quantifiers they contain, nor how intuitively universal they seem. For example,
>
> $$Fa \vee \neg Fa \qquad \forall x(Fx \vee \neg Fx) \qquad \forall x(x = x)$$
>
> are all Particular sentences. The reason is that logical truths are true in all models, and so they fit the anti-fragility condition on Particularity: whenever one is true in any model, it is also true in all that model's extensions.
>
> This is an awkward result. It is counterintuitive to call some of the logical truths above Particular; they are intuitively universal. But it is also hard to see how the sentences could be better classified on our present approach. For first, logically equivalent sentences must be classified in the same ways: $Fa \vee \neg Fa$ cannot be classified differently from $\forall x(Fx \vee \neg Fx)$. And second, changing the classification of the latter to *Universal* would be incompatible with the barrier: no true barrier theorem claims that a logical truth *doesn't* follow from Particular premises – logical truths follow from everything.[48]

Despite the reference, early in this passage, to the less encouraging consequences of the fragility-based definitions, when Universality is eventually recharacterized (p. 204) as $\Subset$-breakability instead of $\Subset$-fragility, this makes no difference to the discouraging classifications

[47] Note that neither of these corresponds to the notion of $\mathcal{R}$-sensitivity extracted in Part I above from *Barriers*, pp. 107, 110, a discussion which comes later than the present excerpt. It would have been better to avoid this terminology and choose something quite different from *sensitive* for the informal notion amenable to the current alternative precisifications – for example, *vulnerable*, a term used in this kind of way in the next but one passage quoted from *Barriers* below.

[48] If one is worried, for the sake of barrier theses, about the fact that, classically, the logical truths follow from every sentence or the fact that every sentence follows from their negations, then this is one place where relevant logic *will* certainly help, as numerous results in Weiss [89] show.

mentioned in the above passage. There is a contrast here with the mark of universality supplied by Łoś–Tarski, to which we return below. That result, in what is sometimes called its dual form (concentrating on the existential rather than the universal) is also relevant to the following passage, from an unnumbered subsection of §3.10 (p. 112) in Chapter 3 on tense, headed "Refinements elsewhere" consisting of the following words and a table (the footnotes here being mine), though it is quoted here for a different reason; note than in it the desirability of a transition from $\Subset$-fragility to $\Subset$-breakability is already on the table:

> It is instructive to repeat our investigation into the distribution of model-theoretic properties with the domain-extension relation from Chapter 2 (see Fig. 3.12[49]). Here the relation that Particular sentences are vulnerable to would presumably be *model reduction* ($\Supset$). The underlying idea is that one deletes objects from the domain to achieve a new model that is a reduction of the old. In practice this is complicated by the fact that our models interpret every primitive individual constant in the language, so that the result of deleting an object with a name is not a model. We will avoid getting into that problem here by restricting our attention to sentences which do not contain individual constants.

Now, hang on a minute. Just after the passage quoted from p. 68 of *Barriers* complained about at the end of our Historico-Philosophical Interlude above, for saying that "(p)articular sentences don't tell you how the whole world has to be, just a particular part of it," on an understanding of "particular sentence" as including existentially quantified sentences, the discussion continued with this:

> Universal sentences, on the other hand, impose global conditions. As a result, adding new objects, with various properties, can make them false.
>
> To illustrate: *Fa* seems like a paradigm particular sentence in our formal language. It has the feature that if it is true in a model *M*, then it is true in all extensions of *M*, where an extension is a model that we get from *M* by adding new objects to the domain, and (optionally) expanding the interpretations of our non-logical predicates to include them (. . .)

So, in Chapter 2, when the particular/universal distinction was in focus, Fa was deemed to be a *paradigm case* of the particular, whereas now in these reflections in Chapter 3 on that same distinction, we are apparently quite happy to avoid a problem by "restricting our attention to sentences which do not contain individual constants." The reviewer feels that this tells in favour of the opinion expressed in the Interlude above, that the whole idea of a particular/universal contrast based on lumping the existential and the specific together, and setting this heterogeneous bunch against the universal, represented the unwise taking over by Russell (the barrier explorer), of an ill-conceived and half-hearted thought bubble on the part of Russell (the logical atomist).

We now resume the passage from p. 112 of *Barriers* above that ended with "individual constants." The observation that $\exists x Fx$, though reduction-breakable, is not reduction-fragile is a neat point which (for me at least) served to restore interest in the concept of fragility in terms of which it is couched, even if that concept has turned out not to be taxonomically fruitful for the study of barriers.

> Several further issues arise because (i) $\Subset$ and $\Supset$ unlike $\curlyvee$ and $\curlywedge$,[50] are not symmetric relations, and because (ii) our classical FOL models are required to have nonempty

[49] This is the table just alluded to.

[50] These are respectively the future-switching and past-switching relations. Future switching permits alterations to (and only to) the interpretations of predicate (or sentence) letters relative to times later than the distinguished present moment n, and past switching does the same for times earlier than that moment. Those relations are actually called basic future (or past) switching on p. 91 (for the propositional case) and p. 102 (for the first-order case), with more general versions on p. 92 and again p. 102, which allow even disruption of the structure of the underlying *frame* in the 'switchable' regions – most imaginative.

> domains. The first feature means that there are no coextensionality results analogous to Propositions 3.2 and 3.3.[51] The second feature means that the results of using model-reduction[52] on the official models for FOL come apart from intuitive expectations. For example, one might intuitively expect existential sentences like $\exists xFx$ to be reduction-fragile (perhaps thinking that one could simply delete from the model all elements that "are F"). But if $\exists xFx$ is true in a model whose domain has only one element, that model has no reduction that makes the sentence false (because its only reduction is the trivial one: itself). So $\exists xFx$ fails to be reduction-fragile. Similarly, $\forall xGx$ is not reduction-constructible[53] – even though one might think intuitively that it should be (or that it would be, if we allowed empty domains).
>
> It is also worth noting, however, that the category of $\Subset$-breakable sentences continues to provide an intuitive class to identify as Universal, and the class of $\curlyvee$-breakable sentences an intuitive class to classify as Future. Here the fact that we only require that the sentence can sometimes lose its truth transitioning to such models means that the existence of some special cases – like singleton models – does not change the overall classification.

The elephant in the room that these recently quoted passages, for all the interesting topics they raise, so conspicuously ignore is that of the comparative merits of (1) looking upward to extensions for spelling out universality, whether in terms of fragility, or in terms of its recent rival, breakability, the discussion in *Barriers* increasingly favours as it proceeds, and (2) looking downward to substructures for the desired explication. This particular elephant has a name – the Łoś–Tarski Theorem – and turns out to be one with which, though it is not mentioned in *Barriers*, Russell is perfectly familiar: see the Aside, below, on her 2011 paper ([72]). The result in question has various nontrivially different formulations or like-named corollaries, the simplest of which is that a first-order sentence is has a prenex normal form equivalent to it with all initial quantifiers being universal if and only if the class of models verifying the sentence is closed under passing to substructures.[54] So what is really wanted is a discussion of the comparative merits of explicating the informal notion of universality in this traditional "look downwards" way, checking for the preservation of a sentence's truth, or in this new "look upwards" style, whether checking for its fragility, *à la* Restall–Russell and early *Barriers* or, as in later (= post-Chapter-7) *Barriers*, for its breakability (non-preservation).[55] Table 1 looks at a few cases to compare verdicts of the

[51] The text shows "co-extensibility results" here but this is a typo, which I have corrected on the basis of the wording of Proposition 3.2, on p. 106. Perhaps *coextensiveness* would have been better still.)

[52] This is the usual *substructure* relation, the converse of the domain-extension relation referred to in the opening sentences. Picking up a point from note 10 above, it looks, from the relative density of the expressions through the course of *Barriers*, as though the decision to recast the prefix from "model-" to "domain-" was made late on – perhaps so that *m-extension* and *d-extension* could neatly distinguish the modal case of adding worlds from the first-order case of adding individuals (to the domain) – but then not fully enforced retroactively.

[53] See note 16.

[54] It is handy to exploit the availability of PNF-equivalents for all formulas, as provided by classical predicate logic (the current default logic in play), though raising, for a more general inquiry, the issue of what to do for logics like intuitionistic predicate logic where no such equivalents are generally available. Perhaps some syntactic criterion is possible, based on the existence of an $\exists$-free equivalent in which all occurrences of $\forall$ are positive, on the usual understanding, according to which in $\forall x(Fx) \vee Rab$, "$\forall x$" occurs positively, in $(\forall x(Fx) \to \forall y(Gy)) \to \neg\forall z(Hz)$, "$\forall z$" occurs negatively (in the scope of one positively occurring $\neg$), "$\forall y$" occurs negatively (in the consequent scope of a negatively occurring $\to$) and "$\forall x$" occurs positively (in the antecedent scope of that negatively occurring $\to$), and so on. Whether or not a version of the Łoś–Tarski Theorem has been investigated for intuitionistic predicate logic along these (or other) lines, I don't know. Of course, one would also have to settle on the appropriate model-theoretic semantics also, since there is more than one option here.

[55] This issue was raised in note 71 of [43], where it is recalled that this result was mentioned (though without being called the Łoś–Tarski Theorem) in a discussion, Humberstone [36], quoted from more than once in *Barriers*, of barrier-like principles in terms of modal strength. (Here necessity corresponds to universality, *à la Barriers*' 'Modal Barrier, Version A' in §5.4.) So it is a bit surprising to see its merits

different approaches.[56]

Example	downward-pres'd, ŁT $\ni$-preserved	upward-frag., RR $\in$-fragile	upward-breakable, R $\in$-breakable
$\forall x(Fx)$	✓	✓	✓
$\forall x(x = a)$	✓	✓	✓
$\forall x(Fx \to Gx)$	✓	✓	✓
$\forall x(Fx \to Fx)$	✓	✘	✘
$\exists y \forall x(Rxy)$	✘	✓	✓
$\forall x \exists y(Rxy)$	✘	✓	✓
$\exists x(Fx)$	✘	✘	✘
Fa	✓	✘	✘
$Fa \lor \forall x(Gx)$	✓	✘	✓
$Fa \land \forall x(Gx)$	✓	✓	✓

Table 1: Illustrating preservation-related properties for Universality
ŁT: Łoś–Tarski; **RR**: Restall–Russell; **R**: Russell (late *Barriers*)

Barriers often has occasion[57] to observe that sentences are preserved by an intermodel relation $\mathcal{R}$ if and only if their negations are preserved by the converse, $\mathcal{R}^{-1}$, of $\mathcal{R}$. In view of this, and the fact that existentially quantified sentences are equivalent to negated universally quantified sentences, the Łoś–Tarski Theorem is sometimes characterized, or relabelled for this purpose 'the dual Łoś–Tarski Theorem', with the content: a first-order sentence has an equivalent in prenex normal form with all initial quantifiers existential if and only if the class of structures/models verifying the sentence is closed under passing to extensions.[58]

Again, though, where Łos–Tarski looks in one direction for the semantic correlate of the existentially quantified (upwards – for preservation) *Barriers* looks in the reverse direction (downwards – for fragility or breakability), as the previous passage shows: "Here the relation that Particular sentences are vulnerable to would presumably be *model reduction* ($\ni$)." Wanted, then: an examination of the comparative merits of the two approaches, and preferably an account not contrasting the universal with anything including the specific, since Fa (to cite the favoured example again) is preserved on passage from structures to

given no attention in the book – even more so, as it *does* get at least mentioned in Russell [72]: see the Aside, coming up, on that paper, and the discussion following it, which includes some speculation as to why it does not put in an appearance in *Barriers*. (The most active area of current interest in the Łoś–Tarski Theorem has been in the ramifications for 'finite model theory' of the 1959 observation by William Tait that the Theorem is false if the quantification over models is replaced by quantification over finite models. See Atserias et al. ([3], p. 210) for a potted history of the relevant developments, and Atserias et al. [4] for new moves in the area.)

[56] *Barriers*, p. 113, is occupied by a larger diagram with a broadly similar intent. The bottom line of our table incorporates an example of breakability without fragility (w.r.t. $\in$) given in *Barriers* on p. 70, where, as we saw above, between the two inset quotations after the end of our Historico-Philosophical Interlude, it is described as a 'Prior-disjunction'.

[57] See note 119 in Appendix C below, and the text to which it is appended

[58] Textbook presentations of the relevant results and variations on them are given in Chang and Keisler [11] Theorem 3.2.2 (p. 149), which coincides in content with Thm. 5.2.4, p. 307, *q.v.* also for the corresponding results on existential formulas and theories with existential axioms (with Theorem 5.2.3 answering the earlier Exercise 3.2.1 on p. 161). Elaborations along these and further lines – using equivalence relative to a given first-order theory and models of that theory, allowing free variables in the formulas concerned, etc., can be found in Hodges ([33], p. 295*f.*).

substructures no less than in the reverse direction. We conclude this discussion, after noting a strange difference between *Barriers* and Russell [72] on the topic we have been discussing, with a couple of less ambitious thoughts.

Aside on Russell's 2011 Paper. *Barriers* often points to aspects of Russell [72] that advance the discussion of barrier principles from where it stood with Restall and Russell [65], and especially with the move from the General Barrier Theorem to the prototype of the Limited General Barrier Theorem as the relevant result to apply. But there is one part of [72] that does not survive into *Barriers*, and that is precisely the omission complained about above: any reference to Łos–Tarski (beyond the glancing comment quoted in the preceding paragraph about "Particular sentences" being presumably vulnerable to model reduction). It was surprising to read the following, after finishing reading *Barriers*, from p. 146 of [72] and, after the vertical ellipsis, by material from p. 147;[59] note that "$\supseteq$" is being used where "$\Supset$" is used in *Barriers* (*see note 28*), with "A" rather than "ϕ", as a schematic letter for sentences:[60]

> **Definition 2** (Genuine Particularity) A sentence is *genuinely particular* iff for each M, M', if $M \models A$ and $M' \supseteq M$, then $M' \models A$.
> (...)
> **Definition 3.** (Genuine Universality) A sentence is *genuinely universal* iff for each M where $M \models A$, there is some $M' \supseteq M$ where $M' \not\models A$.
>
> ⋮
>
> It is also worth noting that by the Łoś–Tarski Theorem (...) the set of genuinely particular sentences characterized here will be the set of $\exists_1$ sentences, or sentences which in prenex normal form consist of a string of existential quantifiers followed by a quantifier-free formula.

Thus although the existential version of Łoś–Tarski is used here for (genuine) particularity, no consideration is given to (genuine) universality on similar lines, as *downward* replacing *upward* preservation ("$M' \subseteq M$" or "$M \supseteq M'$" replacing $M' \supseteq M$). I'm not sure why this option is not raised in Russell [72], even if only to be ruled out on this or that basis, or why, when it came round to *Barriers*, its prospects would not be reconsidered, given that (in the notation from Hodges [33] use by Russell) $\forall_1$ sentences are equivalent to negated $\exists_1$ sentences and *Barriers* often mentions that ϕ's being preserved by (or "anti-fragile w.r.t") $\mathcal{R}$ is equivalent to $\neg\phi$'s being preserved by $\mathcal{R}^{-1}$. (For detailed page references to *Barriers* on this matter, see the paragraph in Appendix C that follows Example 7 there, including note 119.) ***End of Aside.***

We return to the 'less ambitious' thoughts alluded to before the Aside. First, one can of course ask what the barrier situation is with the Łoś–Tarski version of universality and 'existentiality' – as it seems safer to say than 'particularity' given *Barriers*' tendency to think specific instances $\phi(t)$ of a quantified formula $Qx\phi(x)$ have a greater claim to being co-classified with $Qx\phi(x)$ when Q is $\exists$ then when Q is $\forall$. These 'specific' cases can prove problematic for candidate barriers from the existential to the universal, since Fc, for example, is in both the universal and existential camps from a Łoś–Tarski perspective (and could

[59] I have also put the italics on the *definiendum* rather than on the parenthetical label for the definition.

[60] Russell also uses the notation $\exists_1$ used in Hodges [33] in place of the older notation Σ_1, where a Σ_n formula is one whose initial block of quantifiers is existential, followed by a block of universal quantifiers,. . . ,until after n such homogeneous blocks we run out of quantifiers, so that $\exists v_1 \exists v_2 \exists v_3 \forall w_1 \forall w_2 \forall w_3 \exists x_1 \forall y_1 \forall y_2 \phi$ with quantifier-free ϕ would be a Σ_4-formula. (Similarly, Hodges uses "$\forall_n$" in place of the traditional "Π_n".) *Warning:* these notations are apt to sprout numerical superscripts when higher-order logic is involved, and they are also subject to specialized variations of the meanings explained here when applied in the settings of recursion theory, descriptive set theory, and elsewhere; for more detail, see Odifreddi [57], Chapter IV.

be rewritten to make this syntactically evident as $\forall x(x = a \to Fx)$ and $\exists x(x = a \wedge Fx)$, respectively), making the valid inference from Fa to Fa a violation of the envisaged barrier.[61] Along the lines of the above Aside, with its "genuinely," we can use adverbial modifiers to eke out the terminology here in an effort to respond. What about calling a sentence *properly universal* if it is universal and not also existential, and *properly existential* if it is existential but not also universal, and asking whether a properly universal conclusion can be inferred from properly existential premises? This at least makes the premise class and the conclusion class disjoint, and we would still not be expecting them to be each other's complements. We'd also have to restrict attention to what *Barriers* calls logically synthetic sentences (neither unsatisfiable nor having unsatisfiable negations, that is). Even then, though, the inference from $\exists x(Fx) \wedge Ha$ to $\forall x(Gx) \vee Ha$ gets through, with its properly existential premise and its properly universal conclusion. Further exploration along such lines may be worthwhile, and indeed may (for all I know) already have been undertaken. Perhaps it was difficulties like this that explain why the reference to Łoś and Tarski dropped out (as noted in the previous Aside), in the transition from Russell [72] to (Russell [74] and) *Barriers*. I am inclined, though, to favour a somewhat different explanation, and will offer it in the final paragraphs below, before launching into Part III.

What if one felt that there was much to be said both for the Łoś–Tarski notion of universality (preservation downwards) and for the eventually *Barriers*-approved notion of universality (breakability upwards), feeling that perhaps the latter sensitivity to the addition of new individuals was something essential to the 'characteristic' use of universal quantification, even if there were special exceptions ($\forall x(Fx \to Fx)$ etc.) – and similarly with the dual notions? Then a proposal that comes to mind is to consider again composite notions, as with the "properly universal" (and "properly existential") of the preceding paragraph. Call a sentence *characteristically universal* if it is both preserved downwards (*à la* Łoś–Tarski) and also breakable upwards (*à la Barriers*) – or as *Barriers* itself would formulate this, both $\Subset$-anti-fragile and $\Supset$-breakable – and similarly, *characteristically existential* if it is both preserved upwards and breakable downwards. The hope would be that, with these combinations, one then has the best of both worlds – the Łoś–Tarski world and the *Barriers* world. What would become of a barrier principle w.r.t. this new taxonomy? Well, as the reader has perhaps already anticipated, there is actually nothing new about it except for the terminology: the characteristically universal and the characteristically existential are just the same classes of sentences (with new names) as the preceding paragraph's properly universal and properly existential, respectively, so the same counterexample to a would-be barrier settles that question.[62]

We can conclude with an alternative hypothesis as to why *Barriers* is silent over Łoś–

[61] Note that since universal sentences (resp. existential sentences) are often understood simply to be sentences which in prenex normal form have an initial universal (resp., existential) quantifier, here we mean instead having *all* the quantifiers in the PNF quantifier prefix universal (resp., existential).

[62] My own current hunch is that the intuition of upward-vulnerability as a mark of universality – whether this is Restall–Russell $\Subset$-fragility or (late) *Barriers* $\Subset$-breakability – belongs in the first instance to one's account of what makes us count Qx (say) as a universal quantifier, rather than, initially at least, at the level of individual sentences with only such quantifiers in PNF, perhaps along these lines: for Qx to count as a universal quantifier in such-&-such language it is necessary that there should be some open formula $\phi(x)$ of that language for which $Qx(\phi(x))$ is $\Subset$-breakable. Other sentences would count as universal when their PNF representations comprise only occurrences of Q in their prefixes, whether or not they themselves are $\Subset$-breakable. If the usual logical principles governing $\forall$ – in natural deduction, introduction and elimination rules, or in sequent calculus right and left insertion rules, for example – are satisfied, then $\Supset$-preservation is essentially in place, so the justification for the breakability-based condition just suggested might be that if it is not satisfied, the resulting combination of automatic upward and downward preservation will mean that every Q-formulas has the same truth-value in every model, with perhaps further adverse logical repercussions. But this is the merest sketch of a hunch, leaving much that would need to be worked out (perhaps with the aid of a suitable 'disjoint union'-like notion for structures of the same similarity type, applied to arbitrary M, N to provide a structure to proceed to upwards from M and thence downwards to N).

Tarski. The barrier theses that can be established as applications of the Limited General Barrier Theorem each involve a taxonomy based on a single intermodel relation $\mathcal{R}$. The premise class for the application consists of the $\mathcal{R}$-preserved sentences and the conclusion class, of the remaining sentences – the $\mathcal{R}$-breakable ones. And we have the *unless*-clause to handle the recalcitrant inferences which, though valid, would, in the absence of that clause, transgress the barrier. The word "unless" here, as mentioned in Appendix A (esp. note 93), has a tendency to set off what follows it as in some way a marginal counter-consideration. By way of illustration, compare the plausibility of the following two stretches of dialogue. A proposed outdoor art installation is under discussion, the setting being April in England, and with neither contribution intended sarcastically: (1) Let me assure you, everything should go to plan, unless we have an earthquake within the next four weeks; (2) Let me assure you, everything should go to plan, unless it rains within the next four weeks.

Returning now to the universal and the particular/existential taxa, and imagine a barrier thesis for which the premise class consists of the universal sentences and the conclusion class, of the rest. This would be an unlikely barrier thesis to consider, since a from universally quantified sentence, the corresponding existentially quantified sentence follows – including of course a prenexed case in which the premise has an initial string of "$\forall$"s and the conclusion, a corresponding string of "$\exists$"s. Its counterintuitive content notwithstanding, we press ahead regardless, invoking the Limited General Barrier Theorem. So we need an $\mathcal{R}$ which preserves the truth of universal sentences from a model to an $\mathcal{R}$-related one, and the 'universal' form of Łoś–Tarski supplies one in the shape of $\Subset$, which we will help ourselves to just as *Barriers* uses the 'existential' version of Łoś–Tarski to provide a suitable $\mathcal{R}$ – namely $\Supset$ for the premise class of the more natural oppositely directed appeal to the Limited General Barrier Theorem. So for the premise class now we have the class of $\Subset$-preserved sentences – *downward*-preserved as it would have been put in the labelling of Table 1 if there had been more columns. And whether we call the complementary class the particular, the existential or something else again, the pertinent special case we want to consider is that of $\forall x(Fx) \models \exists x(Fx)$, for which the Limited General Barrier Theorem tells is that if $\forall x(Fx)$ is $\Supset$-preserved (which it is), and $\exists x(Fx)$ is not (as it certainly isn't), then – and now we use the 'argument set' formulation, set out in note 19. Here we are concerned with a one-premise inference so modifying the formulation to replace Γ with $\{\gamma\}$ and thus equivalently with γ itself, we have: If ϕ is $\mathcal{R}$-breakable and γ is $\mathcal{R}$-preserved, then $\gamma \not\models \phi$ unless $\gamma \wedge \phi$ is $\mathcal{R}$-preserved. So, in the present case, with γ as $\forall x(Fx)$, ϕ as $\exists x(Fx)$, and $\mathcal{R}$ as $\Subset$, we have what the *if*-clause asks for: $\forall x(Fx)$ is $\Subset$-preserved and $\exists x(Fx)$ is $\Subset$-breakable. So we conclude that $\forall x(Fx) \not\models \exists x(Fx)$ unless $\forall x(Fx) \wedge \exists x(Fx)$ is $\Subset$-preserved. This is quite correct of course, since it is *not* the case that $\forall x(Fx) \not\models \exists x(Fx)$, but it *is* the case $\forall x(Fx) \wedge \exists x(Fx)$, being equivalent to the premise $\forall x(Fx)$, is $\Subset$-preserved. But here the "unless" hardly directs us to some out-of-the-way mostly marginal possibility, and, perhaps more to the point: do we want the justification of such barrier theses as have some pre-theoretical appeal, to apply quite as widely as this? The Limited General Barrier Theorem seems to have no more difficulty applying to justify the implausibly postulated barrier on inferring existential/particular conclusions from universal premises as it does to justify the more plausible converse barrier (with the intermodel relations differently – but not themselves implausibly – chosen). So here we complete the presentation of our alternative hypothesis as to why the Łoś–Tarski theorem is backgrounded in *Barriers*. Its inclusion would direct attention to an objection, whether or not this objection is in the end justified, along the the following lines: the *unless*-clause was meant to be an escape clause for taxonomically arguable cases rather, than to provide a gaping hole for the mass escape of what would otherwise be seen as the clearest counterexamples.

III

We turn now to aspects of what *Barriers* has to say about normal modal logics (mostly in Chapter 5) and context dependence (in Chapter 7), mainly attempting to convey the general gist, and to set the stage for some discussion in Appendices B and C below, with rather less concentration on the barrier theses in play, whose definitive formulation is in any case spelt out as a series of applications of the Limited General Barrier Theorem in Chapter 9 along, several of which we have already encountered. The discussion concludes with some remarks on the interesting Dedication at the start of *Barriers*. Our own discussion of various modal principles and logics uses *sans serif* font and explicit Segerberg–Lemmon–Scott–Chellas notation (KTB, with "KT4" abbreviated to "S4," etc., though when quoting directly from *Barriers* reproduce the nomenclature used there – e.g., in this last case, sometimes "S4" and sometimes "*S4*"). We begin with modality, for which we first need to back up a little to the tense-logical material in Chapter 3.

After we are introduced to the semantics for (initially, propositional) tense logic, we read "If the only frames we allow are isomorphic to the real numbers, there will be no models with, e.g., right endpoints."[63] Now in fact, in *Barriers* only models have been (or will later be) defined, not frames. Such a definition would have been a useful inclusion not just for those encountering the word at the occurrences listed, but for its ramifications elsewhere. These include emphasizing that what in *Barriers* is called, for example, an S4 model is not just any model verifying all theorems of S4, or even a model in which not just the distinguished point @ – explained below – but *every* point verifies those theorems, but rather: a model *on a frame for* S4, i.e. a frame on which those theorems are valid, so that the model's verifying them does not depend on the particular choice of what *Barriers* records as its I component. It's not that the characterization of the class of S4 (etc.) models in structural terms is not given – a table on p. 122 does say that these models need to have an accessibility relation which is is reflexive and transitive (with similar conditions for the other five normal monomodal logics considered);[64] it's just that its significance – that frames with such accessibility relations are precisely those on which every theorem of S4 is valid – remains unclarified. On a first reading of *Barriers*, I must admit to not noticing any reference to frames at all, and thought that this seemed a good choice pedagogically, to keep things as simple as possible for the more general reader with philosophical rather than specifically logical interests. So I was surprised, looking more closely, to see that they are mentioned, not just on p. 94 but on pp. 116, 126, and 132, thereby creating a need for something by way of supplementary explanation on this front, after all. Given the introduction of frame conditions, the need arises not just from the general desirability of an orderly presentation, but from occasional specific points arising. For example, the discussion of S5 in the mainly (alethic) modal Chapter 5 is in terms of models on equivalence-relational frames, since those are precisely the frames on which all of its theorems are valid, whereas when we come to a deontic–alethic modal logic in Chapter 8 the alethic side of the story is given by a much smaller class of frames – those on which accessibility is not just any old equivalence relation, but, in particular, the universal relation, and no mention is made of the fact that the purely alethic fragment of this logic is again S5, despite the smaller class of frames.[65]

[63] *Barriers*, p. 94, with the commas around "e.g." inserted here.

[64] Strictly, these are conditions on frames and not just on their accessibility relations, since for example, where $w_1 \neq w_2$, frames $F = \langle \{w_1\}, \{\langle w_1, w_1 \rangle\} \rangle$ $F' = \langle \{w_1, w_2\}, \{\langle w_1, w_1 \rangle\} \rangle$ have the same accessibility relation – the same (one-membered) set of ordered pairs, that is – though only F is a serial frame. Nor does this situation arise just because of the existential quantifier in the seriality condition: of the two frames, only the first is reflexive. Related issues are aired in Humberstone ([42], p. 90, n. 81).

[65] This is something that anyone teaching a class on the material could supply for curious students, of

A second example merits a discussion of its own. It arises for what *Barriers* calls the A-version modal barrier thesis, which uses a Kripke-semantic adaptation of the Universal/Particular distinction – m(odal)-universality vs. m-particularity, Remark 5.2 (*Barriers*, p. 152) tells us that $\Box p$ is modally universal, which at the current stage of the discussion is a fragility notion, though in the final reckoning (Chapter 9) it will be recast as the corresponding breakability notion, in both cases w.r.t. the intermodel relation of m(odal)-extension. Informal candidate barrier thesis: no m-universal conclusions from m-particular premises. The same issue arises here as was raised in Part II above: it is not clear what justifies this lumping together of, let's say, the 'm-specific' with the 'm-existential', under one heading. The more even-handed approach was taken in the paper cited in note 55, where the universal, specific, and existential notions were labelled *global*, *local* and *indefinite*, and a version of the current informal barrier thesis appears on p. 323 there, though, as *Barriers* correctly suggests (p. 120*f.*), in a seriously garbled form.[66] This garbling consisted in confusing necessity with necessitativity, where being necessary is the usual notion of being incapable of falsehood, and being necessitative is a matter of being (or amounting to) an attribution of necessity (correct or otherwise) – thus "It is necessarily the case Spanish pigs eat acorns" is necessitative without being necessary (or even true), while "Whatever is known to be the case is indeed the case" is necessary without being necessitative.[67]

For the range of monomodal logics under consideration, models M have the form

$$\langle W_M, R_M, @_M, I_M \rangle,$$

where the ingredients after the non-empty universe W_M (called W when it is clear what model is at issue) are much as in the deontic–alethic models described two paragraphs before the Aside on Notation and Variations in Part I above, except that here R_M is a binary relation on W_M – no longer required to be the universal relation $W_M \times W_M$ – much as described in that Aside, here to serve as an accessibility relation for $\Box$ (and of course $\Diamond$). As to the pertinent intermodel relation: N is an *m-extension* of M ("$M \sqsubseteq N$" on p. 125[68]) when the following conditions are satisfied:

- $W_M \subseteq W_N$
- for all $u, v \in W_M$: uR_Mv iff uR_Nv
- $@_M = @_N$

course. *Barriers* does offer a helpful hint when a similar point comes up in the later discussion of indexical logic, with this remark from p. 165: "Since S5 can be modelled without an accessibility relation, let's assume (for simplicity!) that we are working in S5 and drop R" – though this itself is likely to need some elaboration for those not already familiar with the details. Indeed, there are numerous such 'teaching opportunities' scattered throughout *Barriers*, where the text is intelligible without elaboration, but more or less invites it, if used for teaching purposes – as in the representation, without explicit comment, on p. 82 of "The Queen is wise" using a form of the Russellian theory of descriptions, with Q as the predicate "is a queen".

[66] "Either the necessary conclusion follows without the help of the contingent premise(s), because it is a valid formula outright or a semantic consequence of the necessary premises which convey global information, or it is neither valid nor a semantic consequence of the necessary premises, in which case, any additional merely contingent premises will not be strong enough to make good the lack, conveying, as they do, this purely local information." (A similar passage from the same page is quoted in *Barriers*, p. 120.) Note that the use of "global" here has not exactly the same as its use in note 37.

[67] The confusion is particularly inexcusable as the classic exposé of this confusion in the literature – Routley and Routley [68] – is actually cited in Humberstone [36]. Anderson and Belnap ([2], p. 36) also provide a good discussion, except that *necessitative* is dumbed down to *necessitive*. Yes, it *is* a bit of a mouthful, but come on, let's all make a bit of an effort here – at least in writing the word.

[68] But $\sqsubseteq$ has become $\sqsubset$ in Chapter 9 (beginning in a footnote on p. 200, and then in the main text on pp. 202, 204) – not ideal for a reflexive relation. Perhaps I have overlooked something, since the setting is somewhat different on p. 202 (with expanded versions of the Chapter 5 models in play), and the relation involved gets a different name: w-extension instead of m-extension. But even if this is being differently notated and labelled because of the different setting, the relation is still reflexive.

- for all p_i and all $w \in W_M$, $I(p_i, w) = I_N(p_i, w)$.

Modal particularity and modal universality are then defined much as in the first-order case (without the 'modal'): ϕ has the former property if it is m-extension preserved, and the latter if it is m-extension fragile. (In the end – p. 204 – as in other cases, fragility here will be replaced by breakability.) Remark 5.2 says that $\Box p$ is modally universal and has the following proof, which is quoted nearly verbatim,[69] without adjusting the "V_M" notation to that with "$\models$":

> For suppose $\Box p$ is true in M. Then $V_M(\Box p, @) = 1$ and for all worlds $w \in W$ such that $@Rw$, $V_M(p, w) = 1$. Let N be a model $\langle W_N, R_N, @_n, I_N\rangle$ such that $W_N = W_M \cup \{w'\}$ where $w' \notin W_M$ (so we add a new world to W), we get R_N by adding $\langle @, w'\rangle$ to R_M and closing the result under reflexivity, symmetry and transitivity. $@_M = @_N$ (the actual world is the same one) and $I_N = I_N \cup \langle p, w', 0\rangle$ (we evaluate every world that was in W_M as before, and then add that p is false at w'). The result is a model that meets the frame-constraints on being an S5 model (and with it the constraints on being a model for any of the other five modal logics in our list) and we've added a world where p is false, w', which is accessible from @. Then $V_N(\Box p, @) = 0$, and $\Box p$ is false in N. Since M was an arbitrary model making $\Box p$ true, $\Box p$ is fragile with respect to m-extension and hence modally universal.

The point of quoting this example here is not to discuss either the merits of the proposed characterization of m-universality or the attendant barrier thesis – the current incarnation of the General Barrier Theorem (Thm. 4.1) from Chapter 4 – but simply to note some features of the example and the special status of S5 and its associated frame conditions in the treatment of the example. These conditions (reflexivity, symmetry and transitivity) happen all to be universal strict Horn conditions in the semantic metalanguage, so it makes sense to talk about "closing the result" under these conditions – i.e. passing the smallest extension of the frame relation (on the w'-extended frame) satisfying the conditions. As the last parenthetical comment in the passage just quoted indicates, this will work for all sublogics of S5 – the five alluded to (K, KD, KT, KTB, and S4/ KT4)[70] – all answering to this description. In the case of KD, the class of frames comprises those meeting the $\forall\exists$ condition of seriality is a case for which there is no uniquely natural meaning for "the closure of R under" the condition to have, but the argument in *Barriers* still works, since seriality follows from reflexivity.[71] More often, the verdicts for a given formula vary from logic to logic, as with Remark 5.9 (p. 128):

> The status of $\Box\Diamond p$ depends on the logic. In *S5* models it is m-extension anti-fragile, in *S4* models it is m-extension fragile, and in *B* models it is neither.

Barriers goes on to explain/justify these classificatory remarks and to relate them to the relevant barrier thesis, given by applying the General Barrier Theorem, on which see the 'Optional Aside' in Part I above, where it is suggested that the distance from the notion of classifying not just premise-sets ("Γ" in most of the formulations in *Barriers*) – already a departure from classifying the premises themselves individually – but whole *argument-sets* ("$\Gamma \cup \{\phi\}$") is too great for this treatment to be regarded as a defence of the original

[69] The sole exception is that what appears here as "w'" appears in *Barriers* as "w*," which is potentially confusing as our bibliography cites several publications in which the latter is used in place of the "@" of *Barriers*.

[70] *Barriers* uses the (not uncommon) nomenclature D, T, and B for KD, KT, KTB, respectively.

[71] Several questions may occur to the reader. When might it not work? What if we considered normal modal logics that were not sublogics of S5? Why do we care about the frames anyway? Would there be something wrong with just concentrating on models for the logics concerned?

intuitive barrier theses.[72] The final post-fragility incarnation of the Version A barrier thesis (Chapter 9) applies instead the Limited General Barrier Theorem, and the barrier denies passage from (collectively) non-modally premises to a modally-universal conclusion, subject to the *Unless*-clause shared by all such applications.

We turn our attention to the 'version B' modal barrier thesis, which concerns not relative modal strength, in the sense in which necessity-attributions (m-universal sentences) are stronger claims than (unmodalized truth-attributions, which are stronger than) possibility-attributions,[73] but rather attributions of modal properties – necessity and possibility alike – as opposed to non-modal properties. To address this topic, *Barriers*, p. 130, introduces first an intermodel relation call the *m(odal)-switch* relation, and later on (p. 133) the a(ctual)-switch relation, both of which are symmetric relation (by contrast with m-extension).

Lamentation on Switching and Shifting. One of the hardest things about discussing *Barriers* is trying to remember the intermodel relation terminology for these symmetric (indeed equivalence relational) cases. In Chapter 3, we have future-switching and – briefly: p. 108 – past-switching (see note 50); here in Chapter 5 we have modal-switching and actual-switching, but in Chapter 7 we have (partial) context-*shifting*. Then – just in case you thought indexicality was sufficiently different to merit a separate terminological style of its own – in Chapter 8 (Normativity) we have s(uperb)-*shifting* again. I was at a complete loss to tell when we were shifting and when we were switching, short of looking it up every time, and wishing a single verb had been chosen to handle all the cases. I apologize if I have used one of *switch*, *shift*, when to follow *Barriers* I should have used the other. ***End of Lamentation.***

We will get to a(ctual)-switching in due course, beginning here with m(odal)-switching. The *m-switch relation* is defined to hold between M and N when:

- $@_M = @_N$
- for all p_i, $I_M(p_i, @) = I_N(p_i, @)$.

Because of the first condition, the second can be stated without needing to subscript the occurrences of "@," and, as *Barriers* says (p. 130) "m-switching allows big changes," including even "changing whether the actual world is accessible from itself," though earlier on the page it is mentioned that we "allow the rest of the model to vary, consistently with the restrictions on R for the logic in question." Later (base of p. 131), it is made clear that it is not just the rest of the model, but @ itself, that has to be accessible to @ in the case of extension of KT: "changing whether $@R@$ is only an option with K and D models." In terms of this relation, a sentence is defined to be *Actual* iff m-switch-preserved and *Modal* iff it is m-switch fragile.

As *Barriers* notes on p. 132 (Remark 5.14), this makes $\Diamond p$ come out neither Actual (m-switch fragile) nor Modal (m-switch preserved) in extensions of KT, which, though initially an unwelcome surprise, is to be explained by $\Diamond p$'s being, in those extensions, "a disjunction, one of whose disjuncts *is always about the actual world*" (emphasis in the original). That is, since with the reflexivity condition in force for (the frames of) the models concerned, where

[72] Example: Discussing the B-valid argument from p to $\Box\Diamond p$ (which is of course not just KTB-valid but KB-valid), *Barriers* (p. 129*f.*), writes "For sublogics of S4, the argument is not valid, and so not a counterexample. If the logic is B or S5, then the argument set $\{p, \Box\Diamond p\}$ is m-particular, not m-universal, and so again the argument is not a counterexample to version A of the modal barrier."

[73] Naturally, since we are thinking of alethic interpretations of the formal language with $\Box$ and $\Diamond$, these remarks are tailored to (normal modal) logics containing all instances of $\Box\phi \to \phi$ but not all instances of $\phi \to \Box\phi$.

$R^{\neq}$ is the intersection of the accessibility relation R with the relation of non-identity,[74] we have:

For any sentence ϕ, we have $\Diamond\phi$ true at w just in case either ϕ is true at w or for some w' such that $wR^{\neq}w'$, ϕ is true at w'.

When w is @, as in evaluating non-modally-embedded occurrences of $\Diamond\phi$ in a model, the second disjunct is the part making the genuinely Modal(-as-opposed-to-Actual) contribution to the truth-conditions. Similarly, $\Box\phi$, for these purposes can be thought of as a conjunction of ϕ with $\boxdot\phi$, interpreting $\boxdot$ with $R^{\neq}$ – or, *de novo*, some perhaps non-reflexive relation or other – as its accessibility relation.[75]

What might we do in the absence of any such explicit decomposition of $\Box\phi$ into $\boxdot\phi \wedge \phi$ (and $\Diamond\phi$ into $\diamonddot\phi \vee \phi$ – p. 132's "disjunction one of whose disjuncts is always about the actual world")? The reaction in *Barriers*, begins by recalling a move made in the earlier discussion (Chapter 3) of time and tense, which introduced a distinction between being Purely Future and being Partly Future, with $\mathcal{F}p$ having the former status and $\mathcal{P}p \wedge \mathcal{F}q$ the latter. There seems to be no precise definition of these categories in the course of that earlier discussion (on pp. 98 and 104), though, in which they are mentioned and illustrated with the example, $\mathcal{P}p \wedge \mathcal{F}q$, just mentioned, but a definition is given in the course of the modal discussion on p. 133.[76] This requires the definition of the second intermodel relation foreshadowed above called a(ctual)-switching; although our encounter with it here will be very brief, we shall have occasion to consider the idea behind it in Appendix B (on aboutness). N is defined to be an *a-switch* of M iff:

- $W_N = W_M$
- $@_N = @_M$
- $R_N = R_M$
- for all p_i and all $w \in W \setminus \{@\}$, we have $I_N(p_i, w) = I_M(p_i, w)$.

In view of the first two conditions, we can write "W" and "@" in the final condition without having to disambiguate with subscripts for the models. With this in place *Barriers* defines ϕ to be *Partly Modal* if it is m-switch fragile, and *Purely Modal* if it is (i) m-switch fragile, (ii) m-switch constructible, and (iii) a-switch preserved. (For 'constructible,' see note 16.)

[74] This is sometimes called the irreflexive interior of R (the largest irreflexive subrelation of R, just as $R^{=}$ is the smallest reflexive relation extending R). The contrast is with (say) $R^{=}$, the union of R with the identity relation, or more briefly the reflexive closure of R. *Warning*: the symbol "$\boxdot$" about to be introduced is often used for a Box operator with, as the associated accessibility relation, the reflexive closure of that interpreting $\Box$ – e.g., Boolos [10] (beginning on p. 8), and Jeřábek [45], whereas the use of $\boxdot$ introduced below is interpreted as having for its accessibility relation the irreflexive interior of that interpreting $\Box$.

[75] A treatment along these lines can be found in Humberstone [39], where it is noted that we can have the truth-conditions sketched exactly as here, in which $\boxdot\phi$ checks for ϕ's truth at all R-related worlds distinct from the current world of evaluation (in which case we can drop the parenthetical "non-modally-embedded"), or alternatively, in which what fixes $\boxdot\phi$'s truth is that of ϕ all R-related worlds other than @. (These are two separate options which come apart for embedded occurrences of $\boxdot$, since modal embedding takes us away from the distinguished point and so exempting the then-current world from the ambit of what an undotted $\Box$ would be is different from exempting the model's actual world.) These are easiest to think about in the case in which R is the universal relation $W \times W$, for which case the logic determined with the first ("current world") reading for $\boxdot$ is often called the "logic of elsewhere." See the references supplied in [39], esp. notes 15, 16, 17. A complication for the approach of *Barriers* here comes from the fact that the class of irreflexive frames is not modally definable, and the completeness of axiomatically presented normal modal logics w.r.t. classes of such frames is established using various case-by-case arguments at best (by appropriate surgery on their canonical models), or perhaps not at all ("Kripke-incomplete" modal logics).

[76] Even more strikingly absent from the discussion here is any reference to that in Lewis [52] of the different ways in which to understand what it is for a statement to be *partly about* a given subject on the basis of an account of what it is for a statement to be *entirely about* that subject matter.

So being Partly Modal coincides with being modal *simpliciter* on the previous account (= satisfying (i)), and being Purely Modal imposes two further conditions, (ii) and (iii), making evident the implication from *Purely* to *Partly* Modal. The added condition (ii) amounts to saying that, in addition to ϕ's being m-switch fragile, so is ϕ's negation. But it is (iii) that invokes the newly introduced relation of a-switching. This relation is of considerable interest in its own right, and a first-order non-modal analogue of it has made its way into the literature on *aboutness*, as is explained in Appendix B below. As to the role it is playing here, Remark 5.15 (*Barriers*, p. 133) illustrates its bearing on the Partly-vs.-Purely Modal status of $\Box p$ in extensions of KT, which the interested reader can consult.[77] *Barriers* concludes that the resulting 'Partly Modal' status, for the case at hand, gives the right result, though one can't help wonder whether a simpler Purely/Partly contrast might achieve the desired result for this and other cases – in particular, in respect of the (let's just say) rather complicated spelling-out of the Partly Modal.

One last point, obvious enough but of some interest, to be made before leaving Chapter 5 behind concerns the relation between this new a-switch relation and the earlier m-switch relation – neither of which, incidentally, manages to make it into Chapter 9 despite the latter's title "All the Barriers": there would appear to be some intermodel relations with a kind of 'off Broadway' status, not even getting their own special symbols $\circledcirc$, $\sqsubseteq$, etc. (See the table at the top of p. 201 in *Barriers*, and as well as note 7 above). But, returning to the point: the intersection of the a-switch relation and the m-switch relation is none other than the identity relation (on the class of models over which these relations were defined). They are evidently both reflexive, and conversely, their intersection is included in the identity relation because they are alike in respect of having the same @ and agreeing on the p_i at that @, in virtue of being standing in the m-switch relation, and that have the same W, and R, as well as agreeing on the p_i everywhere other than @ in virtue of standing in the a-switch relation. The features just listed exhaust all there to be being this or that model, so they are the same model. That is the last we shall hear of a-switching until Appendix B.

Turning now to the treatment of context-sensitivity, we have, in Chapter 7, perhaps the most complicated and difficult in the book. Its antecedents are Russell [72] and [73], and it is cast in *Barriers* as playing something of a watershed role in the overall narrative development – though this requires attention only to a small part of the apparatus introduced in the chapter and the relevant point will be explained at the end of our discussion here, as the prototype for eventually favoured (post-fragility) paradigm for barrier theses: the Limited General Barrier Theorem, emerges in the course of handling context-sensitivity in [72]. (This bears on the emergence of the *Unless* clause, other aspects of which are deferred to Appendix A.)

Models for the indexical logic the current chapter is concerned with have the rather elaborate from of sextuples $\langle W, T, D, P, C, I\rangle$ in which W is there for (alethic) modal reasons as in Chapter 5 with the universal relation $W \times W$ as accessibility (so no need to give it separate mention), T for the set of times, as in Chapter 3 (and both of them in Chapter 6 for multimodal combinations – not discussed here), taking moments as integers so that again we avoid piling an accessibility ("earlier than") relation into the ingredients of the model for these either, D is our domain of individuals (fixed domain semantics for simplicity, if nothing else – though with an existence predicate), P is a non-empty set (of *places*); C is the main novelty, housing the *contexts* of the model; and we have I to interpret the primitive

[77] More generally, notice that, because truth in a model M is a matter of truth at @ ($= @_M$), every non-tautologous $\Box$-free formulas is a-switch fragile, since the definition of the a-switch relation constrains I only insofar as truth-values for sentence letters at worlds *other than* @ are concerned; for @ itself, anything goes. So we can always a-switch from M to an N for which $I_N(p_i, @)$ provides a falsifying truth-value assignment to our formula, as long as there is such assignment (as is secured by the assumption that we are dealing with a purely truth-functional formula which is not a tautology).

vocabulary in terms of the rest of the apparatus. In more detail in respect of D and P: we are working in a two-sorted first-order language (with numerous exotic accretions), and all predicate letters come with sortal typing – such as a 'Located at' predicate *Loc* for relating individuals to places. And in respect of C (or more explicitly C_M, with a similar elaboration for the other model components when desired), which is the most significant novelty, this is a quadruple $\langle a_C, p_C, t_C, w_C\rangle$ (or a_{C_M}, p_{C_M} etc.) for, respectively, the agent (potential speaker or writer) of the context, the place, time and world of the context, a_C, p_C, t_C, w_C being elements of D, P, T and W, respectively. And for more detail in respect of atomic predicates (predicate letters), here is the explanation from p. 164:

> If Π is an $\langle m, n\rangle$-place non-logical predicate, then $I(\Pi)$ is a function such that for each $t \in T$ and $w \in W$, $I_\Pi(t, w) \subseteq (D^m \times P^n)$ (i.e. a function from time-world pairs to an sequence consisting of an m-tuple of individuals from D followed by an n-tuple of places from P – the predicate's *intension*).

The language to be interpreted by these models, which *Barriers* called the language of IL (for Indexical Logic – a variation on Kaplan's Logic of Demonstratives) sports indexical terms $\mathfrak{i}, \mathfrak{h}$ (mnemonically: *I*, *here*) – the exotic accretions alluded to above – for denoting a_C, p_C, respectively, as well as ordinary (but 'sorted') variables ranging more generally over D and now also P.[78] Finally, there are sentence operators $\mathcal{N}, \mathcal{A}$ (mnemonically *Now*, *Actually*) directing us, from whatever we are evaluating a sentence, to what is the case at t_{C_M}, w_{C_M}, respectively (formerly n_M, $@_M$). The notion of truth that is defined recursively is that of truth in a model relative to a time $t \in T$ and a world $w \in W$ (and also a variable-assignment visible in the Optional Aside below, but for the sake of simplicity, we ignored here), which are being thought of determining the circumstances under which what is said is true and will therefore be called *circumstantial* parameters in contrast to the *contextual* parameters which make up c_M and are thought of as fixing what is said on the basis of the sentence uttered and the context of utterance. And truth in M *simpliciter* is truth w.r.t. relative to the contextual parameters t_C and w_C, now playing the role of the circumstantial parameters. Some details differ from the master plan laid down in Kaplan's publications are mentioned in *Barriers*, which supplies references to those publications (as well as to Russell's own earlier engagements with this material, mentioned above).

Optional Aside on Semantic Details. For those that would like a fuller spelling out of the semantics of Chapter 7, and because one aspect of it will be needed for Appendix B (on aboutness as it arises in the book), we supply the main points here, essentially as they appear on p. 167, so the initial "We" is Russell's:

We write $[\alpha]_{M_{gtw}}$ for the denotation of the term α in the model M, on the assignment g, at the time t in the world w. We write $V_{M_g}(\phi, w, t)$ for the truth-value of the wff ϕ at a world w and time t, in the model M on the variable assignment g.

$$[\alpha]_{M_{gtw}} = \begin{cases} I(\alpha), & \text{if } \alpha \text{ is a non-logical constant} \\ g(\alpha), & \text{if } \alpha \text{ is a variable} \\ a_C, & \text{if } \alpha \text{ is } \mathfrak{i} \\ p_C, & \text{if } \alpha \text{ is } \mathfrak{h} \end{cases}$$

This is taken almost verbatim from p. 167 of *Barriers*, except that "i-constant or p-constant" (individual constant or place constant, that is) has been abbreviated to "constant" and parenthetical glosses on the third and fourth cases to the effect that $\mathfrak{i}$ (resp. $\mathfrak{h}$) refers to

[78] This means that the predicate letters also have to come with 'sorted' positions. In addition, two of them, the first already mentioned above, *Loc* and *Ex* – for *is located at* and *exists* – are given a special treatment in *Barriers* (p. 166) we do not go into here. There is an evocative description of Russell's first encounter with this material in the 'autobiographical prelude' to her [75].

the model's agent (resp. place). Next we have to define truth in M relative to t, w, for the atomic formulas, which involves the places/individuals sorting (as in the inset quotation from p. 164 above), so that on p. 167, Π is a metalinguistic variable over predicate letters taken as having m places for individuals followed by n places, to be occupied respectively by m i-terms and then n p-terms, but here for brevity we do the case for $m = n = 1$, writing p, t for p_1, t_1:[79], and using the *Barriers* "V"-style notation rather than "$\models$":

$$V_{M_g}(\Pi ip, t, w) = 1 \text{ iff } \langle [i]_{Mgtw}, [p]_{Mgtw} \rangle \in I_\Pi(t, w)$$

The recursive clauses for Boolean connectives, modal and tense operators (invoking the t and w parameters), the actually and now operators $\mathcal{A}$ and $\mathcal{N}$ (invoking w_C and t_C) and quantifiers (invoking g) then proceed (p. 168) as expected, given our independent familiarity with them from earlier chapters of *Barriers*,[80] transposed the current more complex setting. ***End of Aside.***

Matters are complicated by the fact that while *actually* and *now* are given an operator treatment, *here* is treated as a term – even though in most of its occurrences in English *here*, like *now*, behaves adverbially. The complication comes not from that difference but from the question of the relevance of the contrast between that difference and another, described in *Barriers* on p. 170 using an English sentence with adverbial *here*. At the top of that page we have an explanation in terms of the semantics offered, of the validity of the inference from *It is raining* to *It is raining now* (or its formal analogue) from – ϕ to $\mathcal{N}\phi$. Here validity is understood as preservation, for all models M, of truth relative to M with its contextually set circumstantial parameters t_{C_M} and w_{C_M}. Then we get the same story for the inference from *It is raining* to *Actually it is raining* (ϕ to $\mathcal{A}\phi$) and conversely – though it is not clear why only the unilateral case was presented for $\mathcal{N}$, no doubt puzzling some readers since the cases seem entirely parallel in this respect. But not surprisingly, since the indexical parameters to which truth was relativized were just t_{C_M} and w_{C_M}, we do not have the same situation for here, p_{C_M} having been left out in the cold for such purposes. It would have been helpful to have had a bit more explanation for why choices like these are made the way they are. Sentences like "What's the situation here right here and now, though?" suggest that both *here* and *now* should be treated in the same way syntactically (at least). Below, we will see the notion of the indexical generalization of a sentence enters the discussion, as having helped Russell in [72] on the road to formulating the Limited General Barrier Theorem, but then it is only indexical terms that participate in the generalizing involved – since this involves universally quantifying the positions they occupy – so *here*, as $\mathfrak{h}$, gets to be generalized away (along with $\mathfrak{i}$ but not *now*, or for that matter *actually*, since these get the operator treatment.

Intermission. Evidently, this is connected with the listing (on p. 163) of $\mathcal{A}$ and $\mathcal{N}$ as context-sensitive operators, a set regarded as disjoint from the set of tense operators ($\mathcal{F}$, $\mathcal{P}$, $\mathcal{G}$, $\mathcal{H}$) despite the traditional categorization of tense as a species of indexicality or context-sensitivity, to say nothing of the modal operators, which would not traditionally be regarded as indexical. Worlds and times get special treatment – a feature going back to Kaplan's work – in being the contextual parameters which take us, when taken as circumstantial parameters from truth in M relative to t and w to truth in M *simpliciter*.[81] Whether, in a

[79] On p. 167, *Barriers* has the $m + n$-tuple enclosed in round brackets (parentheses) but I will follow the usage it established in this connection 100 pages earlier in the book and use angle brackets for our $(1 + 1)$-tuples (alias ordered pairs)

[80] In the case of w_C and t_C, when they appeared, respectively, as @ and n – or more explicitly, for a given model M, as $@_M$ and n_M.

[81] It may be helpful to quote *Barriers*, from p. 173, just before the definition – given below – of the 'partial context-shift' relation: "Essentially, the indexical sentences we will be looking at are the ones which are sensitive to changes in the parts of the context *that are not also parts of the circumstances of evaluation.*"

theory of context-sensitivity, times belong with places or with worlds or all belong together (and perhaps with agents also) seems to be a subject on which theorists feel strongly but differently: Kaplan's and Lewis's approaches are occasionally contrasted in *Barriers* – e.g., in note 25 on p. 173 and the text to which it is appended – and there is Evans [22],[82] raising an eyebrow at the tense–modality parallel, to say nothing of the ruminations in Chapters 2–4 in Richard [66], let alone the participants in debates over contextualism in epistemology. Clearly, the area is one in which those of us without strong commitments need to venture opinions somewhat tentatively. ***End of Intermission.***

The complication alluded to is that in this case, distractingly from the main message, we do not have the formal parallel to supply, since here has been given a nominal rather than adverbial treatment (as the p(lace)-term $\mathfrak{h}$) so the example making the case for a difference in treatment has to be given purely informally, in the middle of p. 170:

> There is no similar argument from a translation of it is raining to it is raining here. This seems right; if it is raining in Chicago, it is raining – even if it is not raining here (in Carrboro). So *it is raining* does not entail *it is raining here*. (But if it is raining at noon tomorrow, it doesn't follow that it is raining – for that it would need to be raining now.)

To get this 'main message' across more efficiently, a complete syntactic parallel would have been desirable – even if the missing operator had to live alongside the term $\mathfrak{h}$. (Naturally, some imaginative philosophical moves have involved usurping the term role of $\mathfrak{h}$'s companion $\mathfrak{i}$ with a similarly sentence-adverbial role: Prior [62] comes to mind.[83]) The passage quoted above raises an additional issue worth pondering, and that is over the extent to which Russell's project in this Context-sensitivity chapter of *Barriers* takes itself to be answerable to the empirical linguistic facts, whether concerning English in particular or natural language in general. We might wonder what would become of the above passage, for example, if it were translated into a language with spatial tenses – in particular of the part saying that *it is raining* does not entail *it is raining here*?[84] "What language with spatial tenses? There is no such language," you may reply. It may turn out that no natural language has enough of a spatial analogue of what we normally think of as tense to merit that description, though artificial examples certainly exist,[85] and even if they didn't, should that conspicuously contingent fact infect an account of context-dependence *per se*?

Another aspect of the discussion raises a similar worry, and that is the Kaplanesque absence of any discussion of linguistic devices for re-setting the contextual parameters. The last time the reviewer felt the need to point to the counterevidence against taking the ban on 'monsters' as a linguistic universal (let alone an *a priori* datum) was note 36, emanating from p. 260 of [40], referring to one of of the then few cases of published pushback against Kaplan on this matter: Schlenker [77]; since then numerous further voices have sounded a similar warning, and much more has been written on the subject, as may be gleaned by consulting Deal [16], on the empirical front, and Rabern and Ball [64], on the theoretical side of things, as well as the references supplied by both papers.

(Italics as in the source.)

[82] A point of special interest to readers of this review, perhaps: the 'Toomas Kaarmo' thanked in Evans' initial acknowledgments ([22], p. 343) is, the surname spelling notwithstanding, is the Toomas Karmo of Karmo [50] discussed in Part I above, and whose own indexicals-oriented paper links "I" with "here" to throw some light on personal identity puzzles in [49].

[83] And perhaps even at a pinch, though the reviewer has no first-hand acquaintance with the source, Lichtenberg's reaction to Descartes' *Cogito*.

[84] The translation of "It is raining," we may suppose, would obligatorily register in its verb morphology a rough indicator of the distance – and perhaps also, if that indication was more than negligible, the orientation from the point of utterance, of the putative rain.

[85] Lojban is perhaps the most well-known case of such a language; see [13], §§10.2, 3.

Somewhat ironically, even within the history of (some dialects of) English there may be traces of the phenomenon so strongly resisted by Kaplanian teratophobia, one of them showing up entirely incidentally in the course of the discussion of Kaplan in *Barriers*. On p. 163 there, we find:

> Kaplan repurposes $\mathcal{G}$ as an operator whose informal meaning is *one day ago* ([48], p. 542) but in the interests of inter-temporal understanding, I will continue to use it with Prior's meaning: *at all future times.*

Barriers does not raise an eyebrow at Kaplan's classification on the page in question of *one day ago* as a tense operator lying outside the demonstrative (i.e., indexical) vocabulary. For those with whom the reviewer interacts, this is no less indexical than *Yesterday*, placed by Kaplan inside that vocabulary, and if you want it to track the same day when embedded under past or future tense operators you need to use "the previous day" or "the day before." That is, in this dialect, what is communicated by uttering "Tom was then released, after being arrested for armed robbery three years ago," is that the arrest occurred three years before the time of the utterance rather than three years before the time of the release. (Of course you could relate past events in this way: "By the first of July, 1975, Margot had spent three cold nights lost in the Andes, and she woke up to the realisation that she was no closer to being rescued than she had been one day ago." This is, however, free indirect speech – *style indirecte libre*, if you prefer – in which, in this respect, anything goes: for instance one could put "yesterday" in place of "one day ago," or insert "now" before "no closer."[86]) Setting aside such curiosities, let us return to the main agenda.

The discussion of a possible barrier along the lines of "no indexical conclusions from non-indexical premises" begins (§7.2, p. 177) with examples, some of them recalling the above aspects of the favoured logical treatment of $\mathcal{A}$ and $\mathcal{N}$:

$$p \models \mathcal{A}p \qquad p \models \mathcal{N}p \qquad \varnothing \models \mathcal{N}Loc(\mathfrak{i}, \mathfrak{h}) \qquad \varnothing \models \mathcal{A}p \leftrightarrow p$$

They are all reasonable examples to bring to attention at this initial stage, but the comment that follows them, which echoes similar remarks elsewhere in *Barriers*,[87] is not: "Each has at least one indexical expression in the conclusion, but none in the premises." With classical logic in the background, it is never a sufficient condition for placing a sentence into the premise class or the conclusion of a candidate barrier thesis that the sentence contains one of a selected set of expressions, since these classes are plausibly required in *Barriers* to be closed under logical equivalence, and for any expression e, every sentence is classically equivalent to an sentence in which e occurs.[88] The point does not bear particularly on the first two examples, and the fourth is essentially the first $\models$-statement combined with its converse. The third is interesting since it will immediately make some readers wonder about the sign on the door, or the answering machine message, or mobile/cell phone, "I am not here now," presenting the recipient apparently with the negation of an indexical-logical truth, and leaving them disappointed to find no bibliographical entry for Predelli [60].[89]

[86] A similar example: *tomorrow* in the set phase *jam tomorrow.*

[87] For example in Remarks 2.1 and 2.2 on p. 72.

[88] This is even true (and well known to be) for the Anderson–Belnap relevant logic R, taking ψ being a consequence of $\phi_1 \ldots \phi_n$ in this case (for definiteness) to amount to the provability of $(\phi_1 \cdots \wedge \cdots \wedge \phi_n) \rightarrow \psi$. References and discussion of the issues involved here are provided in Humberstone [44].

[89] But the reviewer was pleased to see such an entry in Russell [75]. Rather than getting into the debate over the state of "I am here now," what we have, near the end of §3 of the paper, is a general reflection: "But ultimately it does not matter very much for the thesis that there are contingent logical truths whether that sentence turns out to be one of them. Arguing over whether it is a contingent logical truth is analogous to arguing over whether the law of excluded middle is an ordinary logical truth; even if it is not, that does little to undermine the thesis that there are logical truths, we are merely arguing over the extension. We did not come to see that there are contingent logical truths by generalizing from particular contested instances.

Other examples discussed in these early sections of Chapter 7 also involve $\mathfrak{i}$; the following pair appear on pp. 169 and 171 respectively:

$$B\mathfrak{i} \models \exists x x = \mathfrak{i} \qquad\qquad \forall x Bx \models B\mathfrak{i}.$$

The second of these, *Barriers* includes among apparently recording valid inferences from the non-indexical to the indexical, and the first, as displaying the form of the *Cogito* inference, if Bx is taken to represent x *thinks*. Ideally one might say a little more about the role of the premise, since the conclusion is valid outright on the present account, whereas the premise seemed crucial in Descartes' establishing the conclusion to his satisfaction. (A fuller story would then presumably require the epistemic dimension alluded to in note 25 above.) The first of these two, though valid by the lights of the semantics on offer, presents itself as a potential counterexample to any indexical barrier thesis and we return to it in a moment, after looking at the details of the barrier proposed in *Barriers* at this point.

As usual, to sort things out here, we need a suitable intermodel relation and "c-shift" abbreviating "partial content-shift" is what it's going to be called, with the following definition (p. 173), for which we need to recall that our current models look like this: $\langle W, T, D, P, C, I\rangle$, with the context component C itself having the internal structure $\langle a_C, p_C, t_C, w_C\rangle$. One of these, N is defined to be a *c-shift* of another, M, just in case:

$W_M = W_N$, $T_M = T_N$, $D_M = D_N$, $P_M = P_N$, $I_M = I_N$ and —
within C_M and C_N: $t_{C_M} = t_{C_N}$ and $w_{C_M} = w_{C_N}$.

In other words, what's allowed to shift in a partial context shift are just the agent parameter and the place parameter (a and p): so either or both of $\mathfrak{i}$, $\mathfrak{h}$ can change denotations (within the unchanging D and P).

On p. 156 of Russell [72] sentences were defined to be *constant* if they were c-shift-preserved and *indexical* if they were not. Note that this, already back in 2011, has the conclusion class of the would-be barrier blocking indexical conclusions being entailed by non-indexical (or 'constant') premises characterized by *breakability* rather than *fragility* – which is what allows us to treat constant and non-indexical as synonyms. With the premise class and the conclusion class for the barrier being each other's complements, life becomes simpler since there is now a single class of sentences that the barrier claims can't be got into from outside of it by a deductive transition: the class of indexical sentences. A side-effect of this simplification, though, is that ("Prior-style") cases like $B\mathfrak{i} \vee Cj$, say, where both B and C are monadic predicate letters and j is an individual constant, which, with fragility replaced by breakability are now moved into the indexical class and present something of a problem since the valid inference from the non-indexical ('constant') Cj to the indexical $B\mathfrak{i} \vee Cj$ (because c-shift breakable) violates the no constant-to-indexicals barrier – in the absence of special measures such as the *unless*-clause of the Limited General Barrier Theorem which originated in thinking about the present case. (Our disjunction is not c-shift *fragile*, because the Cj-verifying models can't be c-shifted to models falsifying the disjunction.) Russell notices that although the conclusion includes $\mathfrak{i}$ and potentially depends on who the speaker – or more generally, the agent of the context – happens to be if the premise is true, since it entails what is called the *indexical generalization* in which the position (or more generally positions) occupied by $\mathfrak{i}$ are universally quantified away. That is, we have: $Cj \models \forall x(Bx \vee Cj)$. (So we could equally well call it a 'de-indexicalizing generalization'.) Now we know that these (Prior-like) disjunctions invite all sorts of responses – maybe the barrier needs softening for some 'mixed category' cases, and so on – so *Barriers* (and Russell [72], originally) focuses on a case not inviting any such reactions: the case (already encountered) of $\forall x Bx \models B\mathfrak{i}$.

It is the Kaplanian picture of indexicality itself that makes space for them; once a sentence can express different propositions in different contexts, it is no longer forced to inherit the robustness of its truth from the proposition it expresses. Propositions are the primary bearers of truth, but not of logical truth."

Here there is no messing around with mixed cases: the premise is straightforwardly non-indexical and the conclusion indexical. (There could be, but in the present object there are not, context-sensitive atomic predicates, so no distractions over "B" arise on that front.) But here again, it doesn't matter who the agent of the context is, and even more obviously, since the premise already *is*, the indexical generalization of the conclusion and so certainly entails it. So far, so good, then, for the *Barrier* principle with an escape clause: constant-to-indexical inferences are invalid unless the premises entail not only the given indexical conclusion *but also its indexical generalization.* There may be several – not just indexical tokens but – *types* (e.g., would have occurrences, perhaps more than one of each, of both $\mathfrak{i}$ and $\mathfrak{h}$) together in play, and there may be several premises too.

This raises the question of how to abstract from the rather syntactic details of indexical generalization – find the indexical terms, put variables in their place, then prefix universal quantifiers binding all these variables – and of this solution of formulating an appropriate barrier principle which works across the board to give a promising schema for barrier principles. So, what is it about the fact that the conclusion's indexical generalization follows from the premise(s) that is doing the work? It can be expressed model-theoretically by saying that not only is every model of the premises a model of the original indexical conclusion – the inference is valid – but the inference 'over-shoots' in that respect of validity, in having every other model obtained by c-shifting from that one also verifying the conclusion as well. Now, deleting the references to c-shifting and indexicality and replacing them with arbitrary $\mathcal{R}$ and $\mathcal{R}$-breakability, get the Limited General Barrier Theorem's *unless*-clause, "unless for all models M of Γ *all models N such that $M\mathcal{R}N$ are such that* $N \models \phi$," with the 'over-shooting' aspect italicized here. (This last part is represented in Appendix C more succinctly by writing, in an enhanced modal-style language: $M \models \blacksquare\phi$.[90]) So this move lifts the *unless* clause from its local incarnation – no cross-barrier transitions from constant premises to indexical conclusions *unless the indexical generalization of the conclusion also follows from those premises* – to something generally applicable, rather than specific to the context-sensitivity case.

This concludes our discussion of the treatment of modality and of context-sensitivity in *Barriers*. From the former we will need to return to the a-switching relation, and from the latter to the c-shifting relation, in Appendix B, where we will connect them with differing accounts of aboutness currently on the market. Numerous issues on the context-sensitivity front, in particular, deserve a more detailed discussion than has been provided here: which indexicals deserve what kind of treatment – a place for their interpretations among the model's circumstantial parameters or otherwise, a right not to have their interpretations varied in the course of 'semantically unpicking' sentences containing them, and so on. The main point has been to convey the general idea, and to summarize the 'watershed role' of the treatment of inferences like that from $\forall x(Bx)$ to $B\mathfrak{i}$ in the work leading to *Barriers*' favoured treatment of its subject matter.

It would, however, be a shame to finish the main body of our discussion of *Barriers* without citing (and commenting briefly on) the memorable initial Dedication at the start of the book, which is at least as striking as anything one might hope to find in a dictionary of quotations, expressing a sentiment almost all of us must have felt often enough, without articulating it so precisely:

> *For everyone who wishes they could be finished with what they have to do, so they can get on with what they ought to be doing.*

Of course, what makes it striking is the unexpected contrast between *have to* and *ought*, so, deontic logic having been on the book's agenda, we can pause for a moment to comment on

[90] In fact, to avoid confusion, we use a different verification turnstile and write "$M \models\!\!\!\equiv \blacksquare\phi$", where $\blacksquare$ has $\mathcal{R}$ as its accessibility relation.

this.

The bulk of the attention on alternatives among deontic uses of modal auxiliaries has been on the *ought (to)/should*–vs.–*must* contrast, as in Jones and Pörn [46], and Björnsson and Shanklin [8]. But there has also been some attention to the semantics of *have to*, as well including in the latter paper, and Rubinstein [69], as well as several mentioned in Humberstone ([42], fn. 203). (The assumption made here that there is nothing to choose between *should* and *ought to*, is held to be an oversimplification in Myhill [56]; whether or not that's correct, I have certainly oversimplified in another respect: Russell's dedication doesn't simply contrast *have to do* with *ought to do*, but instead with *ought to be doing*.) Whatever the best account of these niceties ends up being, readers of *Barriers* with any judgment at all will agree in thinking that, when engaged in working on it, Russell was definitely doing what she ought to have been doing.

Errata

There will almost certainly be more typos, errors and infelicities in the present review than there are in *Barriers*, but the latter has a considerably greater chance of seeing a second edition at some future date, for which reason it seems worth including this list.

Main Text.

On p. 6, there is a typo already addressed in note 14 of our discussion.

On p. 51, example (1.11): in the formal rendering, we need to add as a further disjunct "$x = y$" in the consequent (or add "$x \neq y$" as a further conjunct in the antecedent).

On p. 70, Definition $\Subset$-fragile (already mentioned: note 40): should end with "$V(\phi) = 0$ for some N such that $M \Subset N$" rather than "$V(\phi) = 1$ for some N such that $M \Subset N$"

On p. 79 in the first line of the last paragraph under the subsection headed 'Prior's dilemma', the two occurrences of $\forall xFx$ should be $\forall xGx$, as in the earlier lines of that subsection.

On p. 85, what appear in note 1 here as H and G are on the following page introduced as $\mathcal{H}$ and $\mathcal{G}$. (Similarly, for "F" in the third line from the base of the main text on p. 98, read "$\mathcal{F}$".)

On p. 88, the putative definition of "anti-symmetry" is, rather, the definition of *asymmetry*, which is perhaps what is intended by the term in the paragraph preceding the definitions. (The correct definition for antisymmetry is given in lines 2–3 of §9.4 on p. 204.) This mistake is becoming more common, appearing, for instance on p. 7 of Bliss and Priest [9].

For p. 106: see note 51.

On p. 112, third line of 'Refinements elsewhere': "co-extensibility" should be "coextensionality" or "coextensiveness" (with hyphens if desired); this was mentioned above in note 51.

On p. 118, Proposition 4.4: "ϕ-fragile" at the end should read "R-fragile" (the notation used in *Barriers* for "$\mathcal{R}$-fragile").

On p. 129, in the sixth line of the paragraph starting "When we", the second occurrence of "@" should be deleted. (Alternatively, leave this occurrence intact and change "which" to "and".)

On p. 133, two lines after the first Definition on the page: "to distinguish Partly from Fully Modal" should read "to distinguish Partly from Purely Modal." In the second definition after that (of *Purely Modal*) "m-switch constructive" should read "m-switch constructible."

On p. 153, second last line of note 2, for "Aurthor" read "Arthur".

On p. 164, line 3 up: for "an sequence" read "a sequence".

On p. 171, three lines up, near the end of the line, "$\forall xFx$" should be "$\forall xBx$".

On p. 174, in the last paragraph before §7.5, Ap and Np are intended to read $\mathcal{A}p$ and $\mathcal{N}p$.

For p. 175: in the first list of four items, delete the words "isn't valid" from the end of item (3).

On p. 191, what is called the *Limited General Barrier Theorem* everywhere else gets called the *General Limited Barrier Theorem* in the proof of Thm. 8.14 here. 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+).

For p. 197: Prior's tense operators G, H, F, P for future-oriented and past-oriented Box-operators and their respective $\Diamond$-operators, are rendered as $\mathcal{G}, \mathcal{H}, \mathcal{F}, \mathcal{P}$ when they appear in Chapter 3 (see esp. pp. 86 and 89, and, before that, n. 3 on p. 23, as well as later, on p. 163 in the Indexical Logic Chapter 7 – "Context-Sensitivity") but in the very useful Chapter 9 ("All the Barriers") this $\mathcal{P}$ has become $\mathcal{E}$ on p. 197 (as well as p. 199). Perhaps this change was because of the awkward proximity in that discussion of the identically notated deontic permissibility operator $\mathcal{P}$, now that the various cases are being considered and compared in a single discussion. But some notification probably needed to be supplied. (I will have to beg forgiveness if there is after all some point I missed at which an explanation was offered. An index of symbols at the back of the book would have been a nice touch, enabling one to locate at least their initial occurrences or their definitions. The list of symbols mentioned in note 7, welcome though it is, does not include page numbers to lead the reader to the points where they are introduced.)

On pp. 200, 202, 204 (at least) there is something of an issue about "$\sqsubseteq$" and "$\sqsubset$" (as well as the terminology of m-extension and w-extension, remarked on in note 68.

p. 202 the final "s" in "each worldslices" needs to go (and one might like to see either a hyphen or a space after "world" here – it has a rather German look about it, without).

On p. 226, line 4, a hyphen has found its way into "proper subset" (between the two words, rather than, as once used to be the case, in the middle of "subset"); if the parenthetical comment that follows this occurrence, "probably a very proper subset," is intended as a joke, then that hyphen rather spoils the effect.

On p. 234, top line,"when when" needs to be just "when".

On p. 265, in line 7, "there there" can be made sense of (especially if a comma in inserted between these two occurrences, so the first is interpretable as "in that section'[' and the second as part of the *there is* construction) but is rather distracting, since one *there* would have sufficed.

On p. 283, final paragraph, "The Fallacy behind the Fallacies" needs its second "the" to be removed; this title, of a well-known paper by G. J. Massey, appears correctly in the bibliography of *Barriers*.

On p. 287, "conclusion-true," eight lines up the page, needs a space in place of the hyphen.

Bibliography.

On p. 293, the opening page of the bibliography, Ball (1989) and Brown (2015) are sourced as appearing in *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, whereas the words following the colon are usually regarded as part of a clarificatory subtitle, rather as part of the title proper for bibliographical purposes, as indeed they are treated in other entries in the bibliography of *Barriers* – such as Chilovi and Wodak (2021), Davies and Humberstone (1980), on this same opening page, of the bibliography (as are all subsequent *Phil. Studies* references).

Page numbers are missing from the following entries: Brown (2015) – where the volume number is also missing – Kripke (1963), Restall and Russell (2010), Tarski (1986).

If Charles Pigden is going to be Pigden, C. R., for his authored articles, he should probably also be Pigden, C. R., (rather than Pigden, C.), for his edited collection (Pigden 2010), so that all publications by the same person appear in chronological order.

On p. 295, in the title of Sher (2018), "Tarksi" for "Tarski" (at the first occurrence).

Index

It would have been helpful to include in the index, so that a reader encountering them can be reminded of where they were first introduced and with what definition, entries for the terms *constructible* and *anti-constructible*. I'm not sure what principle was used for index references under proper names. If one looks at an adjacent pair entries, Kaplan and Karmo, in the index, one finds the earliest page numbers cited for Kaplan as beginning with the stretch 155–7, whereas Kaplan has mentioned before the start of those already on pp. 6, 12, 35, 41, 57 and 151 (with many subsequent references also not in the index), and for Karmo we have three pages listed in the index, 5, 7, and 186, whereas he enters the discussion also on pp. 2 and 61, the latter in connection with a particularly important example, as, as it happens, not at all on p. 186.

Quasi-errata

The following may not count as errata in the sense of cases all would regard as incorrect:

From p. 76: "The set $\{Fa, \neg Fa\}$ is unsatisfiable, but none of its members is." *Neither* of its members is – I would have thought.

On p. 107: the term *constructible* is defined here for the first time, and then more generally (for arbitrary $\mathcal{R}$, that is) on p. 110. At least one of these should be entered into the index, where there is currently no entry for *constructible*/*constructibility* – much needed, given the numerous subsequent references, so that a reader can find a definition quickly.

From p. 129: "It might be true because p is true at the actual world, like in the model on the left in Fig. 5.4." *As in* the model on the left, surely?

On p. 187: The seldom encountered phrase "truth-functional matrix" might benefit from explanation or replacement – preferably the latter, if the intention is to mirror Quine's early usage of "matrix" for the quantifier-free part of a formula (esp. one in in prenex normal form – not highly relevant here since in the modal logics under discussion, not every formula has an equivalent with every truth-functional connective in the scope of a single outlying prefix of modal operators), abandoned (for other reasons) as early as 1950 even by Quine himself ([63], p. 90, n. 1).

On p. 241, somewhat out of the blue, *Barriers* puts a rhetorical question to the reader: "What happens to the truth-value of *All swans are white* if we keep the language the same but add a black swan to the world?" I felt that most readers familiar with the 17th century discovery – as it was from a European explorer's perspective – of black swans in (what is now) Western Australia, and perhaps especially readers aware that when *Barriers* was published, its author was actually based in Australia, might be somewhat puzzled by this. They could be excused for thinking: "What do you mean, *add* a black swan to the world? You could add another black swan if you really wanted to, but not a lot is going to happen to the truth-value of *All swans are white* if you do." I say "out of the blue" because the last time we heard mention of swans was on p. 84 in a discussion of inferences from the past observations of their colour to conclusions about future such observations (Hume's *other* barrier, as it is put in *Barriers*), and before that, on p. 17, where they figure prominently in an interesting argument about cases of (interworld) supervenience not giving rise to entailments. These earlier occurrences might help put a reader into the frame of mind of those presuming all swans to be white – but they are, as I say, quite a long way behind us in the text by the time that this rhetorical question is posed.

Appendix A: Aspects of the *Unless* Formulation

The formulation at issue is that appearing in the Limited General Barrier Theorem, with the *unless*-clause conveying the limitation in question; here we cite the version of this result given as Theorem 2 in Russell [74], since it uses the more convenient "is $\mathcal{R}$-preserved" in place of the wording of Theorem 8.13 *of Barriers* cited in Part I of our discussion I above; it has the incidental oddity of introducing the variable N not subsequently picked up anywhere in the statement of the result:[91]

8.13a If Γ is $\mathcal{R}$-preserved and ϕ is $\mathcal{R}$-breakable, then $\Gamma \not\models \phi$, unless all models of Γ are such that all $\mathcal{R}$-related models N are models of ϕ.

For the Hume's Law application this would then amount to: if the argument with non-normative – descriptive, factual,... but most importantly *s-shift preserved* – premise-set Γ and conclusion ϕ, then that conclusion is also non-normative, or else all models verifying the premises bear the relation $\mathcal{R}$ only to models verifying ϕ.

Now, the "unless" construction used here may not seem familiar in this kind of setting – especially as it appears in (8.13a) itself – because the word *unless* does not exactly wear its meaning on its sleeve, as we are about to illustrate (mainly in footnotes). In the end, we will be able to by-pass some of the niceties about exactly what this word means in English, concerning which there have, of course, been numerous proposals,[92] since it is more or less clear from the discussion in Russell's 2022 paper ([74]) as well as in *Barriers* (pp. 190,

[91] Though in [74] it is picked up in the proof, once we have a model M (of Γ) on our hands bearing the $\mathcal{R}$ relation to. There is no such gratuitous "N" in the formulation of Theorem 8.13 on p. 190 of *Barriers* itself.

[92] Several are described in pp. 964–975 of Humberstone [41], since the publication of which there has naturally been further work on the topic – such as Vostrikova [85]. [41], p. 967*f.* asks the reader to compare the closeness in meaning between "If you don't know the answer, you shouldn't put up your hand," and "Unless you know the answer, you shouldn't put up your hand," with the semantic distance between "If you don't know the answer, you didn't read Chapter 7," and "Unless you know the answer, you didn't read Chapter 7." Several striking examples from Michael Geis and from Samuel Fillenbaum are also given, such as this one (from Fillenbaum), for which we are to imagine a motorist trying to bribe a police officer: "If you don't give me a ticket, I'll give you $20," (plausible) vs. "Unless you give me a ticket, I'll give you $20" (ridiculous).

192) that – setting to one side certain rhetorical or pragmatic considerations – the truth-conditional aspect of her discussion it suffices to paraphrase "α unless β" as "α if not β" and to take the "if" here *materially*.[93] Thus one could equally well have:

8.13b If Γ is $\mathcal{R}$-preserved and ϕ is $\mathcal{R}$-breakable, then $\Gamma \not\models \phi$ or all models of Γ are such that all $\mathcal{R}$-related models N are models of ϕ.

One may prefer a less overtly negative formulation:

8.13c If Γ is $\mathcal{R}$-preserved and $\Gamma \models \phi$: then ϕ is $\mathcal{R}$-preserved, or each model $\mathcal{R}$-related to any model of Γ verifies ϕ.

Or again, getting rid of the disjunction on the right altogether

8.13d If Γ is $\mathcal{R}$-preserved and $\Gamma \models \phi$, and ϕ is $\mathcal{R}$-breakable, then each model $\mathcal{R}$-related to any model of Γ verifies ϕ.

Of course, another way we could get rid of the "or" from (8.13b, c) would be we could reinstate Russell's "unless" in at this point to replace the "or", thereby achieving the rhetorical effect (see notes 92, 93) – which is perhaps fair enough – of downplaying or marginalizing the possibility that follows "unless". It is in the form (8.13d) that this result is cited and given a proof – not that there is anything wrong with the proof in *Barriers* – in Example 4 of Appendix C (an example in the sense of an illustration of the convenience of the formalism on show there for presenting many of the proofs in *Barriers*). The disadvantage of reinstating "unless" is that the complications in the combined semantics and pragmatics of unless make some points difficult to see. And "or" in a context such as that of formulations (8.13a, b) is immediately going to be taken as the commutative connective of inclusive disjunction; this is what one gets by taking S_1 *unless* S_2 as S_1 *if not-*S_1, and then treating this as $\neg S_1 \to S_2$ and reasoning classically.

Those last three words are essential. If we work with intuitionistic logic, where, as in classical logic, $\phi \to \neg\psi$ and $\psi \to \neg\phi$ are equivalent to each other and to $\neg(\phi \wedge \psi)$, we know that negation on the *antecedent* is quite a different matter intuitionistically: neither $\neg\phi \to \psi$ nor $\neg\psi \to \phi$ has the other as a consequence and both are (in general strictly) weaker than $\phi \vee \psi$. Direction of conditionality seems to bear similarly on *unless*. Suppose that Ken wants to avoid seeing Barbie. Then one might say concerning some upcoming social event, "Ken will be there unless Barbie is." But only if it is Barbie that wants to avoid Ken that would one say "Barbie will be there unless Ken is." A fuller account would no doubt link this issue with those raised by the examples in notes 92 and 93. For the moment, let us just notice that simply reading *unless* as *or* resolves a few issues that might otherwise be puzzling. As written above 8.13c is perhaps punctuated in such a way as to raise worries about relative scope. Writing α for "Γ is $\mathcal{R}$-preserved and $\Gamma \models \phi$", "β for ϕ is $\mathcal{R}$-preserved," and γ for "each model $\mathcal{R}$-related to any model of Γ verifies ϕ," the question arises as to whether which of these is intended, in each of which the *if–then*... construction is symbolized by $\to$:

(1) $(\alpha \to \beta) \vee \gamma$ (2) $\alpha \to (\beta \vee \gamma)$

[93] Even when the truth conditions of *unless* and *if not* sentences coincide (assuming here that appropriate semantics-*vs.*-pragmatics warnings have dissuaded proponents of the *iff not* alternative), there remain differences in suggestiveness as to how safely dismissible some possibilities are. Compare *Unless you win the lottery, you won't be able to afford that holiday* with (the much odder) Unless you fail to win the lottery, you'll be able to take that luxury cruise. Sometimes the replacement of *if not* by *unless* even destroys the grammaticality of the sentence, as noted in Green ([29], p. 214*f.*): *John won't go if you don't go either* vs. * *John won't go unless you go either*, comparing this with the corresponding pair with *too* in place of *either*, for which both are grammatical.

A similar question might arise with 8.13b and 8.13a. But at least with some help if necessary, from a truth-table, we can see that (1) and (2) are classically equivalent. (Alternatively, rewrite $\alpha \to \beta$ as $\neg\alpha \vee \beta$ and appeal to the associativity of $\vee$.) Because of the semantic subtleties of *unless*, to say nothing of those of *if*, this is harder to see directly: we are in that case left pondering the relation between "(If α then β), unless γ" and "If α then (β unless γ)" – a rather trickier task.

We can look at another example of the cognitive simplification effected by bleaching the colour out of *unless* until all we are left with is *or* (and not even that if we proceed to the (8.13d) formulation). I have in mind the discussion on p. 191*f.* of *Barriers* as to whether the Limited General Barrier Theorem and its applications should be considered in some way *trivial*, which raises three worries falling under that rubric, the discussion culminating (p. 192) as follows – and we will get back to the other worries later; the reference to s-shifts registers the fact here the Limited General Barrier Theorem is being applied to the non-normative (s-shift-preserved)/normative (s-shift-breakable: we've seen the last of fragility) barrier:

> The third and final triviality worry was that the *unless*-clause in the theorem could be glossed as *unless it does*, which makes the claim uninformative. But the gloss is wrong. The theorem's *unless* clause doesn't just say *unless every model of* Γ *is a model of* ϕ. It specifies a stronger condition, requiring that all the s-shifts of a model of Γ model ϕ *as well.* To see why the stronger *unless*-clause makes a difference, note that Theorem 8.14 as a whole takes the form of a Not-Unless claim: Not-P, Unless Q. In Not-Unless claims, a stronger *unless*-clause (stronger Q) makes for a stronger (less trivial) barrier. Compare: I'm not going – unless you might need me?, vs I'm not going, unless the sky is falling. One of these RSVPers has a stronger "barrier" to going: the second one with the stronger *unless*-clause.

So *Barriers* is working quite hard here, to get a point about *unless* across to its readers. If I understand this point correctly, it doesn't apply specifically to claims of the form *not*-P *Unless* Q, but to any claim of the form P *unless* Q, and the point is that if we strengthen Q – replacing it by something, Q^+ say, logically stronger than Q, the resulting claim, P *unless* Q^+, is similarly stronger than the original. This is not quite true if "stronger" is taken literally, but if it is taken to mean "at least as strong as", then it amounts to saying that *unless* creates upward-entailing ('monotonic', 'positive', 'entailment-preserving') context after it, something very familiar in the case of inclusive disjunction (and not holding for exclusive disjunction): if Q^+ entails Q, then $P \vee Q^+$ entails $P \vee Q$. In other words, formulations with "or" is simpler than those with "unless" and already familiar, in respect of this issue.[94]

We pass to what *Barriers* describes as the second of the triviality worries on the agenda, arising out of an observation provided by a reader for the publishers (OUP), as we read in note 5, p. 191. The observation is that one of the conditions invoked in the statement of the

---

[94] On the other hand, perhaps I have misinterpreted the passage. One obvious difficulty is that Russell does not seem to have logical strength in mind with the example given, which should be cleaned up first, since we don't want the question mark, and I think we can drop the "might" too, so what is being said to be stronger than the proposition expressed in the context in question by "you (will) need me" is the proposition that the sky is falling, and what seems to be involved is the latter's greater improbability rather than anything about entailment. However, the entailment story fits the original case rather well: that all the s-shifts of any model M of Γ model [= verify] ϕ entails that M itself does (since the s-shift relation is reflexive). Surely this is what Russell had in mind in writing "It specifies a stronger condition," and not just any condition that's less likely to be satisfied. (However, this shows that the discussion in the quotation above from *Barriers* is less than optimal, because the point is supposed to be general and not limited to reflexive intermodel relations. The Limited General Barrier Theorems imposes no such conditions on $\mathcal{R}$, even if all correct claims about which sentences are $\mathcal{R}$-preserved remain correct when made about the reflexive closure of $\mathcal{R}$ – indeed when made about the transitive reflexive closure (= ancestral) of R: see Observation 6 in Appendix C.)

Limited General Barrier Theorem does not need to be there for the proof of the result to go through. All of what were described above as alternative formulations (i.e., 8.13a–d) of this result are not merely equivalent, but were deliberately chosen to as to retain this redundant element. For reach of those formulations, we use wavy underlining here to indicate the part of the formulation that can be omitted without jeopardizing the correctness of the claim made; we omit the "8.13" part of the headings:

(a/b) If Γ is $\mathcal{R}$-preserved and ϕ is $\mathcal{R}$-breakable, then $\Gamma \not\models \phi$, unless/or all models of Γ are such that all $\mathcal{R}$-related models N are models of ϕ.

(c) If Γ is $\mathcal{R}$-preserved and $\Gamma \models \phi$: then ϕ is $\mathcal{R}$-preserved, or each model $\mathcal{R}$-related to any model of Γ verifies ϕ.

(d) If Γ is $\mathcal{R}$-preserved and $\Gamma \models \phi$, and ϕ is $\mathcal{R}$-breakable, then each model $\mathcal{R}$-related to any model of Γ verifies ϕ.

Barriers repeats the underlined part of (a) in the proof of Theorem 8.13 on p. 191, but one can see that it plays no role in the argument. The formalized version of the proof in Appendix C below, appearing as Example 4, does this for the (d) formulation. Here, we make the point informally by means of a parallel example, for which purpose we concentrate on the (c) formulation, setting aside the underlined portion. For this we think of the quantification involved as being over people rather than models, collect the elements of Γ into their conjunction γ, and think, for this analogous case, of γ and ϕ as applying to those afflicted by two medical conditions, with $M\mathcal{R}N$ as meaning that M is a (biological) parent of N.[95] The holding of the verification relation $M \models \phi$ thus means that person M has the ϕ condition, while the consequence-statement $\Gamma \models \phi$ means that anyone with the γ-condition has the ϕ-condition – though we envisage that there may be many additional cases of the ϕ-condition in the non-γ-afflicted position, caused for example by exposure to environmental toxins. Now assume the antecedent of (c):

"Γ is $\mathcal{R}$-preserved and $\Gamma \models \phi$: the γ-condition is always inherited from parent to child, and everyone with the γ-condition has the ϕ-condition."

We want to derive, from this antecedent, the non-underlined part of the consequent of (c), namely

"Each model $\mathcal{R}$-related to any model of Γ verifies ϕ: any child of someone with the γ-condition has the ϕ-condition."

So suppose that N is the child of someone M with the γ-condition. Then, since the γ-condition is inherited, N has the γ-condition, and since everyone with the γ-condition has the ϕ-condition, N has the ϕ condition.

Clearly we did not need the underlined portion of (c), which says that (not only the γ-condition but also) the ϕ-condition is always inherited, and nor does its being inherited

[95] The γ-condition, as envisaged here, may represent an imaginary inheritance pattern rather than any that is actually attested: a deterministic version of being an autosomal dominant disorder, in that the reasoning based on (c) below requires that any child with either parent having the condition has a 100%, rather than 50% chance of inheriting the condition. The choice of an irreflexive $\mathcal{R}$ here – the *parent-of* relation – was made to reinforce the point in the parenthetical sentence at the end of note 94. Since this relation is also non-transitive, it makes for a useful illustration of the point (at the end of the proof of Observation 6 in Appendix C) that $\mathcal{R}$-preservation coincides with $\mathcal{R}^*$-preservation, where the latter is $\mathcal{R}$'s ancestral.

follow from what we did assume – though of course it *will* be – loosely speaking – inherited by the children (and indeed by all future generations) of γ-afflicted parents.[96]

What the etiquette of Theorem formulation requires in a situation like this, when one's previously favoured formulation of a result is discovered to contain a redundant condition is perhaps a moot point.[97] I incline to the view that a trimmed-down formulation avoiding the redundancy is really what is required for the initial statement, with such remarks about the weaker claim being added after (and not necessarily giving the latter the status of a Corollary, either). This forestalls the implicature that the result is the strongest relevant result known to the author/speaker. Instead, *Barriers* supplies the (it turns out, somewhat) bloated formulation and indicates why the condition not required for the proof to go through is one that makes appeal to the result especially salient. The point is made in a lively passage, in preparation for which let me mention that Flinders Street Station is a principal rail hub in the heart of Melbourne:

> But while strength is a virtue in a theorem, in everyday discussion – including in philosophical discussion – the virtues of informativeness and relevance often compete with that of strength. If you are heading in to Flinders Street from Burnley, all trains will take you to your destination. If you're coming back from Flinders Street, however, different trains go to different places, and so we tell visitors: "If you're coming back from the city, make sure to get on a train going to the right destination." Otherwise they could end up in Craigieburn! But of course, it's both stronger and true to say that *wherever* you are going, you should get on a train to the right destination. But who would tell visitors *that*? If you're not coming back from the city, there is no risk of being on the wrong train. We focus the warning where it is relevant. Returning to the theorem: with descriptive conclusions, there is no risk of s-shifting leading to the sentence having the wrong truth-value. With normative sentences, there is. So we say: no *normative* sentence follows from descriptive premises unless all s-shifts of models of the premises are models of the normative conclusion – even though the stronger claim is true, because the claim that makes specific reference to normative conclusions highlights the fact that these conclusions are s-shift breakable.[98]

Readers can make up their own minds whether this ingenious and feisty defence of the way *Barriers* presents the result on which the book eventually settles as the key to a unified treatment of barrier principles, or whether a more natural procedure would be to state the result in this form and explain that because it holds whether ϕ is $\mathcal{R}$-preserved or not – in the case of current interest – whether ϕ is descriptive or normative, the conclusion is of particular interest in salvaging a weakening of Hume's Law in the case in which ϕ is normative.

[96] "Loosely speaking," in the case of the ϕ-condition here, with the word *inherited*, because the connection between the γ- and ϕ-conditions need not itself be biologically based; but the point is that the children end up with not only the γ-conditions but also the ϕ-condition. The reference to future generations is a warm-up for the Appendix C issue mentioned in note 95, of passing from $\mathcal{R}$ to its ancestral, when what is at issue is preservation, the etymological source of "ancestral" being on stage in this example.

[97] In fact, the inclusion of this redundancy observation is the main feature that distinguishes the discussion of the Limited General Barrier Theorem in *Barriers* from the presentation in Russell [74], which certainly would have deserved saying if there had been any mention of the latter in *Barriers*.

[98] *Barriers*, p. 192. I particularly like the exclamation mark after "end up in Craigieburn" a full appreciation of which is enhanced by knowing that Craigieburn is not only a fair distance out, but is also one of Melbourne's less salubrious suburbs (for which reason its residents might not share in this appreciation). Note that this is Melbourne, Australia, which is perhaps worth saying since we were in the US with the passage quoted in Part III above from p. 170, including the words "if it is raining in Chicago, it is raining – even if it is not raining here (in Carrboro)" – Carrboro, North Carolina, that is. There is something appealing about a book in which the different chapters let the reader in on where they first were drafted by sprinkling in such geographical hints among the examples.

Appendix B: Aboutness in *Barriers*

Barriers opens (p. 1) with the informal articulation of five candidate barrier theses for consideration, two of which are sketched with the aid of the preposition *about*:

> The past/future barrier: no claims about the future from claims about the past.
> The is/must barrier: no claims about how things must be from claims about how things are.

In these and other cases, the chapters and sections of the book offer characterizations of what it is for a claim to be about X, for the different choices of X. This Appendix aims to bring out some linkages between aspects of Russell's discussion in *Barriers*, with (some parts of) the now extensive literature on what it is for a claim to be about X, *for variable* X.[99]

Let us recall the notion of a(ctual)-switch from Part III's summary, above, of Chapter 5 ('Modality') of *Barriers*, though here making the simplification often made in other parts of the book by taking the case in which the accessibility relation is the universal relation, so that models M for modal logic need not specify it explicitly. Adapting p. 133 of *Barriers* in this direction, models now take the form $\langle W_M, @_M, I_M\rangle$. With this simplification in force, N is defined to be an *a-switch* of M if and only if the following conditions – labelled with an additional "a" to recall their source in defining this relation:

$$(1_a)\ W_N = W_M \quad (2_a)\ @_N = @_M \quad (3_a)\text{ for all } p_i\text{, all } w \in W \setminus \{@\},\ I_N(p_i, w) = I_M(p_i, w).$$

As we also in Part III above, in play in *Barriers*' Chapter 6 ('Context-Sensitivity') were some rather elaborate models $\langle W, T, D, P, C, I\rangle$, in which C – the *context* component – itself had its own internal structure $\langle a_C, p_C, t_C, w_C\rangle$, and that between such models our attention was on the relation of partial context-shift which considered all models like the given model save perhaps in respect of which elements of their D and P domains were chosen as a_C and p_C (for the interpretation of the indexical terms $\mathfrak{i}$ and $\mathfrak{h}$ – representing I and *here*). For present purposes we can simplify away most of this and suppose that the only addition to a standard (non-modal, non-temporal) first-order language is the term $\mathfrak{i}$, so that models take the simpler form $\langle D, \boldsymbol{c}, I\rangle$ where "$\boldsymbol{c}$" is there to suggest *context*, since this is to replace a_C, $\mathfrak{i}$ now being the only item of primitive context-sensitive vocabulary in this language – a change of notation also being desirable here because the Chapter 7 "a" for agent-of-the-context now clashes with the "a" of "a(ctual)-switch" just recalled from Chapter 5. We can then stick to Chapter 7's term *c-shift* for context-shift to name what's left of that relation in the current simplified setting. One point deserves emphasis: the model-relative denotation of $\mathfrak{i}$ is not handled by I, but by [identity with] $\boldsymbol{c}$ directly, as in Chapter 7 (*mutatis mutandis*). As usual, we subscript the items involved to show which model they originate from. A model $N = \langle D_N, \boldsymbol{c}_N, I_N\rangle$ is a *c-shift* of a model $M = \langle D_M, \boldsymbol{c}_M, I_M\rangle$ if and only if the following two conditions are satisfied, the second of which we label with a "3" to match the a-switch conditions above, leaving a gap to indicate the absence of condition corresponding to (2_a) above:

$$(1_c)\ D_N = D_M \qquad\qquad (3_c)\ I_N = I_M.$$

The gap in question of course means that models M, N stand in the (symmetric) c-shift relation when they differ, if at all, only in that $\boldsymbol{c}_N \neq \boldsymbol{c}_M$.

[99] Recent surveys of the relevant literature: Hawke [32] and Giordani [27].

Thus, both the simplified modal models with their conditions (1_a), (2_a), (3_a) on the a-switch relation, and these simplified context-sensitivity respecting models with their conditions (1_c) and (3_c) on the c-shift relation do something similar: each of them allows changes only in respect of how its distinguished element is treated. The former insists it's the same distinguished element of their common domain (of worlds) in both models, but then in (3_a) allows atomic formulas to vary, though only insofar as which of them are true at the distinguished world. The latter allows instead the distinguished element, again drawn from their common domain (of individuals), to be chosen instead as a different element. This may sound, indeed, like a distinction without a difference, since – putting it somewhat impressionistically – there is really nothing more to being a given element than standing in the relations that element stands in to other elements. The difference will emerge when we have the a-switch and c-shift treatments compared on the same territory, rather than having one of them based on possible worlds models for propositional logic and the other on slightly unorthodox (because of the status of the $\boldsymbol{c}$ element) first-order models for predicate logic. We shall presently modify the possible worlds case in the direction of the context-sensitivity inspired case and find that both of them in that setting are alive and well among proposed accounts of what it is for a sentence to be *about* this or that individual.[100] Indeed, aboutness considerations are often informally raised in the discussion in *Barriers* which are suggestive of both accounts.

For the a-switch-based account, here is how *Barriers* introduces the non-modal vs. modal (as opposed to the 'modal strength') barrier idea in p. 130 of the book:

> On the B version of the modal barrier, the premise class contains sentences about the actual world, and the conclusion class sentences about "the rest of modal space".

And then after introducing m-switching, and just before introducing a-switching, we have a remark (italics as in the original), already quoted in part twice in our discussion:

> But that is to suggest that that $\Diamond p$ is a disjunction, one of whose disjuncts *is always about the actual world* (in logics where R is reflexive). No wonder then, if it turns out not to be m-switch fragile: affairs at the actual world can be sufficient to make such a disjunction true.

In short then, *Barriers* is offering a proposal as to what constitutes a sentence's being about the actual world, on the one hand, and its being about 'the rest of modal space' on the other. When we update the suggestion to be consilient with the post-fragility part of the book, ϕ's being about the actual world emerges as ϕ's being a-switch-breakable.

To find support in *Barriers* also for c-shift breakability as registering aboutness (ϕ's being about the distinguished element, that is, in our current distillation of the models for indexical logic) one might have to work a little harder, but we can certainly discern connections to Nelson Goodman's treatment of this topic, to which end let us first notice Russell making a similar move 'on the fly' in an area where indexicality is not at issue, in discussion of the Limited General Barrier Theorem's *unless* clause (*Barriers*, p. 192*f.*):

[100] Elaborating on note 37 above: if, instead of the diminished notion of context in play here as comprising only what is the fuller picture is the *agent* of the context ("a_C"), what survives in the diminished case is the *world* of the context ("w_C", formerly @), then the associated c-shift relation is the accessibility relation for the "Fixedly" operator of Crossley and Humberstone [14], and §2 of Chapter 5 in Davies [15], as well as Appendix 10 there. Its description in [14] as a 'global' operator rather than a modal operator reflects this model-changing semantic perspective – treated as a "$\blacksquare$" rather than "$\Box$" operator in a presentation along the lines considered in Appendix C below, with a distinctly metalinguistic flavour. But, as advised in note 37, see p. 257*ff.* in Humberstone [40] on reasons for thinking of this 'modal' (model-fixed) vs. 'global' (model-changing) distinction, in this case at least, as a by-product of the *presentation* of the semantics rather than a reflection of the underlying semantic ideas themselves.

> Folk logic tells us "you can't prove a universal". First-order logic tells us otherwise, for special cases. A standard version of Universal Introduction says that to prove $\forall x \phi(x)$ you have to be able to prove $\phi(a)$ *for arbitrary a*. We might think of it like this: the truth of Fa varies with reinterpretations of a. The truth of $\forall x F x$ doesn't. Can you get $\forall x F x$ from a set of premises? *Not unless* you can also get Fa no matter how a is interpreted (for all "interpretation shifts" of a.)

This is close to an analogue of indexical generalization for non-indexical terms and the role we saw that playing in Part C's summary of the evolution of the Limited General Barrier Theorem from the "unless you can also derive the indexical generalization of the conclusion" idea. It is at this point that the first aboutness proposal alluded to can make an entrance. This proposal can be attributed, without too much misrepresentation, to Goodman [28] – where there are various frills and complications we need not go into here.

Here is Goodman ([28], p. 4*f.*) saying the kind of thing we have been reading in *Barriers*, especially in the last two paragraphs of this quotation, though with his attention more specifically on aboutness *per se*, or more specifically on what he calls a statement's being *absolutely* about this or that thing – though the contrast he has in mind with relative aboutness (explained in a later section of [28]) is not something we need to attend to here.

> As a first attempt to explain absolute aboutness in terms of mention, we might say that a statement S is absolutely about Maine if some statement T that mentions Maine follows logically from S. But this will need a good deal of amendment. The statement
>
> (1) It either is or is not the case that Maine prospers
>
> mentions Maine without saying anything about it. Moreover, since (1) follows logically from any statement whatever, even from
>
> (2) Florida is Democratic,
>
> every statement would qualify by this criterion as being absolutely about Maine. The remedy that immediately suggests itself is to require that the T in question must not follow from logic alone, must not be logically true. But this will not suffice; for from (2) follows the conjunction of (1) and (2), which mentions Maine and is not logically true. Thus (2) – or any other sentence that is not logically true – will still qualify as absolutely about Maine.
>
> To exclude as a permissible T every statement that contains a logically true clause, or a 'non-essential' occurrence of an expression designating Maine, will accomplish very little. For this requirement and all others so far imposed will be met by another consequence of (2), namely
>
> (3) Maine or Florida is Democratic.
>
> Clearly some such statement mentioning Maine follows from any S that is not logically true; and every statement that is not logically true will thus satisfy all the stated conditions for being absolutely about Maine.
>
> Why, now, even though (3) – unlike (1), and unlike the conjunction of (1) and (2) – is surely about Maine, are we unwilling to accept the fact that (3) follows from (2) as showing that (2) is genuinely about Maine? Notice that we can put for "Maine" in (3) a name or description of anything (or even a universally quantified variable) and still have a statement that follows from (2). In other words, whatever (2) says about Maine it says about everything else as well. Now we must seriously raise the question whether a statement can properly be regarded as saying about any particular thing what it says about everything else. Or is a statement genuinely about Maine only if it says something about Maine that it does not say about something else?

Goodman's answer to that rhetorical question is *yes*, which is interpreted as meaning that – thinking of m as a constant denoting Maine, $\phi(m)$ is about Maine unless, with $\vdash$ for provability (or consequence) in classical predicate logic with identity, we have $\vdash \forall x(\phi(m) \to \phi(x))$. (Here $\phi(x)$ replaces all occurrences of m in $\phi(m)$ with x.) I have put it this way so as to recall the *unless* formulations in *Barriers*, though it would be even closer to the indexical generalization and "interpretation shift" formulations there, to say (equivalently) that $\phi(m)$ is about Maine unless $\vdash \phi(m) \to \forall x(\phi(x))$. Since the logic in use gives us the converse of that conditional automatically, this means that in $\forall x(\phi(x))$ we have an equivalent of $\phi(m)$ in which m does not appear, meaning that m occurs at most *inessentially* in $\phi(m)$, to pick up Goodman's earlier reference to non-essential occurrence.[101] Conversely, it is not hard to see that if m occurs in $\phi(m)$ at most inessentially, then we again have a valid implication from $\phi(m)$ to $\forall x(\phi(x))$. So the implicit account of aboutness in the remarks bearing on that relation and the c-shift relation in the context-sensitivity discussion in *Barriers* is the simple and natural account of aboutness one finds in Goodman [28].[102]

If the c-shift orientation is aligned to the account of aboutness associated with Goodman, what is the story with the a-switch idea? To answer this question, it will help, as already remarked, to have both approaches contrasted in the same linguistic territory, and can most straightforwardly do that by considering classical non-modal first-order logic, so we shall need to de-modalize the linguistic setting of the a-switch relation, and instead of having a the universe of our models be thought of as a set of possible worlds, one of which (namely @) has a distinguished status as part of the model, we will think of the universe as an ordinary domain of individuals, but with one of them anointed as the distinguished element of the model. Since this is what we had before with models $M = \langle D_M, \boldsymbol{c}_M, I_M\rangle$, we can use these same models, notated just like that, though think of $\boldsymbol{c}_M$ as the denotation of a particular individual c̳onstant c rather than as suggesting the word "context".

To keep the analogy with a-switching as close as possible, we want to mimic the conditions for the relevant intermodel relation called (1_a), (2_a) and (3_a) above, and we label them similarly, as $(1'_a)$ etc. But, what to call the relation in question? It would cruel to choose "c-switch" and inflict on the reader the task of remembering which is which out of c-shifting and c-switching – see the Lamentation in Part III above – so, because of its prominence in various publications with Robert Demolombe as the lead author which will be listed presently, let us just call it $\mathcal{R}_{\mathsf{Dem}}$. In the publications alluded to the authors consider the usual language of first-order logic *without identity* and in the case of special interest with a certainly kind of restricted quantifier as well as the usual unrestricted quantifiers and predicate letters of arbitrary arity, but again in order to keep the modal parallel with a-switching as close as possible we look at a language with only monadic predicate letters. That way, just as the sentence letters in the modal case are, in effect, assigned sets of worlds ('propositions') by the a model $M = \langle W_M, @_M, I_M\rangle$ – we omit any explicit accessibility relation – with the set assigned to p_i being, in the notation of *Barriers*, $\{w \in W_M \mid I_M(p_i, w) = 1\}$, so here a monadic predicate letter is assigned as its extension by I_M a subset of the domain. (We add to the conventions from the end of the 'Aside on Notation and Variations' in Part I above, a stipulation like that concerning the p_i, and suppose that our monadic predicate letters are given in some enumeration $F_1, \ldots, F_n, \ldots$.)

Parallelling the above then, we take one of these models, $M = \langle D_M, \boldsymbol{c}_M, I_M\rangle$ to bear the relation $\mathcal{R}_{\mathsf{Dem}}$ to another $N = \langle D_N, \boldsymbol{c}_N, I_N\rangle$, just in case:

$$(1'_a)\ D_N = D_M \quad (2'_a)\ \boldsymbol{c}_N = \boldsymbol{c}_M \quad (3'_a) \text{ for all } F_i, \text{ all } d \in D \smallsetminus \{\boldsymbol{c}\} : d \in I_N(F_i) \Leftrightarrow d \in I_M(F_i).$$

[101] $\exists x(\phi(x))$ would do equally well as an m-free equivalent of $\phi(m)$ under these conditions.

[102] Absolute aboutness, that is; we do not attend to Goodman's supplementary notions of immediate aboutness and relative aboutness in [28], taken up in Ullian [82] and elsewhere.

Thus, just as an actual-switch of a modal model does not really "switch" the distinguished world for another, but only changes the interpretations of sentence letters in respect of the membership in their extensions insofar as @ is concerned, keeping their interpretations alike for all other worlds, so in a Demolombe variant of a model the interpretations of predicate letters only change insofar as $\boldsymbol{c}$ is concerned, and not making any such changes for for other elements of the domain. Note the contrast with the c-shift relation ((1_c) and (3_c) above), which leaves the interpretations of the predicate letters intact but potentially swaps out the given $\boldsymbol{c}_M$ for a new $\boldsymbol{c}_N$.

While the parallel between ($1'_a$)–($3'_a$) and the original (1_a)–(3_a) is evident enough, the relationship between these sets of conditions in the two cases is of course rather different, since truth in a model here – which will have to be inductively defined via the ancillary device of an assignment of domain elements to the individual variables, anyway – and more significantly, there will be no singling out of a distinguished domain element (with truth in the model stipulated to be truth relative to that element). Such differences notwithstanding, these models and the intermodel relation here considered can be used to characterize a notion of aboutness. This was done, in particular, in Demolombe and Fariñas del Cerro [17], which will enable us to contrast the notion with in question with the Goodman-inspired notion.[103] Demolombe and Fariñas del Cerro [18] is quite close to the same authors' [17]. On the other hand the successor paper [19] takes things in a somewhat different direction, while the earlier paper Demolombe and Jones [20] concerns itself with a sentence's being about a subject matter rather than about an object (individual, entity. . .) in the world (an element of the domain, for the model-theoretic version), as Demolombe and Fariñas del Cerro observe in their joint work.[104]

As mentioned already, Demolombe and Fariñas del Cerro [17] do not restrict attention, as we do here (though not in the paper referred to in note 103), to monadic predicate logic, but they do restrict themselves to a language without the identity predicate, because of difficulties that do not concern us here, over its interpretation, despite including an expressive device that looks as though it features an occurrence of (negated) "=". These are the restricted quantifiers $\forall x_i \neq c$ and $\exists x_i \neq c$, where c is the distinguished constant. Syntactically, the constant c occurs in these restricted quantifiers, even though neither of what would normally be the predicates "=", "≠", does. Rather "≠" is an expression that takes an unrestricted quantifier and an individual constant to make a restricted quantifier. Specifically, for present purposes, that constant is the distinguished constant c and, semantically, the resulting quantifier ranges over $D \setminus \{\boldsymbol{c}\}$, so it ends up interpreted in exactly the way it would be if $\forall x \neq c(\phi(x))$ had been introduced as abbreviating $\forall x(x \neq c \to \phi(x))$, though in terms of these restricted quantifiers, the identity predicate and its complement cannot in turn be defined.

The reason for singling out just one individual constant for the special treatment is to explain aboutness in this simple setting by explaining what it is for ϕ, as Demolombe and Fariñas del Cerro [17] puts it, to be about c. That can't be quite right, though, since c is an individual constant rather than what that constant denotes. Nor is this difficulty easily rectified, since what it denotes depends on the model we are considering, and the aim is to give a model-theoretic account of a model-*independent* notion of aboutness – just as *Barriers* provides a model-theoretic account of normativity, particularity, etc., as model-independent rather than model-relative notions (by contrast with those mentioned in note 24 and the

[103] This way of contrasting them was developed independently of that particular apparatus, in §4 of Humberstone [44], just by considering variations on the theme of the papers by Demolombe and Fariñas del Cerro about to be cited. After working on [44], I came to *Barriers* and was intrigued to find aspects of both Goodman's approach to aboutness and that of Demolombe and Fariñas del Cerro deployed in different parts of that book, and the reason for including this appendix lies in the hope that others will also find this intriguing.

[104] For more on this distinction, see Chapter 2 of Yablo [92].

text to which it is appended). This problem did not arise for Goodman [28] with a single interpreted language – English or a somewhat regimented form thereof – in play, and one could flit without a care between saying that a sentence contained (essentially or otherwise) "Maine" and that the sentence mentioned Maine. To skirt around these difficulties, let's use "about*" to relate a sentence to a term to mean that the sentence is about the denotation of that term, in particular the individual constant c in our current 'toy example' of a language (with just the one distinguished constant) where this description – "the denotation of c" – floats from model to model, denoting, when any one model M is under consideration, the element $\boldsymbol{c}_M$.[105] We can do this for Goodman, too, saying that a sentence mentioning Maine is about* "Maine" when Goodman would simply say that the sentence was about Maine – which, as already noted, is the case just when "Maine" occurs essentially in the sentence.

As Demolombe and Fariñas del Cerro note, it is simpler to explain aboutness, or in the terminology just introduced, about*ness, if we work with the complementary notion "not about", which they abbreviate to *NAbout* (to which we append an asterisk as well as the prefix 'Demolombe'), and the proposal comes to the following, using the apparatus of *Barriers* and our discussion of it.[106] Recall that for simplicity we focus on the case in which the only candidate second argument for this binary relation is c; we do the same for the corresponding Goodmanian notion:

- ϕ is *Demolombe-NAbout** c iff ϕ is $\mathcal{R}_{\mathsf{Dem}}$-preserved
- ϕ is *Goodman-NAbout** c iff ϕ is c-shift preserved.

Of course, to put this into precisely the *Barriers*-favoured terminology, we would be saying "anti-fragile" or "unbreakable" in place of "preserved". In any case, defining the corresponding About*ness relations as the complementary relations, we get cases of breakability – or sensitivity, one could say since the intermodel relations are in both cases symmetric: ϕ's being about the denotation of c is a matter, for Demolombe and Fariñas del Cerro, of ϕ's truth-value in M being capable of being changed as we pass to a different N where only $\boldsymbol{c}$'s participation in – as the authors say – 'atomic facts' involving $\boldsymbol{c}$ is changed in this passage: can safely write "$\boldsymbol{c}$" since $\boldsymbol{c}_M = \boldsymbol{c}_N$.[107] On the other hand, Goodman-style aboutness is a matter of breakability in the face of c-shifting: changing *which individual* is under discussion. The question was alluded to, above, as to whether this makes any difference. The answer is that it makes a difference in both directions. In one direction: $\forall x(Fx)$ is Demolombe-about* c, since its truth in M with $I_M(F) = D_M$ turns to falsity on transition to $\mathcal{R}_{\mathsf{Dem}}$-related N with $I_N(F) = D_N \setminus \{\boldsymbol{c}_N\}$, but it is not Goodman-about* c, since any c-shift N of M just changes which element of D_M gets to be $\mathbf{c}_N$ from the same domain, all of whose elements are already in $I_M(F) = I_N(F) = D_N = D_M$. In the other direction: $\forall x \neq c(Fx)$ is Goodman-about* c, since this will go from true in M to false in N if $I_M(F) = D_M \setminus \{\boldsymbol{c}_M\}$ and $\boldsymbol{c}_N$ is any element of D_N other than $\boldsymbol{c}_M$ but not Demolombe-about* c. This example is mentioned in Demolombe and Fariñas del Cerro [17], though not as an example of where their notion of aboutness differs from Goodman's – just as an example in which c occurs in ϕ but ϕ is not (as we would say) Demolombe-about* c. In fact, they to not even mention it under that heading but they simply cite it at the end of the penultimate paragraph of §2 in [17] to note the contrast with the *unrestricted* quantifier case just repeated here as an example of Demolombe-aboutness without Goodman-aboutness. And even that case is

[105] This policy was that adopted in Humberstone [44].

[106] In Demolombe and Fariñas del Cerro [17] *NAbout* appears simply as *NA*; the fuller form is from their [18].

[107] You can see how a problem would arise if we had atomic *identity* facts to contend with, as we would with "=" in the language. This talk of atomic facts is a way of speaking about the upshot of $(3'_a)$ above, rather than a revival of logical atomism's facts as entities in their own right.

not described in such terms, but merely as a case in which ϕ can be (Demolombe-)about* c without c occurring in ϕ. Similarly, the final paragraph of §2 devotes itself to showing that c can occur in ϕ without ϕ's being (Demolombe-)about* c: but in all the examples given there, c occurs *inessentially* in ϕ, and for a comparison with Goodman's account, we need essential occurrence – as indeed we have in the case of ϕ as $\forall x \neq c(Fx)$ (or indeed ϕ as $\exists x \neq c(Fx)$).[108]

Both notions of aboutness (or about*ness) have their merits (as do some others, not here relevant) there is no need to see them as competing candidates for being *the* correct explication of the notion. And of course for each of them the Limited General Barrier Theorem has its application. By now the reader will be able to complete the relevant *Barriers*-style commentary on these cases: If the premises of a valid argument are collectively Goodman-about*/Demolombe-about* c, then so is the conclusion, unless...

Appendix C: The Modal Logic of Preservation

It is quite helpful, for streamlining some of the reasoning involved in thinking about preservation of truth from one model to another, to follow the pattern of Kripke semantics for modal logic but take things up a level, and introduce what one might think of as meta-models, macro-models or second-level models, though here let's just call them Models (with a capital "M"). Our models (lower case "m"), when they are deployed for intensional semantics, remain, as in *Barriers*, pointed models, truth in any one of which, M, is truth at its distinguished point $@_M$, though the remaining ingredients will depend on the particular application (first-order, deontic, etc.), and we take a *Model*, $\mathcal{M}$, to be a pair $(\mathbb{M}, \mathcal{R})$ in which $\mathbb{M}$ is a set of models of the similarity type (or 'signature') appropriate to any given such application and $\mathcal{R} \subseteq \mathbb{M} \times \mathbb{M}$. Unlike the ground-level models $M \in \mathbb{M}$, with their distinguished points, we will not be singling out a distinguished model from among these, and so work throughout with the doubly relativized notion "relative to $\mathcal{M}$, ϕ is true in M," or, more briefly: "$\mathcal{M},\ M \models \phi$," though the context will usually be relied on to supply the relevant $\mathcal{M}$. To help emphasize this difference between models and Models we use parentheses rather than angle-brackets to enclose their components in the latter case, writing "$(\mathbb{M}, \mathcal{R})$" rather than "$\langle \mathbb{M}, \mathcal{R} \rangle$".[109] As with "$\models$," we will use the same turnstile notation for the associated consequence relation, so that we have, where $\mathcal{M} = (\mathbb{M}, \mathcal{R})$:

Definition 1 $\Gamma \models_{\mathcal{M}} \psi$ iff for all $M \in \mathbb{M}$, if $\mathcal{M}, M \models \phi$ for every $\phi \in \Gamma$, then $\mathcal{M}, M \models \psi$.

Let us begin by illustrating the general idea by considering the language of predicate logic with a new 1-ary connective $\blacksquare$, which will be interpreted in a Model by having the intermodel relation $\mathcal{R}$, and will be written as $\blacksquare$ (with $\blacklozenge = \neg\blacksquare\neg$) to emphasize both the similarity of its logical behaviour with that of $\Box$ among and its difference of 'level' from $\Box$

[108] Recall the notations $\boxdot$ and $\diamonddot$ introduced in Part III to highlight the comment in *Barriers* that, especially in extensions of KT, $\Box\phi$ is partly about the actual world and partly about the rest of modal space, rendering the parts in question as the conjuncts ϕ and $\boxdot\phi$. (See note 75 and the text to which it is appended.) As $\boldsymbol{c}_M$ remains fixed in the evaluation, relative to M, of any sentence, this corresponds to $\boxdot$ as exempting the *actual* world rather than exempting the *current* world (in the sense of that note) from the universal quantification effected by $\Box$.

[109] As well as the evident further difference: for Models, there is no further interpretation or valuation component, such as in the case of (Kripke) models would be required to turn a frame into a model. This is because each model in a Model already comes equipped with an interpretation of the non-logical vocabulary (together its own accessibility relation(s), etc., for the case of any logical vocabulary requiring it).

or the like (with, when it is present – which it is not, in the current, non-modal predicate-logical case – its own *intramodel* accessibility relation).[110] Call the following language $L_\forall^{\blacksquare}$: this is to be the smallest set of formulas containing all closed formulas of the language predicate logic (or indeed any first-order language L with selected non-logical vocabulary) and satisfying the condition that for any ϕ, ψ in $L_\forall^{\blacksquare}$ we have $\phi \wedge \psi$, $\neg\phi$ and $\blacksquare\phi$ also in $L_\forall^{\blacksquare}$. We take any other Boolean connectives to be defined in terms of $\neg$ and $\wedge$ in the usual way, and note that the language is more restrictive than that of modal predicate logic (with $\Box$ rewritten as $\blacksquare$) in the following respect: free individual variables to not occur in the scope of the Box operator, and that operator does not occur within the scope of any quantifier.

To interpret $L_\forall^{\blacksquare}$ in a $\underline{\underline{\text{M}}}$odel, we piggy-back (in the first clause below) on the notion of truth in a $\underline{\underline{\text{m}}}$odel provided by the relevant chapter of *Barriers*, which for the present case is Chapter 2, on the Particular/Universal barrier, and the models in play here are those of that discussion. This means that we are thinking principally of $\mathcal{R}$ as the substructure relation $\sqsubseteq$, though we stick with the generic "$\mathcal{R}$" notation since we might equally well want to consider other intermodel relations instead (such as $\sqsupseteq$, or *is a homomorphic image of*, and so on). Thus, relative to a given $\mathcal{M} = (\mathbb{M}, \mathcal{R})$, for $M \in \mathbb{M}$ we have the following, where, more explicitly we would be writing "$\mathcal{M}, M \vDdash \phi$" instead of just "$M \vDdash \phi$":

$$
\begin{array}{llll}
\text{For all closed } \blacksquare\text{-free } \phi\text{:} & M \vDdash \phi & \text{iff} & M \models \phi. \\
\text{For any formulas } \phi, \psi\text{:} & M \vDdash \phi \wedge \psi & \text{iff} & M \vDdash \phi \text{ and } M \vDdash \psi; \\
 & M \vDdash \neg\phi & \text{iff} & M \not\vDdash \phi; \\
 & M \vDdash \blacksquare\phi & \text{iff} & \text{for all } N \in \mathbb{M}, \text{ if } M\mathcal{R}N \text{ then } N \vDdash \phi.
\end{array}
$$

There is an element of double definition going on here with the second and third clauses since if each of ϕ, ψ in the former case – and if ϕ in the latter case – is already a closed $\blacksquare$-free formula, its case has already been dealt with by the first clause (and the inductive definition of "$M \models$"). But this is harmless, in that there is no clash between the verdicts provided by the two routes to the truth-in-M conditions for these Boolean compounds. It is perhaps worth explicitly noting, in this connection, that for any sentence ϕ in which $\blacksquare$ (and the defined $\blacklozenge$) does not appear and any model $M \in \mathbb{M}$ for some Model $\mathcal{M} = (\mathbb{M}, \mathcal{R})$ relative to which the l.h.s. here is to be understood and with the r.h.s. exactly as in the main body of our discussion, we have:

$$M \models \phi \text{ if and only if } M \vDdash \phi.$$

The *raison d'être* of the present apparatus lies, of course, with the 'meta-modal' Box-operator $\blacksquare$, as treated in the final clause of the inductive definition of $\vDdash$ above, excluded from the sentences ϕ in the observation just made. With its aid we can say, now exploiting our enhanced object-language, that a formula ϕ is $\mathcal{R}$-preserved (or in the formulations preferred

[110] The idea of pursuing modal logic against a semantic backdrop in which the role usually played by points in a Kripke model – representing possible worlds, moments of time, etc. – is played instead by whole models (of this or that logic or theory) is not new. A famous case is one in which it is models of set theory that are involved – Hamkins and Löwe [30], relatedly, Esakia and Löwe [21], and references supplied in those papers. (In fact, Hamkins and Wołoszyn [31], which appeared after the present review – apart from this sentence – was completed, is perhaps even closer to the present material. Likewise with the somewhat earlier Saveliev and Shapirovsky [76].) In the case of Berger, Block and Löwe [7], these internal models are instead (models of the theory of) Abelian groups, providing a nice contrast with §3 of Garson [25] in which individual groups, Abelian and otherwise, play the role of (frames of) Kripke models rather than model-elements. A similar tack is taken in the textbook [83], pp. 151–153 (pp. 232–236 in the Nousoul Digital edition) with, however, no explicit comment on the connections between his transformation groupoids and groups – despite the inclusion of a section on groups in the opening 'mathematical preliminaries' chapter of the book. Garson's §2 is also relevant to this approach, and was taken up in more general model-theoretic work by Giangiacomo Gerla cited in two entries in the bibliography of Ghilardi [26], which itself takes the ideas further in that direction.

in *Barriers*: $\mathcal{R}$-anti-fragile, or $\mathcal{R}$-unbreakable) over a class $\mathcal{M}$ of models by means of the consequence statement on the left here, with that on the right to express the corresponding fragility claim:

$$\phi \models_{\mathcal{M}} \blacksquare\phi \qquad \text{(Preservation)} \qquad\qquad \phi \models_{\mathcal{M}} \blacklozenge\neg\phi \qquad \text{(Fragility)}$$

or indeed, omitting "$\varnothing$" on the left, by means of: $\models_{\mathcal{M}} \phi \to \blacksquare\phi$, for preservation, and similarly for the case of fragility.[111] And in the latter case, of course, one could write "$\neg\blacksquare$" if one prefers. What can't be expressed by means of $\models_{\mathcal{M}}$ (or the outright $\mathcal{M}$-valid formulas or sentences, i.e., those ϕ for which $\varnothing \models_{\mathcal{M}} \phi$), are the denials of such claims; for a given sentence ϕ to be breakable (or makeable) over $\mathcal{M} = (\mathbb{M}, \mathcal{R})$ is just for it to be the case that $\phi \not\models_{\mathcal{M}} \blacksquare\phi$ (or $\neg\phi \not\models_{\mathcal{M}} \blacksquare\neg\phi$, respectively, the latter being equally well put by saying that $\blacklozenge\phi \not\models \phi$). In what follows, we pay closest attention to formulations in terms of preservation.

Interlude on a Complication. The mention of what was called $L_{\forall}^{\blacksquare}$ above is meant to be illustrative rather than to force attention onto a single example of the $\blacksquare$-expansions of the languages and intermodel relations in play in *Barriers*. And the reason that the formation rules were given in a more restricted way than is customary was not that it would be impossible to make sense of applying $\blacksquare$ to open sentences, but that this would present a distraction in that it is not immediately clear what the best way of doing so would be. To illustrate the problematic feature of allowing cases of quantifying into $\blacksquare$-contexts, consider the (non-$L_{\forall}^{\blacksquare}$) sentence $\exists x(Fx \wedge \blacklozenge Gx)$,[112] and we are thinking of our intermodel relation as $\sqsupseteq$, wanting to evaluate it relative to a particular M in some $\mathbb{M}$. For it to be true in M (relative to $(\mathbb{M}, \sqsupseteq)$) we need to assign an object, presumably from D_M, to the variable x and then ask about that object's membership in $I_M(F)$ – so far, so good – and also in $I_N(F)$ for the various $N \in \mathbb{M}$ for which $N \sqsupseteq M$: but we may have lost the object in question on passing to N, concerning which to ask whether or not it belongs to $V_N(F)$. ***End of Interlude.***

Next, let us observe, regardless of what the universe $\mathbb{M}$ and the relation $\mathcal{R}$ of our Models $\mathcal{M}$ may be, we have a basic 'normality' condition familiar (with $\Box$ in place of $\blacksquare$) from conventional modal logic, satisfied by the (local) consequence relation of any normal modal logic, written in rule notation here as would be more appropriate if the consequence relation notation were replaced by a sequent-separator, as a marker of which we will at least remove the reference to $\mathcal{M}$, the claim being that the associated conditional holds regardless of how $\mathcal{M}$ is chosen:

$$\frac{\phi_1, \ldots, \phi_n \models \psi}{\blacksquare\,\phi_1 \ldots, \blacksquare\,\phi_n \models \blacksquare\,\psi} \qquad (\blacksquare\text{-Norm})$$

If we are pursuing modal logics as sets of formulas, this is the 'normality' rule, taking us from $\bigwedge_{i=1}^{n} \phi_i \to \psi$ to $\bigwedge_{i=1}^{n} \blacksquare\phi_i \to \blacksquare\psi$, where these conditionals are identified with their consequents when $n = 0$.

Before proceeding with further details, let us pause to see how anyone with a bit of a modal-logical background might find this apparatus helpful in spelling things out. As an

[111] Strictly, yes, it is the universe $\mathbb{M}$ that has the models as elements rather than the Model $\mathcal{M}$. The example just given, illustrating the 'Deduction-Detachment Theorem' property of $\to$ – namely: $\Gamma, \phi \models_{\mathcal{M}} \psi$ if and only if $\Gamma \models_{\mathcal{M}} \phi \to \psi$ – shows that $\models_{\mathcal{M}}$ for the case in which (the universe of) $\mathcal{M}$ is the class of all models under consideration is not what has been called the *model-consequence* or *global consequence* relation in modal logic, which holds between Γ and ψ when ψ is true throughout (i.e. at all points in) any model throughout which all formulas in Γ are true. This relation famously holds between ϕ and $\Box\phi$, giving rise to many counterexamples to the "only if" half of the DDT. The key difference, evident already in the "if" half of the above clause for $\neg$, is that here truth is truth in a pointed model rather than truth at every point in a model on the more egalitarian conception of what a model looks like for modal logic.

[112] Here F and G are monadic predicate letters.

initial example, take the distinction (in Part I) between collective and distributive readings of the attribution of properties to (esp. finite) sets of formulas, stressed in *Barriers*.

Example 2 For the property of being $\mathcal{R}$-preserved as applied to the representative case of a three-element set $\{\phi_1, \phi_2, \phi_3\}$, we have, expressed compactly as formulas:

Distributive reading:	$(\phi_1 \to \blacksquare\phi_1) \wedge (\phi_2 \to \blacksquare\phi_2) \wedge (\phi_3 \to \blacksquare\phi_3)$
Collective reading:	$(\phi_1 \wedge \phi_2 \wedge \phi_3) \to \blacksquare(\phi_1 \wedge \phi_2 \wedge \phi_3)$.

That the collective reading follows from the distributive reading is in this setting easily seen. From the distributive reading, we have as a truth-functional consequence:

$$(\phi_1 \wedge \phi_2 \wedge \phi_3) \to (\blacksquare\phi_1 \wedge \blacksquare\phi_2 \wedge \blacksquare\phi_3),$$

and the normality rule delivers

$$(\blacksquare\phi_1 \wedge \blacksquare\phi_2 \wedge \blacksquare\phi_3) \to \blacksquare(\phi_1 \wedge \phi_2 \wedge \phi_3).$$

Putting these together gives us the collective reading.

To see that the converse implication fails, note that the collective reading is always correct if any two – or indeed all three – of ϕ_1, ϕ_2, ϕ_3, are together though not individually inconsistent in which case that reading is vacuously true, and let $\mathcal{M}$ be collect some of the deontic–alethic models from *Barriers*' Chapter 8 ("Normativity"), discussed in Part I above, with $\mathcal{R}$ as the s-shifting relation. For example, choose $\phi_1 = \phi_2 = \mathcal{O}p$ and $\phi_3 = \neg\mathcal{O}p$ ($\mathcal{O}\neg p$ would do equally well here) so that the collective reading is vacuously true but the first conjunct of the distributive reading, $\mathcal{O}p \to \blacksquare\mathcal{O}p$ is not true in any $M \in \mathbb{M}$ with p true throughout S_M and not true in some $\mathcal{R}$-related N with S_N including some S_M-absent point at which p is false. $\lhd$

This last case, of the sentence $\mathcal{O}p \to \blacksquare\mathcal{O}p$ whose invalidity in deontic (or alethic-deontic) Models was just noted, provides us with the occasion to elaborate on a tricky point: our references to normality ($\blacksquare$-Norm, etc.) in all this pertain just to the behaviour of $\blacksquare$. The question arises as to whether we are dealing here with a normal (perhaps multi-)modal logic, which involves more than this. Set aside for the moment the fact that the language, if we stick with the restricted language in which we are tentatively working (no $\blacksquare$ in the scope of $\forall$, $\Box$, $\mathcal{O}$, ...) disallows unrestricted embedding. The most common definition of a normal modal logic will build in, not only the required normality constraint(s) on the $\Box$ operators in play – or let us, to reduce the risk of confusion, one the Box operators in play (including the $\Box$, $\mathcal{G}$, $\mathcal{O}$, etc. of *Barriers* and $\blacksquare$ too) and classicality conditions for the non-modal logical vocabulary, but also but also a substitution-related condition. The latter takes the form of a condition of substitution-invariance on the associated consequence relation (here any of the various $\models_{\mathcal{M}}$), if that is how the logic is being conceived, or, in case we are identifying logics with sets of formulas, that the set of formulas be closed under the rule of Uniform Substitution (or a predicate-logical analogue of this rule in the quantified case). In fact this is often taken as not only a precondition not just for being a normal modal logic, but for being a modal logic, and indeed even more fundamentally for being a logic at all. Other logicians make no such demands.[113] If we want to continue to speak of the present topic as

[113] Differences on this terminological front are illustrated in the second last paragraph on p. 88 of Williamson [91], where the fact that Chellas and Segerberg [12] explicitly refrain from requiring closure under Uniform Substitution is noted. (These last two papers are in our bibliography for a quite different reason below.) Numerous examples of what their proponents have regarded as logics but which are not substitution-invariant

the modal logic of preservation, we had better align ourselves with the latter, at least for the duration of this Appendix.

The relevant point is made clearly in *Barriers*, p. 139, note 1: the status of a sentence as $\mathcal{R}$-preserved (-fragile, or -breakable, etc.) is not automatically inherited by its substitution instances. For a simple example, consider the ■-enhancement of the language of Chapter 7's deontic-alethic language (with its primitive non-Boolean operators $\mathcal{O}$, $\Box$) and the associated models and truth-definition as summarized in Part I above, before and after the 'Aside on Notation and Variations' there, with $\mathcal{R}$ as the s-shift relation. Then for any Model $\mathcal{M} = (\mathbb{M}, \mathcal{R})$, with $\mathbb{M}$ housing the deontic–alethic models $M = \langle W, S, @, I\rangle$ in play there, we have $p \models\!\equiv_{\mathcal{M}} \blacksquare p$, since no shift of S_M to S_N of a new model N can affect the M-truth (the truth-value at $@_M$, that is) of a sentence free of deontic vocabulary, as the simplest available sentences, like p, are. But of course $\mathcal{O}p \not\models\!\equiv_{\mathcal{M}} \blacksquare\mathcal{O}p$: a failure of substitution invariance. In terms of the 'logics as sets of formulas' approach, the set of $\mathcal{M}$-valid formulas includes $p \to \blacksquare p$ but not, as we noted in the preceding paragraph, its substitution instance $\mathcal{O}p \to \blacksquare\mathcal{O}p$, violating the potentially contestable 'closure under Uniform Substitution' condition on normal modal logics.[114] (Since, as we began by noting, the restricted embedding aspects of such languages as $L^{\blacksquare}_{\forall}$, and what by analogy would be called $L^{\blacksquare}_{\mathcal{O}}$, $L^{\blacksquare}_{\Box}$,... means that the set of sentences is itself not closed under Uniform Substitution, we can make the point better by saying that, even for the cases in which ϕ has a well-formed substitution instance ϕ', ϕ's being in the logic does not imply that ϕ' is – as with the case just cited.)

The issue about Uniform Substitution may already have caught the reader's attention at the point at which ϕ's fragility w.r.t. $\mathcal{M}$ was introduced above in terms of the consequence statement $\phi \models\!\equiv_{\mathcal{M}} \blacklozenge\neg\phi$. This becomes more obvious if we concentrate on the set of $\mathcal{M}$-valid formulas, for $\mathcal{M} = (\mathbb{M}, \mathcal{R})$ with all sentence letters are all $\mathcal{R}$-fragile, such as the alethic modal logic models with the a-shift relation as $\mathcal{R}$.[115] This brings us face to face with the formula version, with ϕ as p: $p \to \blacklozenge\neg p$, increasing the chances of seeing that no consistent normal modal logic in the usual substitution-closed sense, can contain this formula. (Just substitute the nullary connective $\top$ – or the formula $p \to p$ – for p and appeal to Modus Ponens.)

Note on Two Technicalities. The failure of Uniform Substitution (or of substitution-invariance if we are thinking of the consequence relations $\models\!\equiv_{\mathcal{M}}$) just remarked on may prompt some to deny the this discussion deserves to be called the *logic* (modal or otherwise) of preservation, but let us consider here a couple of qualms about its being called the (modal) logic *of preservation* – to acknowledge the points in question rather than to change this way of putting things.

First Qualm. Alethic modal logic is thought of as the logic of operators representing necessity and possibility, and deontic logic as the logic of operators (= 1-ary connectives, for present purposes) representing obligatoriness and permissibility, and so on. But what we are calling the modal logic of preservation doesn't similarly present the logic of a preservation

can be found in the top third of p. 192 of Humberstone [41]; see also Holliday, Hoshi and Icard [35]. (One of the former examples has already been mentioned, though not by name: the logic $\mathsf{S5A}\mathcal{F}$ of – Necessity, Actuality and 'Fixity' discussed in the references in note 100, detailing the connection with ■.) The discussion in [41] also describes two proposed weakenings of substitution-invariance that have been proposed for maintaining different aspects of the original motivation: one, urged in that work, is that we require closure under uniform (propositional) variable-for-variable substitution, while another, urged by Gerhard Schurz ([79] and elsewhere) is the somewhat more restrictive condition of closure under invertible substitution. Thus, if logics are taken as sets of formulas in a sentential language for which # is a binary connective and $p \# (q \# p)$ is in a logic, on the former account we require that $p \# (r \# p)$, $q \# (r \# q)$, and $q \# (p \# q)$, for instance, are also in the logic, whereas on the latter account we require that these, and in addition $p \# (p \# p)$, $q \# (q \# q)$, etc., are in the logic too.

[114] Of course, instead of substituting $\mathcal{O}p$ for p, we could make the point by substituting for p any formula not guaranteed to be similarly insensitive to s-shifts, such as $\mathcal{P}(q \wedge r)$, or $q \vee \mathcal{O}p$.

[115] This is a special case of the point made in note 77.

connective. The point is not simply that $\blacksquare\phi$ does not, when interpreted relative to $\mathcal{M} = (\mathbb{M}, \mathcal{R})$, mean that ϕ is preserved on passage from one element of $\mathbb{M}$ to any other $\mathcal{R}$-related element. A necessary and sufficient condition for that to be the case is of course that the $\phi \models_{\mathcal{M}} \blacksquare\phi$, or equivalently that $\varnothing \models_{\mathcal{M}} \phi \to \blacksquare\phi$. But that does not mean that if we defined $\mathcal{P}$, say, by setting $\mathcal{P}\phi = \phi \to \blacksquare\phi$, that $\mathcal{P}$ would be, in the sense intended, a preservation connective.[116] If it were, then $\mathcal{M} \models \neg\mathcal{P}\phi$ would be a way to say that ϕ was not $\mathcal{R}$-preserved over $\mathcal{M}$, whereas what it says is something much stronger, namely that $\models_{\mathcal{M}} \phi \wedge \neg\blacksquare\phi$, or equivalently $\models_{\mathcal{M}} \phi$ and $\models_{\mathcal{M}} \neg\blacksquare\phi$. Not just stronger, but inconsistent, since we have the necessitation rule for $\blacksquare$ (i.e., the $n = 0$ case of the more general ($\blacksquare$-Norm) above, relative to any $\mathcal{M}$). Instead, to deny that ϕ is preserved one has keep the negation metalinguistic and say that $\phi \not\models_{\mathcal{M}} \blacksquare\phi$ (or $\not\models_{\mathcal{M}} \phi \to \blacksquare\phi$). One could consider remedying this. One way might be to add a new primitive "universal modality" $\blacksquare^+$, understanding $\mathcal{M}, M \models \blacksquare^+\phi$ to hold iff for all $N \in \mathbb{M}, \mathcal{M}, N \models \phi$, with $\mathcal{P}\phi$ now defined as (a Box-variegated version of the noncontingency connective mentioned at the end of note 116): $\blacksquare^+(\phi \to \blacksquare\phi)$.

Second Qualm. The second respect in which the present logic of preservation might be complained of as not living up to its name is that many preservation theorems in mainstream model theory don't provide conditions for the truth of a sentence to be preserved on passage from a *single model* (or single structure, as we might say to remove any suggestion that the discussion is restricted to models of this or that non-logical theory) to other suitably related models. An early preservation result in this style was that in McKinsey [54] to the effect that if each of a set of models verifies a universally quantified Horn sentence, then so does their direct product. (This was later shown to hold without the restriction to universal quantifiers by Alfred Horn. For all unexplained terminology involved here and an "if and only if' version of the result as well as variations, see Hodges [34] or – Chapter 9 of – [33].) McKinsey and Horn used the older terminology of 'direct union' for direct products. ***End of Note.***

As our next illustration of the present apparatus, let us look at the General Barrier Theorem – no initial "Limited" – mentioned in Part I in the Optional Aside (on that result):

Example 3 The result in question reads as follows:

If Γ *is satisfiable and* $\mathcal{R}$*-preserved but* $\Gamma \cup \{\phi\}$ *is* $\mathcal{R}$*-fragile, then* $\Gamma \not\models \phi$.

Let us label some ingredients here, the three hypotheses (a–c) of the result and the negation (d) of the conclusion it claims can be drawn from them, with a view to deriving a contradiction: (a) Γ is satisfiable; (b) Γ is $\mathcal{R}$-preserved; (c) $\Gamma \cup \{\phi\}$ is $\mathcal{R}$-fragile; (d) $\Gamma \not\models \phi$. All sentences involved here – i.e., all elements of $\Gamma \cup \{\phi\}$ – are taken to be free of $\blacksquare$ (and thus also of $\blacklozenge$), and for illustrative purposes we take Γ as $\{\gamma_1, \gamma_2\}$ though feel free to rewrite this as (or to regard it as an eccentric notation for) $\gamma_1 \wedge \ldots \wedge \gamma_n$. Lifting the restriction to Γ finite would require a somewhat different presentation, since ($\blacksquare$-Norm), and, more to the point for the formula-to-formula presentation here, the implicational formulation given of this condition was couched in finitary terms. The idea of the proof is as a selective contraposition of the line of reasoning summarized above. We assume the $\mathcal{M}$-validity, for the ϕ, γ_1, γ_2 in question, of (b), (c), and *not*-(d) (taken as the claim that $\Gamma \models \phi$ – or rather, in the formula form: $(\gamma_1 \wedge \gamma_2) \to \phi$) and conclude with *not*-(a), which appears at line 11.

[116] The connective $\mathcal{P}$ understood in normal modal logic conventionally interpreted and with $\Box$ in place of $\blacksquare$ used to define it (i.e., with $\mathcal{P}\phi$ as $\phi \to \Box\phi$) is introduced in Marcos [53], where $\mathcal{P}\phi$ is written as $\bullet\phi$ and glossed it terms of ϕ's being essentially true, despite his own recognition of the result that "a formula is said to have an essentially true status in case it is simply false" – showing the inappropriateness of the gloss. A modalized variant $\Box(\phi \to \Box\phi)$ is suggested in Humberstone ([42], p. 102) as supplying one of several noncontingency-like notations distinguishable in S4 though coinciding in S5.

1	$(\gamma_1 \wedge \gamma_2) \rightarrow \blacksquare(\gamma_1 \wedge \gamma_2)$	by (b)
2	$(\gamma_1 \wedge \gamma_2) \rightarrow \phi$ by	not-(d)
3	$\blacksquare(\gamma_1 \wedge \gamma_2) \rightarrow \blacksquare\phi$	from 2 by Normality
4	$(\gamma_1 \wedge \gamma_2) \rightarrow \blacksquare\phi$	1, 3 TF
5	$(\gamma_1 \wedge \gamma_2) \rightarrow (\blacksquare(\gamma_1 \wedge \gamma_2) \wedge \blacksquare\phi)$	1, 4 TF
6	$(\gamma_1 \wedge \gamma_2) \rightarrow \blacksquare(\gamma_1 \wedge \gamma_2 \wedge \phi)$	from 5 by Normality
7	$(\gamma_1 \wedge \gamma_2 \wedge \phi) \rightarrow \blacksquare(\gamma_1 \wedge \gamma_2 \wedge \phi)$	6, TF
8	$(\gamma_1 \wedge \gamma_2 \wedge \phi) \rightarrow \neg\blacksquare(\gamma_1 \wedge \gamma_2 \wedge \phi)$	by (c)
9	$\neg(\gamma_1 \wedge \gamma_2 \wedge \phi)$	7, 8 TF
10	$(\gamma_1 \wedge \gamma_2) \rightarrow \neg\phi$	9, TF
11	$\neg(\gamma_1 \wedge \gamma_2)$	2, 10 TF

◁

We turn, next, to performing a similar service in the case of the eventually favoured modification of that observation, the Limited General Barrier Theorem in one of its several alternative formulations, namely (8.13d) from Appendix A.

Example 4 Using the $\blacksquare$ operator, relative to a class of models for the language under consideration and intermodel relation $\mathcal{R}$, we give a proof of the result in question for the special – but entirely representative – case in which $\Gamma = \{\gamma_1, \gamma_2\}$ by way of illustration. The redundant hypothesis is indicated in square brackets here – added into Appendix A's (8.13d) formulation – and in the formalized version:

If Γ is $\mathcal{R}$-preserved and $\Gamma \models \phi$, [and ϕ is $\mathcal{R}$-breakable,] then each model $\mathcal{R}$-related to any model of Γ verifies ϕ.

Since the formulas in $\Gamma \cup \{\phi\}$ are themselves $\blacksquare$-free, where $\mathcal{M}$ is the Model concerned, and writing "$\models$" for "$\models_{\mathcal{M}}$", spelling out the hypotheses amounts to:

(1) $\gamma_1, \gamma_2 \models \blacksquare(\gamma_1 \wedge \gamma_2)$ and (2) $\gamma_1, \gamma_2 \models \phi$,

and, for the "ϕ is $\mathcal{R}$-breakable" (i.e., is not $\mathcal{R}$-preserved) part:

$\big[$(3) $\phi \not\models \blacksquare\phi\big]$.

From these hypotheses we need to show (4): $\gamma_1, \gamma_2 \models \blacksquare\phi$.

To that end, we rewrite (2) as $\gamma_1 \wedge \gamma_2 \models \phi$ and invoke ($\blacksquare$-Norm), getting: $\blacksquare(\gamma_1 \wedge \gamma_2) \models \blacksquare\phi$, from which, taken together with (1), we reach (4) – with no appeal to (3). ◁

Having now seen several illustrations of the current logic $\blacksquare$ in action, we turn to the issue of possible strengthenings of that logic. First, notice that the r.h.s. in the clause for $\blacksquare$ in the inductive definition, above, of (not the consequence relation, but the verification relation) $\models$, "for all $N \in \mathbb{M}$, if $M\mathcal{R}N$ then $N \models \phi$", is equivalent to "for all $N \in \mathbb{M}$, if $M\mathcal{R}^*N$, then $N \models \phi$," where $\mathcal{R}^*$ is the ancestral of whatever $\mathcal{R}$ we might originally have been thinking of.[117] More explicitly, suppose that we first define the notion of the preservation-of-truth relation for a Model $\mathcal{M}$ and a set Φ of sentences of the language being interpreted by $\mathcal{M}$:

Definition 5 Where $\mathcal{M} = (\mathbb{M}, \mathcal{R})$, for $M, N \in \mathbb{M}$, we define:

$$\mathrm{Pres}^{\mathcal{M}}_{\Phi}(M,N) \text{ iff for all } \phi \in \Phi,\ \mathcal{M}, M \models \phi \ \Rightarrow\ \mathcal{M}, N \models \phi.$$

[117] This issue was foreshadowed somewhat informally in Appendix A (in the second half of note 95, and ambient main-text discussion).

For the proof that follows, we use the fact that given binary relations R_1, R_2 (on any set) the ancestral-forming operator is (upward-)monotone, in the sense that $R_1 \subseteq R_2$ implies ${R_1}^* \subseteq {R_2}^*$.

Observation 6 *For $\mathcal{M}$, $\text{Pres}^{\mathcal{M}}_{\Phi}$, as in Def. 5, we have: If $\mathcal{R} \subseteq \text{Pres}^{\mathcal{M}}_{\Phi}$ then $\mathcal{R}^* \subseteq \text{Pres}^{\mathcal{M}}_{\Phi}$.*

Proof. Suppose that $\mathcal{R} \subseteq \text{Pres}^{\mathcal{M}}_{\Phi}$. Then by the monotone property alluded to above: $\mathcal{R}^* \subseteq (\text{Pres}^{\mathcal{M}}_{\Phi})^*$. But since, as the form of Def. 5 makes evident, $\text{Pres}^{\mathcal{M}}_{\Phi}$ is already transitive and reflexive, $(\text{Pres}^{\mathcal{M}}_{\Phi})^* = \text{Pres}^{\mathcal{M}}_{\Phi}$, giving the result claimed. ◁

So, relative to any choice of $\mathcal{M}$, if our attention is focussed on *preservation*, we have the T and 4 principles associated with reflexive and transitive accessibility, often encountered as implicational axioms or axiom-schemata, though here written in the above (connective-minimizing) style, and again with "$\models$" for "$\models_{\mathcal{M}}$ (for arbitrary $\mathcal{M}$)":

$$\blacksquare\phi \models \phi \qquad (\mathsf{T}) \qquad\qquad \blacksquare\phi \models \blacksquare\blacksquare\phi \qquad (4)$$

For the moment, let us just park these modal principles here with a view to returning to them below, since, like the discussion of the collective/distributive distinction above (Example 2), the subject of the 'redundancy aspect' of the Limited General Barrier Principle remarked on in pp. 191–193 of *Barriers* and treated here in the (a/b)–(c)–(d) discussion in Appendix A, does not require appeal to the new principles. (We return to the $\mathcal{R}$-vs.-$\mathcal{R}^*$ issue after the Aside on the MacIntosh Rule below.)

These illustrations of the current formalism exploit only the K/normality aspects of $\blacksquare$'s behaviour, rather than the fuller strength of KT4 = S4 alluded to above and put temporarily on hold. So let us include one example exploiting this greater strength, and, in particular, the availability of T. This allows us to simplify the representation of fragility:

Example 7 The representation of ϕ's fragility w.r.t. $\mathcal{M}$ as the $\mathcal{M}$-validity of $\phi \to \neg\blacksquare\phi$ (or, if preferred, of $\phi \to \blacklozenge\neg\phi$) can be simplified by dropping the antecedent of this condition altogether. Evidently the conditional follows from its consequent, so what needs to be checked is that from $\phi \to \neg\blacksquare\phi$, we can deduce $\neg\blacksquare\phi$ *simpliciter*; "TF" indicates the appeal to purely truth-functional reasoning in deriving consequences:[118]

1	$\phi \to \neg\blacksquare\phi$	hypothesis
2	$\blacksquare\phi \to \phi$	T
3	$\neg\phi \to \neg\blacksquare\phi$	2, TF
4	$\neg\blacksquare\phi$	1, 3, TF

◁

What about the prospect, for certain choices of $\mathcal{M}$ at least, of going beyond S4? This question brings us to a recurrent theme of *Barriers*, first encountered at pp. 13 and 33, by way of anticipation, then again in more detail later.[119] The observation of interest, appropriately stressed by Russell, is that whenever an intermodel relation preserves the truth of ϕ, its converse preserves the truth of $\neg\phi$, and therefore when that relation is

[118] Repeated appeals to this, as in the justifications for lines 4 and 5 here, could be collapsed to single appeals – with 5 announced as following by TF from 1 and 3 – but the aim here is to make such proofs clearer rather than simply shorter.

[119] In particular at pp. 106, 107, in Propositions 3.2, 3.3, respectively, as well as in a couple of sentences in the top half of p. 128, and most extensively in §9.4 (p. 204*f.*), and also p. 255, in the "going informal" part of the book, where the intermodel relations $\mathcal{R}$ have become relations between things called *Combinations*.

symmetric (the relation then coinciding with its converse), the class of all sentences whose truth it preserves is closed under negation, as is the class of sentences it does not preserve – with respect to which it is breakable, in Russell's most commonly used terminology: and indeed this gives her favoured formulations: the normative sentences as those which are s-shift breakable, the Future sentences as those which are future-shift breakable, the universal sentences as those which are extension-breakable, for example. Of these three examples, the first two cases involve symmetric relations while the third does not, but before returning to the symmetric cases let us pause to note that because all of these classes are defined model-theoretically, all are closed under (logical) equivalence, so we have the following point for any such class of Φ of sentences: If Φ is closed under negation then so is its complement $L \setminus \Phi$.

The argument is just a re-run of part of that in note 3 and makes essential use of $\neg$ behaving classically.[120] So whether $\mathcal{R}$ is symmetric or not, the class of $\mathcal{R}$-preserved (alias $\mathcal{R}$-anti-fragile or $\mathcal{R}$-unbreakable) sentences is closed under negation if and only if the class of $\mathcal{R}$-breakable sentences is closed under negation. What the symmetry of $\mathcal{R}$ does is to convert this "iff" into an "and," in view of the argument given in *Barriers* and summarized above.

From the 'meta-modal' perspective of this appendix, what this does is to strengthen what we have so far with T and $\mathsf{4}$ – namely, though subject to a qualification we shall get to presently – the modal logic $\mathsf{S4}$ ("$\mathsf{KT4}$" in the 'anatomy-of-an-axiomatization' nomenclature of Lemmon–Scott and Chellas) – up to $\mathsf{S5}$ (with its most appropriate anatomical label for current purposes being $\mathsf{KTB4}$) by adding what would in the Kripke semantics (with "$\blacksquare, \blacklozenge$" understood as their white counterparts are) be the symmetry-expressing condition on accessibility, (B): $\blacklozenge\blacksquare\phi \models \phi$. Here "$\models$" means, more explicitly: $\models_{\mathcal{M}}$ for any $\mathcal{M} = (\mathbb{M}, \mathcal{R})$ with $\mathcal{R}$ symmetric. In this setting, let us see how the argument from symmetry to closure under negation works out:

Example 8 To reduce clutter and increase familiarity (given the usual presentation of modal deductions), we run the argument using implicational sentences, as with the 'simplification of fragility' example above, rather than $\models$-claims. We'll need the $n = 1$ case of ($\blacksquare$-Norm) above, with $\blacklozenge$ in place of $\blacksquare$: call this $\blacklozenge$-Mono(tonicity), which is an easy consequence of ($\blacksquare$-Norm) itself and the definition of $\blacklozenge$.[121]

1	$\phi \to \blacksquare\phi$	hypothesis that ϕ is $\mathcal{R}$-preserved
2	$\blacklozenge\phi \to \blacklozenge\blacksquare\phi$	1, $\blacklozenge$-Mono
3	$\blacklozenge\blacksquare\phi \to \phi$	B
4	$\blacklozenge\phi \to \phi$	2, 3 TF
5	$\neg\phi \to \blacksquare\neg\phi$	TF (contraposition) and def. $\blacklozenge$

$\lhd$

Gratifying as it may no doubt be to see this recurrent theme from *Barriers* from the current (meta-)modal perspective, there is a bit more to say about it, confined here in an *Aside* so that it may be skipped by readers with less interest in such matters.

An Aside on the MacIntosh Rule. The core of the proof in Example 8 lies in lines (1)-(4), since line (5) can be regarded as an insignificantly different rendering of (4). The transition thereby effected is the passage from premise to conclusion of the rule referred to

[120] Suppose $\psi \in L \setminus \Phi$ but $\neg\psi \notin L \setminus \Phi$. The latter implies that $\neg\psi \in \Phi$, so by the assumption that Φ is closed under negation, $\neg\neg\psi \in \Phi$ and therefore (by the point about equivalence just made) $\psi \in \Phi$, contradicting that our supposition that $\psi \in L \setminus \Phi$.

[121] As those familiar with modal logic will be aware, 1 is the only value of n for which a version of ($\blacksquare$-Norm)-with-$\blacklozenge$-for-$\blacksquare$ which *does* follow from ($\blacksquare$-Norm) in this way.

in the title of this excursus, to which we shall return in a moment after pausing to note one aspect of the B-schema in play here. Let us use the more conventional language of ordinary modal logic with its $\Box$ and $\Diamond$ so that things look a little more familiar for the moment. As the form in which it appeared above, this in this notation would be $\Diamond\Box\phi \to \phi$, there is also the familiar dual version of the schema equally deserving to be called B, namely $\phi \to \Box\Diamond\phi$. Slightly less familiar, perhaps is the fact that imposing the condition on a normal modal logic (conceived of as a set of formulas) that it be closed under either of the rules:

$$\frac{\Diamond\phi \to \psi}{\phi \to \Box\psi} \qquad\qquad \frac{\phi \to \Box\psi}{\Diamond\phi \to \psi}$$

These rules are interderivable with each other on the basis of any normal modal logic as well as with the either of the B-schemas mentioned above.

Either of the corresponding special cases – again interderivable – in which ψ is taken to be ϕ itself has been called the Macintosh Rule in the literature.[122]

$$\frac{\Diamond\phi \to \phi}{\phi \to \Box\phi} \qquad\qquad \frac{\phi \to \Box\phi}{\Diamond\phi \to \phi}$$

Blackening the box and diamond, we see the formulation on the right is none other than the transition from (1) to (4) in the proof above, and thus, essentially, from (1) ("ϕ is $\mathcal{R}$-preserved") to (5) ("$\neg\phi$ is $\mathcal{R}$-preserved"). We already knew that this could be done with the aid of B as an axiom(-schema) for symmetry, but the current recasting of the derivation presents the process rather more directly as getting us from (1) to (5) as an application of a rule which is a special case of a rule for symmetry. And since the MacIntosh rule is a *very* special case of that rule – the $\psi = \phi$ case, it raises a question about whether a weaker condition than the symmetry of $\mathcal{R}$ would suffice for the implication from ϕ's being $\mathcal{R}$-preserved to $\neg\phi$'s being similarly preserved. Chellas and Segerberg [12] present us with examples of normal modal logics for which the MacIntosh rule is admissible but in which B is not provable; for example KD – [12], Theorem 3.8 – is one such logic, which is certainly not an extension of KB.

We need to remember that if the basic modal logic of preservation is taken to be $\mathsf{S4}$, the more pertinent question to ask in this connection is whether there are, specifically, extensions of $\mathsf{S4}$ for which the Macintosh Rule is admissible but B remains unprovable. And in this case the answer is readily seen to be negative: applied to an instance of the $\mathsf{4}$-schema $\blacksquare\neg\phi \to \blacksquare\blacksquare\neg\phi$ the MacIntosh Rule yields the conclusion $\neg\blacksquare\neg\phi \to \blacksquare\neg\blacksquare\neg\phi$, i.e., $\mathsf{5}$ (sometimes called E – for "Euclidean": see note 9) $\blacklozenge\phi \to \blacksquare\blacklozenge\phi$. And, as is well known, $\mathsf{KT5} = \mathsf{KTB4} = \mathsf{KDB4} = \mathsf{S5}$, so there is after all no way of avoiding B here once we have the MacIntosh Rule, despite the initial appearance of only exploiting a special case of the latter rule. ***End of Aside.***

Not that we are left speechless on matters of backward preservation when the intermodel $\mathcal{R}$ under consideration is not symmetric, since we can do for $\blacksquare$ what we might do for $\Box$ when its (intramodel) accessibility relation is not symmetric: boost the language to a 'tense-logical' style expansion with an operator for the converse relation in either (intra- or intermodel) case.[123] Here shudder quotes are used so as not to offend those who would prefer tense

[122] For the etymology of this name and further information on the rule, see Chellas and Segerberg [12], as well as the companion article Williamson [91], and briefly also on p. 96 ((a) and (c) of Corollary 2, and the last full paragraph) of Williamson [90], as well as in the Digression beginning on p. 98 of Humberstone [42]. (The final sentence of that Digression is typographically marred, and should read: "From this, we know via Exercise 2.5.53 and (1)–(10) following it, that we can derive B.")

[123] Cf. the parenthetical remark in *Barriers* four lines above the mid-page ornament on p. 128 there: "They can be made true by $\mathcal{R}^{-1}$-ing."

logic to be construed as dealing with time-related considerations, as opposed to having been merely motivated in its evolution by such considerations. For them we could call the bimodalization which consists in adjoining new operators $\Box_i^{-1}$ for antecedently available operators $\Box_i$ "conjugated" modal logic, after the description of such pairs (or their modal-algebraic analogues) *conjugate* operators in Jónsson and Tarski ([47], p. 903). Thus Prior's G and H (in *Barriers*: $\mathcal{G}$ and $\mathcal{H}$), are a pair of such operators in the case of tense logic (i.e., $H = G^{-1}$ and – equivalently – $G = H^{-1}$).[124] Here, we would be considering a conjugate operator $\blacksquare^{-1}$ for the current $\blacksquare$. The point remarked on as a leitmotif in *Barriers* in note 119 above, and the text to which that note is appended, can then be put by saying that relative to $\mathcal{M} = (\mathbb{M}, \mathcal{R})$, we always have $\phi \models_{\mathcal{M}} \blacksquare\phi$ iff $\neg\phi \models_{\mathcal{M}} \blacksquare^{-1}\neg\phi$, regardless of whether $\mathcal{R}$ is symmetric (though of course if it is, there is no point in introducing "$\blacksquare^{-1}$" as distinct from "$\blacksquare$").[125]

The final paragraph of the Aside above opened with the claim there was nothing to choose between B and (closure under) the MacIntosh Rule, the latter being a 'rule' incarnation of the claim that if the negation of any $\mathcal{R}$-preserved sentence is again $\mathcal{R}$-preserved, "if the basic modal logic of preservation is taken to be $\mathsf{S4}$," so we pause to reflect on the work this "if" is doing here, and on what would happen if we thought, instead, of the basic logic of $\blacksquare$ itself, as currently interpreted i.e., in terms of Models $\mathcal{M} = (\mathbb{M}, \mathcal{R})$. Recall that for a sentence ϕ, relative to such a Model, it is not $\blacksquare\phi$ that says that ϕ's truth is preserved by $\mathcal{R}$ throughout $\mathcal{M}$ (in the sense that the latter holds iff we have $\mathcal{M} \models \phi$), but rather $\phi \to \blacksquare\phi$ that says this (with what we have after the "iff" being $\mathcal{M} \models \phi \to \blacksquare\phi$, or equivalently $\phi \models_{\mathcal{M}} \blacksquare\phi$).

The issue of symmetry raises a further question on which we shall conclude the present Appendix. The B axiom schema in conventional modal logic, whether taken in the form $\phi \to \Box\Diamond\phi$ or in the dual form $\Diamond\Box\phi \to \phi$, is valid (i.e., has each of its instances valid) on all symmetric frames, but we know something stronger: the schema is valid on all *and only* such frames. Can we match this 'modal definability' result in the present setting? The only thing stressed about the analogous case of $\phi \to \blacksquare\blacklozenge\phi$ is that for any $\mathcal{M} = (\mathbb{M}, \mathcal{R})$, if $\mathcal{R}$ is symmetric, then $\models_{\mathcal{M}} \phi \to \blacksquare\blacklozenge\phi$ (or, if preferred: that $\phi \models_{\mathcal{M}} \blacksquare\blacklozenge\phi$). This naturally suggests the question of whether we can similarly boost this "if" to an "iff". Can we, that is, have $\models_{\mathcal{M}} \phi \to \blacksquare\blacklozenge\phi$ for all ϕ, without $\mathcal{R}$ being symmetric? This is not an easy question to settle.

To ponder it, we could begin with the familiar case, thinking of Kripke models the way this is done in *Barriers* except considering unpointed models (no distinguished world for truth at which to count as being true in the model): $\langle W, R, I\rangle$, and for a simpler exposition, where I maps each p_i to a subset of W, the set *Barriers* picks out instead as the set of all $w \in W$ with $I(p_i, w) = 1$. The current interest concerns the frames $\langle W, R\rangle$ and the notation of validity on a frame as truth at all points in every model $\langle W, R, I\rangle$, and more specifically the "only if" direction for the familiar claim that any representative instance of the B-schema – $\Diamond\Box p \to p$ will do – is valid on $\langle W, R\rangle$ if and only if R is a symmetric relation on W. One supposes that R is not symmetric, so that for some $x, y \in W$, xRy and not yRx, with a view to finding a model on the frame and a point in that model at which the chosen formula is false. With the x, y just invoked in mind, one sets I to be any

[124] A non-temporal illustration involving metaphysical necessity, appears as Example 4.1.11 in Humberstone [42], p. 209. Another is given at p. 268 for deontic logic with $q \to \mathcal{O}\mathcal{P}^{-1}q$, to use *Barriers*'s notation, with the antecedent to consequent inference as a contender for the status of a counterexample to Hume's Law, which the interested reader is invited to ponder with the apparatus of *Barriers* to hand. (Here of course, $\mathcal{P}^{-1}$ is $\neg\mathcal{O}^{-1}\neg$.) See also the case of a conjugate for the belief operator in doxastic logic: [42], 439–446.

[125] An economical axiomatic presentation of the logic we get by adding a conjugate operator is described on p. 180*f.* of [42] using a disjunctive variation on the two-way rule above with schematic letters ϕ and ψ, allowing the transition in either direction between (provable) $\Diamond\phi \to \psi$ and $\phi \to \Box^{-1}\psi$; the disjunctive form in question, writing G and H for $\Box$ and $\Box^{-1}$ then permits passage in either direction between $\phi \vee G\psi$ and $H\phi \vee \psi$. The normality of either G or H, whichever we start with, is then inherited by the other with no further (syntactic) ado.

interpretation satisfying $I(p) = \{w \in W \mid yRw\}$. This secures the result that $\Box p$ is true at y in the resulting model $\langle W, R, I\rangle$, and hence that $\Diamond\Box p$ is true at x, since xRy, while at the same time p is false at x since *ex hypothesi* not yRx – thereby falsifying the implication $\Diamond\Box p \to p$ at x in the model concerned. If we call the model M, we can turn it into a pointed model by supplying x as the desired w_M (adapting *Barriers*, Chapter 7, which corresponds to the "$@_M$" of Chapters 5 and 8) but for present purposes we will regard this distinguished point w_M as external to the model, and denote the pointed version of the model as $\langle M^-, w\rangle$ and consider different choices of $w \in W$ as 'alternative pointings' of the unpointed model $\langle W, R, I\rangle$. We will use "M^-" as variable ranging over such (unpointed) models to avoid notational confusion with the use of $\mathcal{M}$ elsewhere (when Kripke-style semantics is at issue) for pointed models. Think of the superscripted minus sign as an indication that something – namely the specification of distinguished point – is missing.) The process of converting a model of the kind in play here into a Model (for the language with $\blacksquare$ but in this case no other non-Boolean connectives) is in any case a potentially useful illustration of a rather different way for such Models to arise than our usual preoccupation with $\mathcal{R}$-preservation:

Example 9 The construction we have in mind converts a single (unpointed) Kripke model $M^- = \langle W, R, I\rangle$ into a Model $\mathcal{M}^{M^-} = (\mathbb{M}^{M^-}, \mathcal{R}^{M^-})$, where $\mathbb{M}^{M^-} = \{\langle M^-, w\rangle \mid w \in W\}$ and for $w, w' \in W$, $\langle M^-, w\rangle \mathcal{R} \langle M^-, w'\rangle$ iff wRw'. Informally summarizing, then, we duplicate the single unpointed Kripke model M^- by means of a Model comprising all the alternative pointings of M^-, with one of these being accessible (by $\mathcal{R}$) to another just in case the distinguished element of the former is accessible (by the original relation R) to the distinguished element of the latter. We have $\mathcal{M}, \langle M^-, w\rangle \models \phi$ just in case $M^-, w \models \phi$. $\lhd$

We can use the Models derived in this way to answer many questions about the possibilities for $\mathcal{R}$. For instance, consider the famous question of whether there can be a binary relation which is transitive and symmetric without being reflexive – famous because of early incorrect negative answers to it in print. The empty relation on a one-element structure provides an example realizing the possibility in question, as would, say a three-element structure with universe $\{0, 1, 2\}$ and $R = \{1, 2\} \times \{1, 2\}$, so, taking the latter, say, we can easily decorate it with a distribution of truth-values to formulas at its elements so as to obtain a model with the structure in question as its frame, perhaps one refuting the claim that $\Box p \to p$ is a theorem of the logic KB4, which the construction described in Example 9 converts into a Model illustrating the possibility of have a Model $(\mathbb{M}, \mathcal{R})$ with $\mathcal{R}$ transitive and symmetric (over $\mathbb{M}$) without being reflexive (over $\mathbb{M}$). It would of course similarly settle the question about the modal formulas involved (now written in blackened form). And we know, returning to our original question, we can convert any Kripke model with a non-symmetric R and falsifying $\Diamond\Box p \to p$ as some point, into a (shall we say?) Kripke Model not verifying $\blacklozenge\blacksquare p \to p$ relative to one of its models. But the question was whether for an *arbitrary* $\mathcal{M} = (\mathbb{M}, \mathcal{R})$, and not just one arising from the special construction featuring in Example 9, if $\mathcal{R}$ was not symmetric, then $\mathcal{M} \not\models \blacklozenge\blacksquare\phi \to \phi$ for some ϕ.[126]

This leaves us with a question mark over one of the possible implications among the condition on Models $\mathcal{M} = (\mathbb{M}, \mathcal{R})$ in play in our discussion (which does not presume transitivity or reflexivity for $\mathcal{R}$):

$$\mathcal{R} \textit{ is symmetric} \tag{1}$$

$$\textit{For all } \phi:\ \blacklozenge\blacksquare\phi \models_{\mathcal{M}} \phi \tag{2}$$

$$\textit{For all } \phi:\ \phi \models_{\mathcal{M}} \blacksquare\phi \implies \neg\phi \models_{\mathcal{M}} \blacksquare\neg\phi \tag{3}$$

[126] We can't limit this to p, even if the language in question contains sentence letters, in view of the failure of uniform substitution.

$$(1) \Rightarrow (2) \Rightarrow (3) \qquad (3) \not\Rightarrow (2) \text{ [and so, } (3) \not\Rightarrow (1)] \qquad (2) \overset{?}{\Rightarrow} (1)$$

Much of this is implicit in the Aside on the MacIntosh rule above, (3) being a version of that rule for the 'formula logic' of $\mathcal{M}$ (i.e., the set of $\models$-consequences of $\varnothing$).[127] The reasoning for the implication (1) $\Rightarrow$ (2) for Models $\mathcal{M}$ mimics that for ordinary Kripke models, though we have seen that the same cannot be said for the converse implication, (2) $\Rightarrow$ (1), which remains an open question. For (2) $\Rightarrow$ (3), the reasoning in the Aside above suffices, translated into the current format. That is, we begin by supposing (2) and that $\phi \models_{\mathcal{M}} \blacksquare\phi$, (remembering that this means that for all $M \in \mathbb{M}$: if $\mathcal{M}, M \models \phi$, then $\mathcal{M}, M \models \blacksquare\phi$), from which we infer that $\blacklozenge\phi \models_{\mathcal{M}} \blacklozenge\blacksquare\phi$. In fact this shows something stronger, namely that *for any given* ϕ, the ('B-axiom') condition in (2) implies the conditional formulation (3).

It remains to check the claimed non-implication (3) $\not\Rightarrow$ (2), since we helped ourselves to T and (more significantly) 4 in the earlier defence (in the Aside above) of this implication. To that end we retrieve the notion of *$*$-symmetry* from Chellas and Segerberg [12]. A binary relation R has this property just in case $xRy \Rightarrow yR^*x$, for all x, y. This is a generalization of the 'safe return' property of van Benthem ([6], pp. 333, 348).[128] It is not hard to show that the accessibility relation in the canonical frame of any consistent normal modal logic (in the conventional sense) closed under the MacIntosh rule is $*$-symmetric, though for present purposes all we need to do is provide a case of $*$-symmetry without symmetry in the current $\blacksquare$ setting:

Example 10 Consider the simple (if 'artificial') case of $\mathcal{M} = (\mathbb{M}, \mathcal{R})$ in which $\mathbb{M} = \{M_1, M_2, M_3\}$ and $\mathcal{R} = \{\langle M_1, M_2\rangle, \langle M_2, M_3\rangle, \langle M_3, M_1\rangle\}$, noting that $\mathcal{R}$ is non-symmetric (and is, in fact, asymmetric). We check that despite this failure of condition (2) – and so, equivalently, of (1) – condition (3) is satisfied. For a contradiction, suppose that it is not: that for some ϕ we have (a) $\phi \models_{\mathcal{M}} \blacksquare\phi$ while (b) $\neg\phi \not\models_{\mathcal{M}} \blacksquare\neg\phi$, say (WLOG) because (c) $\mathcal{M}, M_1 \models \neg\phi$ although (d) $\mathcal{M}, M_1 \not\models \blacksquare\neg\phi$. (d) implies that $\mathcal{M}, M_2 \models \phi$, so by (a), $\mathcal{M}, M_2 \models \blacksquare\phi$. Thus, as $M_2\mathcal{R}M_3$, and by (a) again $\mathcal{M}, M_3 \models \blacksquare\phi$, and so, since $M_3\mathcal{R}M_1$, $\mathcal{M}, M_1 \models \phi$, contradicting (c). Note that in the counterexample just considered, $\mathcal{R}$ is not just $*$-symmetric without being symmetric, it is what one could call *uniformly* $*$-symmetric: it's not just that for all $M, N \in \mathbb{M}$, $M\mathcal{R}N$ implies that for some k, $N\mathcal{R}^kM$; rather, there is some k such that for all $M, N \in \mathbb{M}$, $M\mathcal{R}N$ implies that $N\mathcal{R}^kM$. (In our case, the relevant k is 2; picking such a specific value turns the second-order condition of $*$-symmetry into a simple first-order condition.) $\lhd$

As remarked at the end of note 128, a binary relation's $*$-ancestrality is equivalent to its having a symmetric ancestral. Already in Observation 6 we saw how concern with preservation turns attending to a binary relation into attending to its ancestral, independently of our recent focus on symmetry. If we remove $\langle M_3, M_1\rangle$ from the $\mathcal{R}$ of Example 10, for instance, we see that the truth of $\phi \to \blacksquare\phi$ at M_1 and M_2 is enough to secure ϕ's truth at M_3 given

[127] Recall also that we can also replace (1) by an equivalent plain ϕ on the left of the $\models$ and $\blacksquare\blacklozenge\phi$ on the right.

[128] It is, as it were, just the 'return property': if you can get from x to y in one R-step, then you can return from y to x via some number of intermediate points; *safe return* demands, further, that x bears the relation R to each of these intermediate points in at least one such return route. (Van Benthem observes that this condition holds for all elements x, y in a standard Kripke frame iff the frame validates a conditional schema related to the MacIntosh rule – a rule formulation of (3) – with premise $\phi \to \Box\phi$ and conclusion $\Diamond\phi \to \phi$, namely the implication whose consequent is the conclusion of this rule and whose antecedent is the necessitation of its premise. This is done in [6] as part of showing that the smallest normal modal logic extending KT and containing all instances of this schema as well as of the McKinsey schema $\Diamond(\Box\phi \vee \Box\neg\phi)$ is Kripke-incomplete, not containing $p \to \Box p$ despite the latter formula's validity on every frame validating the axioms in question.) For $*$-symmetry (the mere 'return' property), we note – though we do not need this for our immediate purposes – that it is equivalent to the condition that xR^*y implies yR^*x, for all x, y.

its truth at M_1, even though we do not have $M_1 \mathcal{R} M_3$. Ordinary mathematical induction – we set transfinite induction to one side here – works like this, with $\phi \to \blacksquare\phi$ handling the inductive step, so if we have an ω-chain of models $M_0, M_1, \ldots, M_n, \ldots$ comprising $\mathbb{M}$, with $\mathcal{R} = \{\langle M_i, M_i + 1\rangle \mid i \in \omega\}$, then for $\mathcal{M} = (\mathbb{M}, \mathcal{R})$, the assumption that $\mathcal{M}, M_0 \models \phi$ and $\models_{\mathcal{M}} \phi \to \blacksquare\phi$ secure the conclusion that $\models_{\mathcal{M}} \phi$. So, illustrating the point, if the relation of d-extension in *Barriers* is restricted to a form allowing the addition of at most countably many 'new' objects to the domain, then this can equivalently be restricted still further to a form permitting only the addition of one new object at a time, this does not affect the set of sentences that are extension-preserved.[129] Not that any such restrictions are being suggested here – or indeed that the status of S4 as the basic modal logic of preservation be rescinded: the point has just been to clarify some of the issues involved.

Acknowledgments. I would like to thank Gill Russell for clarifying a point of interpretation for me, Shawn Standefer for assisting with various aspects of the production of this review, and Rohan French for helpful suggestions and numerous corrections.

References

[1] A. R. Anderson, 'The Logic of Norms', *Logique et Analyse* **1** (1958) 84–91.

[2] A. R. Anderson and N. D. Belnap, *Entailment: the Logic of Relevance and Necessity, Vol. I*, Princeton University Press, Princeton, NJ 1975.

[3] Albert Atserias, Anuj Dawar, and Phokion Kolaitis, 'On Preservation under Homomorphisms and Unions of Conjunctive Queries', *Journal of the ACM* **53** (2006), 208–237.

[4] Albert Atserias, Anuj Dawar, and Martin Grohe, 'Preservation Under Extensions on Well-Behaved Finite Structures', *SIAM Journal on Computing* **38** (2008), 1364–1381.

[5] Matthew S. Bedke, 'Non-descriptive Relativism', pp. 48–70 in R. Shafer–Landau (ed.), *Oxford Studies in Metaethics: 13*, Oxford University Press, Oxford 2018.

[6] Johan van Benthem, 'Correspondence Theory', D.M. Gabbay and F. Guenthner (eds.), 325–408 in *Handbook of Philosophical Logic*, (2nd Edition), *Volume 3*, Kluwer, Dordrecht 2001.

[7] Sören Berger, Alexander Christensen Block and Benedikt Löwe, 'The Modal Logic of Abelian Groups', (as of 2024) to appear in *Algebra Universalis.*

[8] Gunnar Björnsson and Robert Shanklin, '"Must", "Ought" and the Structure of Standards', pp. 33–48 in F. Cariani, D. Grossi, J. Meheus, and X. Parent (eds.), *Deontic Logic and Normative Systems* (12th International Conference, DEON 2014), Springer, Cham, Switzerland 2014.

[9] Ricki Bliss and Graham Priest, 'The Geography of Fundamentality: An Overview', pp. 1–33 in R. Bliss and G. Priest (eds.), *Reality and its Structure: Essays in Fundamentality*, Oxford University Press, Oxford 2018.

[10] George Boolos, *The Logic of Provability*, Cambridge University Press, Cambridge 1993.

[129] We take adding one new object to be a matter of adding *exactly* (rather than *at most*) one new object, for the sake of giving an illustration with an irreflexive intermodel relation; here we pick up the trivial preservation of ϕ's truth on the passage from M to itself not by multiple applications but by zero applications ($\mathcal{R}^0$ being the identity relation), or, to put it differently, because the ancestral of a relation is its not its transitive closure but its reflexive transitive closure.

[11] C. C. Chang and H. J. Keisler, *Model Theory*, 3rd Edn., Elsevier, Amsterdam 1990.

[12] Brian Chellas and Krister Segerberg, 'Modal Logics with the MacIntosh Rule', *Journal of Philosophical Logic* **23** (1994), 67–86.

[13] J. W. Cowan, *The Complete Lojban Language*, A Logical Language Group Publication, 2016

[14] John N. Crossley and Lloyd Humberstone, 'The Logic of "Actually"', *Reports on Mathematical Logic* **8** (1977), 1–29.

[15] Martin Davies, *Meaning, Quantification, Necessity: Themes in Philosophical Logic*, Routledge and Kegan Paul, London 1981.

[16] Amy Rose Deal, 'Indexiphors: Notes on Embedded Indexicals, Shifty Agreement, and Logophoricity', pp. 59–86 in R. Ivan (ed.), *The Leader of the Pack: A Festschrift in Honor of Peggy Speas*, University of Massachusetts Occasional Papers in Linguistics #40, GLSA Amherst 2018.

[17] Robert Demolombe and Luis Fariñas del Cerro, 'Towards a Logical Characterization of Sentences of the Kind "Sentence p is about object c"', pp. 85–99 in S. Hölldobler (ed.), *Intellectics and Computational Logic: Papers in Honor of Wolfgang Bibel*, Kluwer, Dordrecht 2000.

[18] Robert Demolombe and Luis Fariñas del Cerro, 'How to Produce Information About a Given Entity Using Automated Deduction Methods', *Electronic Notes in Theoretical Computer Science* **256** (2009), 19–31.

[19] Robert Demolombe and Luis Fariñas del Cerro,'Information About a Given Entity: From Semantics Towards Automated Deduction', *Logic Journal of the IGPAL* **20** (2010), 1231–1250.

[20] Robert Demolombe, Andrew J. I. Jones, 'On Sentences of the Kind "Sentence p is about topic t": Some steps toward a formal-logical analysis', pp. 115–133 in H. J. Ohlbach and U. Reyle (eds.), *Logic, Language and Reasoning: Essays in Honor of Dov Gabbay*, Kluwer, Dordrecht 1999.

[21] Leo Esakia and Benedikt Löwe, 'Fatal Heyting Algebras and Forcing Persistent Sentences', *Studia Logica* **100** (2012), 163–173.

[22] Gareth Evans, 'Does Tense Logic Rest upon a Mistake?', Chapter 12 in Evans, *Collected Papers*, Oxford University Press, Oxford 1985.

[23] Philippa Foot, 'Moral Arguments', *Mind* **67** (1958), 502–513 (appearing also as Paper 7 in Philippa Foot, *Virtues and Vices, and other essays in moral philosophy*, Basil Blackwell, Oxford 1978).

[24] Philippa Foot, 'Moral Beliefs', *Procs. of the Aristotelian Society* **59** (1958), 83–104 (appearing also as Paper 8 in Foot, *Virtues and Vices, and other essays in moral philosophy*, Basil Blackwell, Oxford 1978).

[25] James W. Garson, 'Two New Interpretations of Modality', *Logique et Analyse* **15** (1972), 443–459.

[26] Silvio Ghilardi, 'The Invariance Modality', pp. 165–175 in A. Citkin and I. M. Vandoulakis (eds.), *V. A. Yankov on Non-Classical Logics, History and Philosophy of Mathematics*, Spinger Nature, Switzerland 2022.

[27] Alessandro Giordani,'On the Notion of Aboutness in Logical Semantics', pp. 407–467 in F. L. G. Faroldi and F. Van De Putte (eds.), *Kit Fine on Truthmakers, Relevance, and Non-classical Logic*, Springer, Cham (Switzerland) 2023.

[28] Nelson Goodman, 'About', *Mind* **70** (1961), 1–24.

[29] Georgia M. Green, 'The Lexical Expression of Emphatic Conjunction', *Foundations of Language* **10** (1973) 197–248.

[30] Joel David Hamkins and Benedikt Löwe, 'The Modal Logic of Forcing', *Transactions of the American Mathematical Society* **360** (2008), 1793–1817.

[31] Joel David Hamkins and Wojciech Aleksander Wołoszyn, 'Modal Model Theory', *Notre Dame Journal of Formal Logic* **65** (2024), 1–37.

[32] Peter Hawke, 'Theories of Aboutness', *Australasian Journal of Philosophy* **96** (2018), 697–723.

[33] Wilfrid Hodges, *Model Theory*, Cambridge University Press, Cambridge 1993.

[34] Wilfrid Hodges, 'Logical Features of Horn Clauses', pp. 449–503 in D. Gabbay, C. J. Hogger and J. A. Robinson (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 1: Logical Foundations*, Clarendon Press, Oxford 1993.

[35] Wesley H. Holliday, Tomohiro Hoshi and Thomas F. Icard III, 'Information Dynamics and Uniform Substitution', *Synthese* **190** (2013), 31–55.

[36] Lloyd Humberstone, 'Necessary Conclusions', *Philosophical Studies* **41** (1982), 321–335.

[37] Lloyd Humberstone, 'First Steps in Philosophical Taxonomy', *Canadian Journal of Philosophy* **12** (1982), 467–478.

[38] Lloyd Humberstone, 'A Study in Philosophical Taxonomy', *Philosophical Studies* **83** (1996), 121–169.

[39] Lloyd Humberstone, 'Modal Logic for Other-World Agnostics: Neutrality and Halldén Incompleteness', *Journal of Philosophical Logic* **36** (2007), 1–32.

[40] Lloyd Humberstone, 'Can Every Modifier be Treated as a Sentence Modifier?', *Philosophical Perspectives* **22** (2008), 241–275.

[41] Lloyd Humberstone, *The Connectives*, MIT Press, Cambridge, MA 2011.

[42] Lloyd Humberstone, *Philosophical Applications of Modal Logic*, College Publications, London 2016.

[43] Lloyd Humberstone, 'Recent Thought on *Is* and *Ought*: Connections, Confluences and Rediscoveries', *Journal of Applied Logics* **6** (2019), 1372–1446.

[44] Lloyd Humberstone, 'Exploring Essential Occurrence', presented at Melbourne Logic Workshop December 2023, Monash University (Caulfield campus), sponsored by PLEXUS and ARC project "Substructural logics for limited resources" and organized by Paul Égré and David Ripley.

[45] Emil Jeřábek, 'Cluster Expansion and the Boxdot Conjecture', *Mathematical Logic Quarterly* **62** (2016), 608–614.

[46] Andrew J. I. Jones and Ingmar Pörn, "Ought" and "Must", *Synthese* **66** (1986), 89–93.

[47] B. Jónsson and A. Tarski, 'Boolean Algebras with Operators. Part I', *American Journal of Mathematics* **73**, (1951), pp. 891–939.

[48] David Kaplan, 'Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives', pp. 481–563 in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*, Oxford University Press, New York 1989.

[49] Toomas Karmo, ' "I" and "Here" ', *Philosophical Quarterly* **30** (1980), 289–293.

[50] Toomas Karmo 'Some Valid (but no Sound) Arguments Trivially Span the "Is"–"Ought" Gap', *Mind* **97** (1988), 252–257.

[51] David Lewis, 'Semantic Analyses for Dyadic Deontic Logic', pp. 1–14 in S. Stenlund (ed.), *Logical Theory and Semantic Analysis*, Reidel, Dordrecht 1974.

[52] David Lewis, 'Statements Partly about Observation', *Philosophical Papers* **17** (1988), 1–31.

[53] João Marcos, 'Logics of Essence and Accident', *Bulletin of the Section of Logic* **34** (2005), 43–56.

[54] J. C. C. McKinsey, 'The Decision Problem for Some Classes of Sentences without Quantifiers', *Journal of Symbolic Logic* **8** (1943), 61–76.

[55] Charles E. Mitchell, 'The Non-commutativity of Subtraction', *School Science and Mathematics* **83** (1983), 133–139.

[56] John Myhill, '*Should* and *Ought*: the Rise of Individually Oriented Modality in American English', *English Language and Linguistics* **1** (1997), 3–23.

[57] Piergiorgio Odifreddi, *Classical Recursion Theory*, First Edition, Elsevier, Amsterdam 1989, or Second Edition (as *Classical Recursion Theory: Vol. 1*), 1999.

[58] Charles R. Pigden (ed.), *Hume on 'Is' and 'Ought'*, Palgrave Macmillan, NY 2010.

[59] Stephen Pollard (ed.), *Ernst Schröder on Algebra and Logic*, Springer, Cham, Switzerland 2022.

[60] Stefano Predelli, 'I am not here now', *Analysis* **58** (1998), 107–115.

[61] A. N. Prior, 'The Autonomy of Ethics', *Australasian Journal of Philosophy*, **38** (1960), 199–206. (Reprinted as Chapter 10 in P. T. Geach and A. J. Kenny (eds.), *A. N. Prior: Papers on Logic and Ethics*, Duckworth, London 1976, and as Chapter 1.1 in Pigden [58].)

[62] A. N. Prior, 'Egocentric Logic', *Noûs* 2 (1968), 191–207 (1968). Reprinted as Chapter 2 in A. N. Prior and Kit Fine, *Worlds, Times and Selves*, Duckworth, London 1977.

[63] W. V. O. Quine, *Methods of Logic*, Revised edn., Holt, Rinehart and Winston, New York 1950 (reprinted 1959, 1966).

[64] Brian Rabern and Derek Ball, 'Monsters and the Theoretical Role of Context', *Philosophy and Phenomenological Research* **98** (2019), 392–416.

[65] Greg Restall and Gillian Russell, 'Barriers to Implication' pp. 243–259 in Pigden [58].

[66] Mark Richard, *Truth and Truth Bearers: Meaning in Context, Volume II*, Oxford University Press, Oxford 2015.

[67] Francis Ritchie, *First Steps in Latin*, revised by R. D. Wormald, Longmans, Green and Co. Ltd., London and New York 1959. (First publ. 1881.)

[68] R. Routley and V. Routley, 'A Fallacy of Modality', *Noûs* **3** (1969), 129–153.

[69] Aynat Rubinstein, 'On Necessity and Comparison', *Pacific Philosophical Quarterly* **95** (2014) 512–554.

[70] Bertrand Russell, *The Philosophy of Logical Atomism*, Routledge Classics republication of the 1972 edition (published by Fontana), Routledge, Abingdon 2010.

[71] Gillian Russell, 'In Defence of Hume's Law', pp. 151–161 in Pigden [58].

[72] Gillian Russell, 'Indexicals, Context-sensitivity and the Failure of Implication', *Synthese* **183** (2011), 143–160.

[73] Gillian Russell, 'Lessons from the Logic of Demonstratives: what indexicality teaches us about logic, and vice versa', pp. 190–216 in G. Restall and G. Russell (eds.), *New Waves in Philosophical Logic*, Palgrave Macmillan, Basingstoke 2012.

[74] Gillian Russell, 'How to Prove Hume's Law', *Journal of Philosophical Logic* **51** (2022), 603–622.

[75] Gillian Russell, 'The *I* in Logic', *Theoria* ('early view') as of 2024.

[76] Denis I. Saveliev and Ilya B. Shapirovsky, 'On Modal Logics of Model-Theoretic Relations, *Studia Logica* **108** (2020), 989–1017.'

[77] Philippe Schlenker, 'A Plea for Monsters', *Linguistics and Philosophy* **26** (2003), 29–120.

[78] Gerhard Schurz, *The Is-Ought Problem: An Investigation in Philosophical Logic*, Kluwer, Dordrecht 1997.

[79] Gerhard Schurz, 'Relevance as Difference-making: a Generalized Theory of Relevance and its Applications', *Philosophical Studies*, in press.

[80] J. M. Shorter, 'Professor Prior on the Autonomy of Ethics', *Australasian Journal of Philosophy* **39** (1961), 286–287. (Reprinted as Chapter 1.2 in Pigden [58].)

[81] John G. Slater (ed.), *Bertrand Russell, Collected Papers, Vol. 8: The Philosophy of Logical Atomism and other essays, 1914–19*, Allen & Unwin, London 1986.

[82] J. S. Ullian, 'Corollary to Goodman's Explication of "About"', *Mind* **71** (1962), 345.

[83] Bas van Fraassen, *Formal Semantics and Logic*, Macmillan, London 1971. (2016 reissue as an eBook, mostly unchanged, by Nousoul Digital Publishers, 2016.)

[84] Bas van Fraassen, 'The Logic of Conditional Obligation', *Journal of Philosophical Logic* **1** (1972), 417–438.

[85] Ekaterina Vostrikova, 'On the similarity between *unless* and *only-if-not*', *Sinn und Bedeutung* **21** (2016), 1271–1288.

[86] Peter Vranas, 'Comment on "Barriers to Implication"', pp. 260–267 in Pigden [58].

[87] Elise Wege, Sophie Batchelor, Matthew Inglis, Honali Mistry, and Dirk Schlimm, 'Iconicity in Mathematical Notation: Commutativity and Symmetry', *Journal of Numerical Cognition* **6** (2020), 378–392.

[88] Paul Weingartner and Gerhard Schurz, ‘Paradoxes Solved by Simple Relevance Criteria’, *Logique et Analyse* **29** (1986), 3–40.

[89] Yale Weiss, ‘A Relevant Framework for Barriers to Entailment’, manuscript 2024.

[90] Timothy Williamson, ‘An Alternative Rule of Disjunction in Modal Logic’, *Note Dame Journal of Formal Logic* **33** (1992), 89–100.

[91] Timothy Williamson, ‘Non-genuine MacIntosh Logics’, *Journal of Philosophical Logic* **23** (1994), 87–101.

[92] Stephen Yablo, *Aboutness*, Princeton University Press, Princeton, NJ 2014.