Notre Dame Journal of Formal Logic
Volume 50, Number 2, 2009

Collapsing Modalities

Lloyd Humberstone

Abstract Sections 1 and 2 respectively raise and settle the question of whether, if an affirmative modality collapses (reduces to the null modality, that is) in a normal modal logic, then all modalities of the same length collapse in that logic, while Section 3 considers some special cases of an analogous phenomenon for congruential modal logics, closing with a general question about collapsing modalities in this broader range of logics.

1 Collapsing Modalities in Normal Logics: A Question

Notation and terminology here is as in [9], which mainly follows [4] on the nomenclature for normal modal logics. Thus, for example, B is the formula $p \to \Box\Diamond p$ (or the schema $\varphi \to \Box\Diamond\varphi$), KB is the smallest normal modal logic containing this formula (or containing all instances of this schema), B_c is the converse formula (or schema), KB_c is the smallest normal modal logic containing it (or its instances), B! is the corresponding biconditional, and KB! is the smallest normal modal logic containing it (or all its instances), and thus the smallest normal modal logic extending KB and KB_c, and similarly for other candidate modal axioms with conventional transcription in implicational form (such as D, T, 4).[1] A further addition from [9]: C_n is the formula (or the corresponding schema) $p \leftrightarrow \Box^n p$, where the superscript indicates n-fold iteration. (Thus KC_1 = KT!.) This "C", which has nothing to do with the use of the same letter in [4], is intended to recall "cyclic", for semantical reasons (explained in [9]). Here however, our reasoning will be exclusively syntactic,[2] as we explore one aspect of the perennially popular theme[3] of the structure of modalities. From here until Section 3, in which we broaden our horizons, the term 'logic' means *normal monomodal logic* (though we retain the full formulation in the statement of results), and we shall use interchangeably the phrasing "the logic S contains the formula φ" and "S proves φ" (or "$\vdash_{\mathsf{S}} \varphi$").

Received June 10, 2008; accepted November 17, 2008; printed March 20, 2009
2000 Mathematics Subject Classification: Primary, 03B45
Keywords: modal logic, modalities
 10.1215/00294527-2009-001

On a fairly widespread usage of the phrase, a *collapsing identity* for an equational theory, or for a class of algebras, is an equation with a variable on one of its two sides, provable in the theory or satisfied in each algebra in the class.[4] The analogous definition in the case of modal propositional logic would be that of an equivalence $\varphi \leftrightarrow p$ where φ is any formula other than a variable and p is a propositional variable; however, we restrict attention here to φ of the form $\mathrm{O}_1 \dots \mathrm{O}_n p$ ($n \geq 1$), where each O_i is either $\Box$ or $\Diamond$, that is, to the case of applying a nonnull *affirmative modality* to p. (The letter "O" here simply abbreviates "Operator".) In this case we call the equivalence in question, if provable in a given modal logic, a *collapsing equivalence* (for that logic), and refer to $\mathrm{O}_1 \dots \mathrm{O}_n$ itself as a *collapsing modality* (for the logic in question) and more explicitly as a collapsing modality of length n. By way of stylistic variation, we also put this by saying that the modality in question "collapses" (in the given logic). In this section and the next, collapsing modalities in normal modal logics will be our concern, while in Section 3, the topic is pursued with respect to arbitrary congruential modal logics. Where O_i is $\Box$ (respectively, $\Diamond$), we write $\widetilde{\mathrm{O}}_i$—read "dual of O_i"—to denote $\Diamond$ (respectively, $\Box$). Similarly, if X is a nonnull affirmative modality (henceforth, modality) then $\widetilde{X}$ denotes the dual modality, that is, for $X = \mathrm{O}_1 \dots \mathrm{O}_n$, the modality $\widetilde{\mathrm{O}}_1 \dots \widetilde{\mathrm{O}}_n$. Note that X is a collapsing modality for a given logic just in case $\widetilde{X}$ is.

This last observation has several convenient consequences, three of which are the following. First, if some modality collapses in a logic, then that logic has a collapsing modality beginning with $\Box$, a collapsing modality beginning with $\Diamond$, a collapsing modality ending with $\Box$, and a collapsing modality ending with $\Diamond$. (Here we are thinking of $\mathrm{O}_1 \dots \mathrm{O}_n$ as beginning and ending with O_1 and O_n, respectively.) The reason is that if X is a collapsing modality which doesn't begin with $\Box$ (alternatively, with $\Diamond$), then we have available $\widetilde{X}$ as one which does—and similarly in the case of ending with $\Box$ (or $\Diamond$). Let us put one aspect of this to use at once.

Lemma 1.1 *Any normal modal logic with a collapsing modality extends* KD.

Proof Suppose our logic contains $p \leftrightarrow Xp$ for a (nonnull) affirmative modality X. By the above considerations, X can be chosen to begin with $\Diamond$, so putting $\top$ for p we can derive D, using the provability of $\top \to \Diamond\varphi$ in this logic, for any φ, and that of $\Diamond\varphi \to$ D in any (normal modal) logic. $\Box$

Secondly—though this is really a variation on the first point—the fact that every modality can be written (equivalently) in the form $X\Box$ means that a logic with a collapsing modality is one in which $\Box$ has X as a left inverse—an indigenous left inverse, as it is put in [9]—while in fact that every modality can be written in the form $\Box Y$ means that $\Box$ also has an indigenous right inverse (namely, Y) in any logic with a collapsing modality. The same points apply to $\Diamond$. More generally, whenever a modality XY collapses, the collapsing equivalence tells us that X is a left inverse of Y and Y a right inverse of X. These points explain the bearing of [9] on the question raised as a conjecture below, though in fact the 'inverses' theme will not occupy us explicitly further here.[5]

Thirdly, the fact that a modality collapses just in case its dual does means that if any modality of length 1 collapses in a logic, then so does every modality of length 1 (i.e., so do both such modalities)—this being essentially the familiar point that KT! $\supseteq$ KD!. Slightly less familiarly, if any modality of length 2 collapses in a logic then so do all modalities of length 2. The crucial—and readily verified (semantically)—fact here, noted in passing at p. 20 of [7], is that $\mathsf{KB}_c = \mathsf{KB}! = \mathsf{KC}_2$, and in particular that the second equation holds. (We return to the first in the following section.) This suggests the following conjecture, which will be our main concern in what follows: if a logic has any collapsing modality of length n, then all modalities of length n collapse in that logic. We extract two results from [9] bearing on this question, combining them as Proposition 1.2. Part (i) is Proposition 6.5 of [9], and part (ii) is Lemma 6.7 there.

Proposition 1.2

(i) *If* S *is a normal modal logic containing* $p \leftrightarrow \Box\varphi$ *for some formula* φ *of modal degree* m*, then* S *contains* $p \leftrightarrow \Box^{(m+1)!}p$*.*

(ii) *For all* $n \geq 1$, $\mathsf{KC}_n \supseteq \mathsf{KD}!$.

Part (ii) here has the consequence that if a logic has a homogeneous collapsing modality of length n (i.e., a modality $\mathrm{O}_1 \dots \mathrm{O}_n$ with $\mathrm{O}_1 = \cdots = \mathrm{O}_n$), then every modality of length n collapses in that logic, since the hypothesis here places the logic among the extensions of KC_n, so by the cited result, it also extends $\mathsf{KD}!$, which renders occurrences of $\Box$ and $\Diamond$ interchangeable, allowing us to trade in our homogeneous modality for an arbitrarily heterogeneous modality of the same length. But we cannot confirm the conjecture mooted above using just these resources, since if we begin by supposing that a logic, S say, has a collapsing modality of length n, we would need to know that it had a collapsing homogeneous modality of that length in order to draw the desired conclusion that all modalities of length n collapse in S. Proposition 1.2(i) offers some assistance in this direction: if S has a collapsing modality of length n, then, as already noted, it has a collapsing modality of length n which begins with a $\Box$. Let Y be the remainder of the modality in question. So $\vdash_{\mathsf{S}} p \leftrightarrow \Box Yp$. Taking Yp as the φ of 1.2(i), we conclude that $\vdash_{\mathsf{S}} p \leftrightarrow \Box^{(m+1)!}p$, where m is the modal degree of φ. In the current case of $\varphi = Yp$, the modal degree of φ is $n-1$, so putting in this as the value of m, we get that $\vdash_{\mathsf{S}} p \leftrightarrow \Box^{n!}p$. This is not quite what we needed, though, since we wanted n rather than $n!$ on the right. As it stands, it gives us only that $\mathsf{S} \supseteq \mathsf{KC}_{n!} \supseteq \mathsf{KD}!$, and hence by the previous reasoning that all modalities of length $n!$ collapse in S, rather than that all modalities of length n collapse in S. The earlier examples of cases favorable to the conjecture, for collapsing modalities of lengths 1 and 2, are conspicuously cases of n for which $n = n!$, and thus are equally compatible with the revised 'conjecture' (already known to be correct, in view of these considerations) that if a logic has a collapsing modality of length n, then all modalities of length $n!$ collapse in that logic. We want the stronger result that on the same hypothesis we can draw this conclusion for all modalities of length n itself. (This is stronger in the sense that, because $n!$ is a multiple of n, from the fact that all modalities of length n collapse we can conclude that the same goes for any modality X of length $n!$, since X can be written as $Y_1 \dots Y_{(n-1)!}$ with each Y_i a modality of length n. This can then be seen to collapse by $(n-1)!$ successive appeals to the collapsing equivalences for the Y_i.) So, for instance, while the estimate derived from Proposition 1.2 for the case of $n = 4$ is just that if some modality of length 4 collapses then all modalities of length 24 $(= 4!)$ collapse, whereas according to the conjecture we are keen to verify on this same hypothesis we should be able to conclude that (already) all modalities of length 4—and therefore also for any multiple thereof—collapse.

2 Settling the Question

To settle the conjecture raised in Section 1 for the case of a collapsing modality X of length n in a logic S, we need to attend separately to the two halves of the equivalence C_n, and to notice that already in S, one direction, namely, $\Box^n p \to p$ is already provable by appeal to D and the fact that $\mathsf{S} \subseteq \mathsf{KD}$ (by Lemma 1.1): just begin with $Xp \to p$ and note that $\vdash_{\mathsf{KD}} \Box^n p \to Xp$. This leaves the other direction of C_n to be shown to belong to S, before we can invoke Proposition 1.2(i) to conclude that all modalities of length n, being provably equivalent in S to each other and to the homogeneous $\Box$ (or for that matter $\Diamond$) modality of that length, and therefore all collapse.

Lemma 2.1 *If a normal modal logic* S *has a collapsing modality and* $\vdash_{\mathsf{S}} \Diamond\varphi$ *then* $\vdash_{\mathsf{S}} \varphi$.

Proof Suppose X collapses in S. As noted at the start of Section 1, without loss of generality we may assume that X is of the form $Y\Diamond$. Since S extends KD, by Lemma 1.1, the closure of S under Necessitation also implies its closure under 'Possibilitation' (prefixing a $\Diamond$), so if $\vdash_{\mathsf{S}} \Diamond\varphi$ we can successively prefix the last, then the second last, ..., and finally the first operator in Y to $\Diamond\varphi$, showing $Y\Diamond\varphi$, alias $X\varphi$ to be provable. The S-provability of φ then follows from the $X\varphi \to \varphi$ half of the collapsing equivalence for X (and closure under modus ponens). $\Box$

Lemma 2.2 *For any* $n \geq 1$ *and any normal modal logic* S, *if* $\vdash_{\mathsf{S}} \Box^n p \to p$ *then* $\vdash_{\mathsf{S}} \Diamond^n(p \to \Box^n p)$.

Proof Suppose that $\vdash_{\mathsf{S}} \Box^n p \to p$; thus by duality $\vdash_{\mathsf{S}} p \to \Diamond^n p$, and so by uniform substitution of $\Box^n p$ for p, we get $\vdash_{\mathsf{S}} \Box^n p \to \Diamond^n\Box^n p$, which by the K-equivalence of the formulas involved, gives us the desired conclusion: $\vdash_{\mathsf{S}} \Diamond^n(p \to \Box^n p)$. $\Box$

We are now in a position to conclude that either no modalities of a given length collapse in a logic or else all modalities of that length do, confirming the conjecture raised in Section 1.

Theorem 2.3 *If* S *is a normal modal logic with a collapsing modality of length n, then all modalities of length n collapse in* S.

Proof Suppose a modality of length n collapses in S. By the reasoning of the opening paragraph of this section, $\vdash_{\mathsf{S}} \Box^n p \to p$. Therefore we also have $\vdash_{\mathsf{S}} \Diamond^n(p \to \Box^n p)$ by Lemma 2.2, and so by n appeals to Lemma 2.1, we get $\vdash_{\mathsf{S}} p \to \Box^n p$. Thus $\mathsf{S} \supseteq \mathsf{KC}_n$ and thus S proves the equivalence of any two modalities of length n (by Proposition 1.2(ii)), and thereby that all modalities of length n collapse in S, since $\Box^n$ collapses in S. $\Box$

Thus the normal extension of K by collapsing equivalence for a modality of length n produces the logic KC_n in which all modalities of length n collapse (since C_n collapses the homogeneous modalities of this length while Proposition 1.2(ii) assures us we are in an extension of KD!, which brings all the heterogeneous modalities of the same length along). Having settled the question raised in Section 1, our main business is over. But another observation lies sufficiently close to hand to be worth making. It was mentioned in passing between Lemma 1.1 and Proposition 1.2 that *one half* of a collapsing equivalence, namely, the B_c-half of B! suffices for all modalities of length 2 to collapse. As it was put above, $\mathsf{KB}_c = \mathsf{KB!} = \mathsf{KC}_2$. Here we give a simple generalization of this point. The fact that B is K-deducible from B_c is the $k = 1$ case (though here we are thinking of the dual form of B, as described in the opening paragraph of Section 1).

Proposition 2.4 *Any normal modal logic containing the formula $p \to \Diamond\Box^k p$ (any $k \geq 1$) also contains the converse of this formula.*

Proof First note that if normal S proves $p \to \Diamond\Box^k p$, by (closure under) uniform substitution $\vdash_{\mathsf{S}} \varphi \to \Diamond\Box^k\varphi$ for all φ; by taking the case of φ as $\top$ we conclude that $\mathsf{S} \supseteq \mathsf{KD}$. From this we conclude, for later use, that $\vdash_{\mathsf{S}} p \to \Diamond^k\Box p$, by weakening each of the first $k - 1$ occurrences of $\Box$ in the S-provable $p \to \Diamond\Box^k p$ to a $\Diamond$. Next take φ as $\Diamond\Box^k p \wedge \neg p$:

$$(\Diamond\Box^k p \wedge \neg p) \to \Diamond\Box^k(\Diamond\Box^k p \wedge \neg p),$$

so by normality, which allows us to distribute the outer "$\Box^k$" in the consequent across to the conjuncts, we get the following (in which the underscoring is for later reference), as S-provable:

$$(\Diamond\Box^k p \wedge \neg p) \to \Diamond(\underline{\Box^k\Diamond}\Box^k p \wedge \Box^k\neg p).$$

Recall that $\vdash_{\mathsf{S}} p \to \Diamond^k\Box p$, and so by duality (and uniform substitution), we have $\Box^k\Diamond\varphi$ provably implying φ in S, for any φ, which allows us to replace the former by the latter in any "positive" context, as in the first conjunct after the "$\Diamond$" in the consequent of the last inset formula above. That is, we can drop the underscored modality, concluding to the S-provability of

$$(\Diamond\Box^k p \wedge \neg p) \to \Diamond(\Box^k p \wedge \Box^k\neg p).$$

But the negation of the consequent here is evidently KD-provable, and hence S-provable, so the negation of the antecedent is S-provable, giving us the desired conclusion, that $\vdash_{\mathsf{S}} \Diamond\Box^k p \to p$. $\Box$

Thus we have pleasantly economical axiomatizations for the logics KC_n ($n \geq 2$), which are the weakest logics in which every modality of length n collapses.

Corollary 2.5 *For $n \geq 2$, KC_n is the smallest normal modal logic containing $p \to \Diamond\Box^{n-1}p$.*

Proof Clearly $\vdash_{\mathsf{KC}_n} p \to \Diamond\Box^{n-1}p$, since KC_n proves the corresponding $\leftrightarrow$-formula (for $\Diamond\Box^{n-1}$ as for all modalities of length n). It remains to show that no weaker normal modal logic proves the formula in question. Letting S be any such logic, then by Proposition 2.4, since $\vdash_{\mathsf{S}} p \to \Diamond\Box^{n-1}p$, we conclude that $\vdash_{\mathsf{S}} p \leftrightarrow \Diamond\Box^{n-1}p$. So by Theorem 2.3, every modality of length n collapses in S, including the homogeneous ones, giving us the conclusion that $\mathsf{S} \supseteq \mathsf{KC}_n$. $\Box$

3 Collapsing Modalities in Congruential Logics

We have been concentrating on normal modal logics and turn here to the broader class of congruential modal logics (those containing $\Box\varphi \leftrightarrow \Box\psi$ whenever they contain $\varphi \leftrightarrow \psi$ (for any formulas φ, ψ). Surprisingly many of the "If X collapses then Y collapses" claims from the former arena turn out to be correct in this broader setting, as we shall see, even though an exact analogue of Theorem 2.3 is not forthcoming.

The only modalities of length 1 are $\Box$ and $\Diamond$, and as before, either of these collapses just in case the other does. When we turn to length 2, duality again delivers the conclusion that $\Box\Diamond$ collapses just in case $\Diamond\Box$ does, and similarly for the homogeneous cases $\Box\Box$ and $\Diamond\Diamond$, but now the relation between the homogeneous and heterogeneous pairs is disrupted. Let us observe first that if the homogeneous length-2 modalities collapse in a congruential modal logic, this does not imply that the heterogeneous modalities of the same length collapse. As in Chellas [4], $\mathsf{E}\lambda$ is the smallest congruential modal logic containing the formula (or all instances of the schema) to which the label λ has been assigned.

Example 3.1 To see that $\nvdash_{\mathsf{EC}_2} \mathsf{B!}$, consider models $\mathcal{M} = \langle W, w^*, V\rangle$ with $w^* \in W$ and V assigning subsets of W to the propositional variables, providing the basis case of the usual inductive definition of truth of a formula φ at a point $x \in W$ in the model $\mathcal{M}$ (notation: $\mathcal{M} \models_x \varphi$). Although each of these models has a distinguished element (namely, w^*) we use this only in the course of the definition of truth, and not in the account of validity, which is truth at all points in all models.[6] The inductive clauses in the truth-definition are as usual for the Boolean connectives, while for $\Box$ we say (for $x \in W$, with $\mathcal{M}$ as above)

$$\mathcal{M} \models_x \Box\varphi \text{ if and only if either } x = w^* \text{ and } \mathcal{M} \models_x \varphi,$$
$$\text{or } x \neq w^* \text{ and: } \mathcal{M} \models_x \varphi \text{ iff } \mathcal{M} \models_{w^*} \varphi.$$

Clearly, the set of formulas valid in the sense just indicated is a congruential modal logic and with a little work one sees that it contains C_2 but not $\mathsf{B!}$, showing that $\nvdash_{\mathsf{EC}_2} \mathsf{B!}$ and thus that homogeneous collapse does not imply heterogeneous collapse (for length 2 modalities) among congruential modal logics.

To see the converse, that heterogeneous collapse does not yield homogeneous collapse in arbitrary congruential modal logics, for modalities of length 2, stick with the above semantics for $\Box$ and define the operator $\Box'$ by $\Box'\varphi = \Box\neg\varphi$ and observe that now the valid formulas comprise a congruential modal logic for $\Box'$ and include $\mathsf{B!}$ but not C_2.

To consider further the collapse relations among modalities (in congruential logics), some abbreviative notation will be useful. For modalities X and Y, let us write

$$X \gg Y$$

to mean that in any congruential modal logic in which X collapses, Y also collapses. (Think of "$\gg$" as reminiscent of a kind of implicational arrow, rather than as any kind of *greater than* symbol.) Let us collect here some simple observations concerning duality, the modalities in which one modality is sandwiched between two copies of another, those in which one modality is sandwiched between another and the dual of the latter, and finally (parts (iv) and (v)) cases in which only a modality and its dual figure.

Proposition 3.2 *For any modalities X and Y,*

(i) $X \gg \widetilde{X}$;

(ii) $XYX \gg YXX$ *and* $XYX \gg XXY$;

(iii) $XY\widetilde{X} \gg XY\widetilde{Y}$ *and* $XY\widetilde{X} \gg \widetilde{Y}YX$.

(iv) $XX\widetilde{X} \gg X\widetilde{X}X$.

(v) $X\widetilde{X}\widetilde{X} \gg XX\widetilde{X}$.

Proof (i) is clear—and was already mentioned in the second paragraph of this section; (ii) and (iii) just involve straightforward (uniform) substitutions and replacements (of equivalents by equivalents), using (i) in addition for the proof of (iii). We illustrate with the second assertion under (ii), and with (iv) and (v), beginning with the first of these. Suppose for congruential S that XYX collapses in S, with a view to showing that XXY does also. The supposition means that we have as S-provable the formula $XYXp \leftrightarrow p$. Substitute YXp for p, getting $XYXYXp \leftrightarrow YXp$, and then replace the $XYXp$ in left-hand formula with (the S-provably equivalent) p, giving $XYp \leftrightarrow YXp$. Now embed both sides in the scope of X, getting an equivalence between $XXYp$ and $XYXp$, and conclude that since XYX collapses in S, so does XXY. Turning to (iv) we record the initial assumption and its dual (available by (i)) as (1) and (1δ):

(1) $\vdash_{\mathsf{S}} XX\widetilde{X}p \leftrightarrow p$ (1δ) $\vdash_{\mathsf{S}} \widetilde{X}\widetilde{X}Xp \leftrightarrow p$.

Substituting $\widetilde{X}Xp$ for p in (1),

(2) $\vdash_{\mathsf{S}} XX\widetilde{X}\widetilde{X}Xp \leftrightarrow \widetilde{X}Xp$,

and making a replacement licensed by (1δ) in (2),

(3) $\vdash_{\mathsf{S}} XXp \leftrightarrow \widetilde{X}Xp$,

we now substitute $\widetilde{X}p$ for p:

(4) $\vdash_{\mathsf{S}} XX\widetilde{X}p \leftrightarrow \widetilde{X}X\widetilde{X}p$.

Thus by (4) and (1),

(5) $\vdash_{\mathsf{S}} p \leftrightarrow \widetilde{X}X\widetilde{X}p$,

and so by duality (i.e., invoking (i)),

(6) $\vdash_{\mathsf{S}} p \leftrightarrow X\widetilde{X}Xp$.

Finally for (v) we begin similarly,

(1) $\vdash_{\mathsf{S}} X\widetilde{X}\widetilde{X}p \leftrightarrow p$ (1δ) $\vdash_{\mathsf{S}} \widetilde{X}XXp \leftrightarrow p$.

Substituting XXp for p in (1),

(2) $\vdash_{\mathsf{S}} X\widetilde{X}\widetilde{X}XXp \leftrightarrow XXp$.

So by a replacement in (2) justified by (1δ),

(3) $\vdash_{\mathsf{S}} X\widetilde{X}p \leftrightarrow XXp$.

Substituting $\widetilde{X}p$ for p,

(4) $\vdash_{\mathsf{S}} X\widetilde{X}\widetilde{X}p \leftrightarrow XX\widetilde{X}p$,

so in view of (1), $XX\widetilde{X}$ collapses in S. □

Such substitution-and-replacement proofs as are illustrated above, in what we might call the 'calculus of (congruential) modalities', could equally be notated as proofs of quasi-identities (conditional equations) for monoids—whose unit (two-sided identity element) we write as 1; stopping short at the point in the proof of the second claim in (ii) above where the hypothesis that XYX collapses is seen to lead to the conclusion that $XYp \leftrightarrow p$ is provable, we should have a proof of the quasi-identity $xyx \approx 1 \Rightarrow xy \approx yx$ (semigroup operation indicated by juxtaposition, parentheses omitted in view of associativity) and, continuing on, of the quasi-identity $xyx \approx 1 \Rightarrow x^2y \approx 1$. (To accommodate duality (as in 3.2(i) and (iii)) we need monoids expanded by an involution $\widetilde{\cdot}$ of period 2 satisfying the identities $\widetilde{1} \approx 1$ and $\widetilde{xy} \approx \widetilde{x}\widetilde{y}$.) The fact that in the calculus of modalities we are dealing with a monoid with a single generator ("$\Box$"), since we are considering monomodal logic, does not affect the proofs of such "$\gg$"-statements as those in Proposition 3.2. We shall continue to write such proofs in the conventional notation of modal propositional logic.

Although Proposition 3.2 is very general, let us look at one special case for modalities of length 3, to show that just as if either of the two heterogeneous modalities of length 2 collapses (in a congruential modal logic), so does the other—trivially in that case, by duality—so at length 3 also, if any of the six modalities of length 3 collapses, then so do the other five. We use this result to obtain something more general and contrasting with the length 2 case, below. (The more general result, which drops the condition that Y is heterogeneous, could be obtained now, but it suits our expository purposes to postpone it slightly.) In each case the variables ranging over arbitrary modalities are instantiated to the special case of modalities of length 1 (i.e. modal operators).

Lemma 3.3 *If X and Y are heterogeneous modalities of length 3 then $X \gg Y$.*

Proof Since the relation $\gg$ is transitive, it suffices to establish the cycle

$$\mathrm{OO}\widetilde{\mathrm{O}} \gg \mathrm{O}\widetilde{\mathrm{O}}\mathrm{O} \gg \mathrm{O}\widetilde{\mathrm{O}}\widetilde{\mathrm{O}} \gg \mathrm{OO}\widetilde{\mathrm{O}},$$

where we write "O" for either of $\Box$, $\Diamond$ (so that each length-3 heterogeneous modality is of one of the forms $\mathrm{OO}\widetilde{\mathrm{O}}$, $\mathrm{O}\widetilde{\mathrm{O}}\mathrm{O}$, $\mathrm{O}\widetilde{\mathrm{O}}\widetilde{\mathrm{O}}$). The first $\gg$-link is given by Proposition 3.2(iv), X as O. For the second link, use the first assertion under Proposition 3.2(ii), with X as O and Y as $\widetilde{\mathrm{O}}$. For the third, take X as O in Proposition 3.2(v). $\Box$

We return to the simple case of modalities of length 3 presently, first looking at something much more general, illustrated in that special case by Proposition 3.2(v), according to which $X\widetilde{X}\widetilde{X} \gg XX\widetilde{X}$.

Proposition 3.4 *For any modality X, $X^m\widetilde{X}^n \gg X^n\widetilde{X}^m$.*

Proof We consider two cases: (i) $m \leq n$; (ii) $m > n$.

Case (i) Let $n = m + k$ ($k \geq 0$); thus we are to suppose that

(1) $\vdash_{\mathsf{S}} X^m\widetilde{X}^{m+k}p \leftrightarrow p$ for some congruential modal logic S, so by duality,

(1δ) $\vdash_{\mathsf{S}} \widetilde{X}^m X^{m+k}p \leftrightarrow p$.

Substituting $X^{m+k}p$ for p in (1) we get

(2) $\vdash_{\mathsf{S}} X^m\widetilde{X}^{m+k}X^{m+k}p \leftrightarrow X^{m+k}p$, that is,

(3) $\vdash_{\mathsf{S}} X^m\widetilde{X}^k\widetilde{X}^m X^{m+k}p \leftrightarrow X^{m+k}p$,

so, dropping the $\widetilde{X}^m X^{m+k}$, in accordance with (1δ), from (3) we have

(4) $\vdash_{\mathsf{S}} X^m \widetilde{X}^k p \leftrightarrow X^{m+k} p$.

Substituting $\widetilde{X}^m p$ for p in (4),

(5) $\vdash_{\mathsf{S}} X^m \widetilde{X}^k \widetilde{X}^m p \leftrightarrow X^{m+k} \widetilde{X}^m p$, that is,

(6) $\vdash_{\mathsf{S}} X^m \widetilde{X}^{m+k} p \leftrightarrow X^{m+k} \widetilde{X}^m p$, whence by (1) we conclude

(7) $\vdash_{\mathsf{S}} p \leftrightarrow X^{m+k} \widetilde{X}^m p$,

which is to say $X^n \widetilde{X}^m$ collapses in S, completing the proof for Case (i).

We turn to Case (ii), letting $m = n + k$ $(k > 0)$. Our supposition is that for a congruential modal logic S we have (starting the numbering afresh)

(1) $\vdash_{\mathsf{S}} X^{n+k} \widetilde{X}^n p \leftrightarrow p$, and therefore by duality also,

(1δ) $\vdash_{\mathsf{S}} \widetilde{X}^{n+k} X^n p \leftrightarrow p$.

Substituting $\widetilde{X}^k X^n p$ for p in (1),

(2) $\vdash_{\mathsf{S}} X^{n+k} \widetilde{X}^n \widetilde{X}^k X^n p \leftrightarrow \widetilde{X}^k X^n p$, that is,

(3) $\vdash_{\mathsf{S}} X^{n+k} \widetilde{X}^{n+k} X^n p \leftrightarrow \widetilde{X}^k X^n p$,

so exploiting (1δ) to drop the "$\widetilde{X}^{n+k} X^n$",

(4) $\vdash_{\mathsf{S}} X^{n+k} p \leftrightarrow \widetilde{X}^k X^n p$.

Dualizing (4), we get

(5) $\vdash_{\mathsf{S}} \widetilde{X}^{n+k} p \leftrightarrow X^k \widetilde{X}^n p$,

applying X^n to both side of which yields

(6) $\vdash_{\mathsf{S}} X^n \widetilde{X}^{n+k} p \leftrightarrow X^{n+k} \widetilde{X}^n p$.

So by appealing to (1), we conclude

(7) $\vdash_{\mathsf{S}} X^n \widetilde{X}^{n+k} p \leftrightarrow p$.

$\square$

The "swapping of exponents" phenomenon described in Proposition 3.4 can be used to convert a collapsing heterogeneous modality into a collapsing homogeneous modality of the same length under certain conditions. In particular, under Case (i) of the proof, when $k = 1$ and X itself is homogeneous, such a conversion is available.

Proposition 3.5 $\mathrm{O}^m \widetilde{\mathrm{O}}^{m+1} \gg \mathrm{O}^{2m+1}$ *for* $\mathrm{O} \in \{\Box, \Diamond\}$.

Proof Our starting hypothesis is that, for some congruential S,

(1) $\vdash_{\mathsf{S}} \mathrm{O}^m \widetilde{\mathrm{O}}^{m+1} p \leftrightarrow p$.

If $m = 0$ then we can stop here, since (1) then says that $\widetilde{\mathrm{O}}$ collapses, in which case by duality so does $\mathrm{O} = \mathrm{O}^{2m+1}$, as required. So we may assume that $m \geq 0$. According to line (4) of the proof, under Case (i) of Proposition 3.4 above, we have

(2) $\vdash_{\mathsf{S}} \mathrm{O}^m \widetilde{\mathrm{O}} p \leftrightarrow \mathrm{O}^{m+1} p$,

and by Proposition 3.4 itself, from (1) we have

(3) $\vdash_{\mathsf{S}} \mathrm{O}^{m+1}\widetilde{\mathrm{O}}^{m} p \leftrightarrow p$,

which we can write as (4), using parentheses to highlight the key submodality,

(4) $\vdash_{\mathsf{S}} \mathrm{O}(\mathrm{O}^{m}\widetilde{\mathrm{O}})\widetilde{\mathrm{O}}^{m-1} p \leftrightarrow p$.

Replacing the parenthesized portion in accordance with (2), we get

(5) $\vdash_{\mathsf{S}} \mathrm{O}\mathrm{O}^{m+1}\widetilde{\mathrm{O}}^{m-1} p \leftrightarrow p$, that is,

(6) $\vdash_{\mathsf{S}} \mathrm{O}^{m+2}\widetilde{\mathrm{O}}^{m-1} p \leftrightarrow p$.

If $m-1 = 0$, we can stop, since we have reached a homogeneous collapsing modality of the desired length (as in the $m = 0$ case above). If $m - 1 \geq 1$, we continue in the same way, thinking of (6) as (7):

(7) $\vdash_{\mathsf{S}} \mathrm{O}^{2}(\mathrm{O}^{m}\widetilde{\mathrm{O}})\widetilde{\mathrm{O}}^{m-2} p \leftrightarrow p$,

and invoke (2) again

(8) $\vdash_{\mathsf{S}} \mathrm{O}^{2}\mathrm{O}^{m+1}\widetilde{\mathrm{O}}^{m-2} p \leftrightarrow p$, that is,

(9) $\vdash_{\mathsf{S}} \mathrm{O}^{m+3}\widetilde{\mathrm{O}}^{m-2} p \leftrightarrow p$,

and so on, until finally the exponent on the "$\widetilde{\mathrm{O}}$" reaches 0, giving us the desired homogeneous collapsing modality. □

We could very simply have observed earlier—see the discussion after the proof below—that our various heterogeneous modalities of length 3 stand in the $\gg$ relation to the homogeneous modalities of the same length, but Propositions 3.4 and 3.5 seem of sufficient general interest and broader applicability to take the detour through them to draw this conclusion for the 'toy' case of length 3.

Proposition 3.6 *If X and Y are modalities of length 3 and X is heterogeneous, then $X \gg Y$.*

Proof According to Lemma 3.3, all the heterogeneous modalities of length 3 stand in the $\gg$ relation to each other, so it remains only to find one, X, of length 3, and show that $X \gg \Box\Box\Box$, to establish the present result. By Proposition 3.5, we can choose this X as $\Box\Diamond\Diamond$. □

We turn to the promised simpler (though perhaps less widely generalizable) route to the above conclusion, which shares the stress on $\mathsf{D!}$ that we found in Section 2, only now it is $\mathsf{ED!}$ rather than $\mathsf{KD!}$ that is pertinent.

Proposition 3.7 *If the modality $\Box\Diamond\Box$ collapses in a congruential modal logic S (or equivalently, if $\Diamond\Box\Diamond$ collapses in S, then $\mathsf{S} \supseteq \mathsf{ED!}$ and thus (since $\Box$ and $\Diamond$ are now freely interreplaceable in S) all modalities of length 3 are S-equivalent and all collapse in S.*

Proof It suffices to show that on the hypothesis described, $\vdash_{\mathsf{S}} \mathsf{D!}$. So make the hypothesis, along with its dual form:

(1) $\vdash_{\mathsf{S}} \Box\Diamond\Box p \leftrightarrow p$ (1δ) $\Diamond\Box\Diamond p \leftrightarrow p$,

and substitute $\Diamond p$ for p in (1),

(2) $\vdash_{\mathsf{S}} \Box\Diamond\Box\Diamond p \leftrightarrow \Diamond p$.

By (1δ) we can drop the "$\Diamond\Box\Diamond$" from left-hand side, ending up with $\mathsf{D!}$:

(3) $\vdash_{\mathsf{S}} \Box p \leftrightarrow \Diamond p.$

$\Box$

Thus a quick proof of Proposition 3.6 would consist of a proof of part of Lemma 3.3 showing that for every heterogeneous length 3 modality, X, $X \gg \Diamond\Box\Diamond$; this together with Proposition 3.7 then yields Proposition 3.6. It is interesting to ponder how to generalize the above proof of Proposition 3.7. An analysis of the argument shows that to obtain a proof of $\mathsf{D!}$ along these lines we must have a collapsing modality of the form $\mathrm{O}X\mathrm{O}$ in which $X\mathrm{O} = \widetilde{\mathrm{O}}\widetilde{X}$ (where this "=" means syntactic identity, not mere provable equivalence). Thus letting X be $\mathrm{O}_1 \ldots \mathrm{O}_m$, we must have

$$\mathrm{O}_1 \ldots \mathrm{O}_m\mathrm{O} = \widetilde{\mathrm{O}}\widetilde{\mathrm{O}}_1 \ldots \widetilde{\mathrm{O}}_m;$$

so for k with $1 \le k \le m$, we must have $\mathrm{O}_{k+1} = \widetilde{\mathrm{O}}_k$, understanding O_{k+1} to be O. Thus it never happens that $\mathrm{O}_k = \mathrm{O}_{k+1}$, so X must be of odd length and consist of alternating occurrences of $\Box$ and $\Diamond$, beginning and ending with $\widetilde{\mathrm{O}}$, the initial collapsing modality $\mathrm{O}X\mathrm{O}$ being of the form $(\mathrm{O}\widetilde{\mathrm{O}})^{\ell}\mathrm{O}$. (In terms of k as the length of X, $\ell = (k+1)/2$.) The main point we may extract from this discussion of the proof of Proposition 3.7 we state as the following corollary.

Corollary 3.8 *For every odd m there is a modality X of length m such that if X collapses in a congruential modal logic* S, *then* $\mathsf{S} \supseteq \mathsf{ED!}$ *and every modality of length m collapses in* S.

According to Proposition 3.6, whichever style of proof for it one prefers, we can drop the heterogeneity restriction on Y in Lemma 3.3 and still have the conclusion that $X \gg Y$, but can we drop the restriction on X, which persists in Proposition 3.6? The answer is *no*, as a simple example given here shows. (Unlike the case of Example 3.1, this one did not seem amenable to as simple a presentation in model-theoretic terms, though a somewhat convoluted formulation along similar lines is certainly possible.)

Example 3.9 Consider the four-valued matrix which is the product of the two-element matrix (for a functionally complete set of Boolean connectives) with itself, expanded by a table (described below) for $\Box$. Thus the values are $\langle T, T\rangle$, $\langle T, F\rangle$, $\langle F, T\rangle$, $\langle F, F\rangle$ which we call 1, 2, 3, and 4, respectively. The designated element is 1. For the function (associated with) $\Box$ we have

$$\Box : 1 \mapsto 3, 2 \mapsto 1, 3 \mapsto 2, 4 \mapsto 4.$$

From this it is clear that $\Box^3(x) = x$ ($x = 1, 2, 3, 4$), so the homogeneous collapsing equivalence (in $\Box$) of length 3 is valid in the matrix. But $\Box\Box\neg\Box\neg p \leftrightarrow p$ is not similarly valid, since when p is assigned the value 1, the formula on the left receives the value 2, giving the whole equivalence the (undesignated) value 2. After checking that the logic determined by the matrix (i.e., the set of formulas valid therein) is congruential, one concludes that here we have a congruential logic refuting the hypothesis that $\Box\Box\Box \gg \Box\Box\Diamond$. (Note that we are concerned here only with the formula logic determined by the matrix, not with the consequence relation determined by it. There is a notion of congruentiality for consequence relations, or more specifically for connectives (here $\Box$) with respect to consequence relations, which in the present case, writing "$\models$" for the consequence relation determined by the above matrix, would demand that $\varphi \dashv\models \psi$ should imply $\Box\varphi \dashv\models \Box\psi$. This demand is not met in the present instance. Since the values of φ and $\Box\varphi$ are always different and we have only one designated value, they can never both be designated, which suffices to guarantee that $\varphi \wedge \Box\varphi$ never assumes a designated value, implying that $p \wedge \Box p \dashv\models q \wedge \Box q$. But we do not have $\Box(p \wedge \Box p) \dashv\models \Box(q \wedge \Box q)$: for example, assign p and q, respectively, the values 2 and 1 to refute the forward, or "$\models$", direction. By contrast, another congruentiality-related condition—sometimes called extensionality—namely, that $\varphi \leftrightarrow \psi \models \Box\varphi \leftrightarrow \Box\psi$ for all φ, ψ, *is* satisfied, which has the consequence that the congruentiality condition of interest in this example—that $\varnothing \models \varphi \leftrightarrow \psi$ should imply $\varnothing \models \Box\varphi \leftrightarrow \Box\psi$—is also satisfied.)

The intersection of $\gg$ with its converse is an equivalence relation on the set of modalities.[7] At the start of this section, we saw that the length 2 modalities form two equivalence classes—the heterogeneous modalities comprising one and the homogeneous modalities comprising the other—with respect to this relation, with no $\gg$ relations from the one class to the other. (For length 1, there is just one equivalence class, as in normal modal logics for any n, according to Theorem 2.3, though here we are of course reconstruing the definition of $\gg$ so that "$X \gg Y$" means the collapse of X guarantees that of Y in any normal modal logic.) We have now seen that the situation at length 3 is rather different. There are again two equivalence classes, heterogeneous and homogeneous, but the $\gg$ relation holds in one direction between them (or more accurately, their elements)—from the former to the latter. The obvious question with which to close is this: What is the general picture for modalities of length n—that is, how many equivalence classes are there (as a function of n) and what are the $\gg$ relations among them? Corollary 3.8 suggests there may be a significant difference between the case of odd and even n, but this—and much else—remains to be seen.

Notes

1. Below, we also use "!" in its arithmetical sense, for the factorial function. No confusion will arise. Note that for present purposes D should be understood as the formula $\Box p \to \Diamond p$—or the corresponding schema—rather than the formula $\Diamond(p \vee \neg p)$ or $\Diamond\top$ ($\top$ a truth constant). $\Diamond$ here is regarded as abbreviating $\neg\Box\neg$.

2. Exception: Examples 3.1 and 3.9 are described semantically, one using model-theoretic and the other using matrix-theoretic apparatus, in order to show the unprovability of certain formulas.

3. Pursued, for example, in [10] and [3] for specific logics (listed in chronological order), and more generally in [5], [2], and [1].

4. An example of this usage may be found in [6]; the phrase 'collapse identities' is also used in the same sense, as is the phrase 'absorptive identities'.

5. This theme surfaced originally in [11], and, after [9], received further attention in [12].

6. The situation in this respect is as for the treatment of modal actuality logic described, for instance, in §2 of [8], where one has an "actually" operator A for which $A\varphi$ is deemed to be true at any point in a model just in case φ is true at w^*. A definitionally equivalent alternative would use an operator L with $L\varphi$ meaning that things stand with respect to

φ as they do in the actual world; that is, $L\varphi$ is true at x just in case φ is true at both x and w^* or at neither of them. (Thus A can be recovered from primitive L by defining $A\varphi = (L\varphi \leftrightarrow \varphi)$.) The condition for $\Box$ described in the present example is a variation on L, behaving like L at all $x \neq w^*$ but being an 'identity connective' at w^* itself. (L itself behaves like the 'constant true' operator at w^*.)

7. The word "equivalent" in the phrase "equivalent modalities", as customarily used—for example, in the references cited in note 3—does not stand for this relation, of course, but for the logic-relative relation of the provable equivalence of the results of attaching the modalities in question to an arbitrary formula; the current equivalence relation is rather that of being 'equi-collapsing' over a range of logics (for the purposes of this section, over the class of congruential modal logics.

References

[1] Bellissima, F., and M. Mirolli, "A general treatment of equivalent modalities," *The Journal of Symbolic Logic*, vol. 54 (1989), pp. 1460–71. Zbl 0706.03017. MR 1026610. 130

[2] van Benthem, J. F. A. K., "Modal reduction principles," *The Journal of Symbolic Logic*, vol. 41 (1976), pp. 301–12. Zbl 0337.02014. MR 0409111. 130

[3] Chellas, B., "Modalities in normal systems containing the **S5** axiom," pp. 261–65 in *Intention and Intentionality: Essays in Honour of G. E. M. Anscombe*, edited by C. Diamond and J. Teichman, Harvester Press, Brighton, 1979. 130

[4] Chellas, B. F., *Modal Logic. An Introduction*, Cambridge University Press, Cambridge, 1980. Zbl 0431.03009. MR 556867. 119, 124

[5] Fitch, F. B., "A correlation between modal reduction principles and properties of relations," *Journal of Philosophical Logic*, vol. 2 (1973), pp. 97–101. Zbl 0259.02012. MR 0414319. 130

[6] Giesl, J., and D. Kapur, "Dependency pairs for equational rewriting," pp. 93–107 in *Rewriting Techniques and Applications (Utrecht, 2001)*, edited by A. Middeldorp, vol. 2051 of *Lecture Notes in Computer Science*, Springer, Berlin, 2001. Zbl 0981.68063. MR 1892391. 130

[7] Humberstone, L., "Negation by iteration," *Theoria*, vol. 61 (1995), pp. 1–24. Zbl 0883.03007. MR 1619613. 121

[8] Humberstone, L., "Two-dimensional adventures," *Philosophical Studies*, vol. 118 (2004), pp. 17–65. MR 2068363. 130

[9] Humberstone, L., and T. Williamson, "Inverses for normal modal operators," *Studia Logica*, vol. 59 (1997), pp. 33–64. Zbl 0887.03015. MR 1470208. 119, 120, 121, 130

[10] Parry, W. T., "Modalities in the *survey* system of strict implication," *The Journal of Symbolic Logic*, vol. 4 (1939), pp. 137–54. Zbl 0023.09902. MR 0000809. 130

[11] Williamson, T., "Two incomplete anti-realist modal epistemic logics," *The Journal of Symbolic Logic*, vol. 55 (1990), pp. 297–314. Zbl 0706.03019. MR 1043559. 130

[12] Williamson, T., "Continuum many maximal consistent normal bimodal logics with inverses," *Notre Dame Journal of Formal Logic*, vol. 39 (1998), pp. 128–34. Zbl 0968.03027. MR 1671746. 130

Acknowledgments

For numerous corrections, I am grateful to Steve Gardner.

School of Philosophy and Bioethics
Monash University
Clayton VIC 3800
AUSTRALIA
lloyd.humberstone@arts.monash.edu.au