How Not to Think About Modal Definability: A Modal Axiom from G. E. Hughes

Lloyd Humberstone
Monash University

Abstract

In a 1990 paper, George Hughes axiomatized the logic determined by the class of all frames in which each point has a reflexive successor, and raised various questions along the way, one of which is answered *incorrectly* here by means of an interestingly fallacious argument.

1 Background

Hughes [3] considers the following axioms

$\Diamond\mathbf{T}_n$ $\quad\Diamond((\Box p_1 \to p_1) \land \ldots \land (\Box p_n \to p_n))$

and the following modal logics: $\mathbf{K}\Diamond\mathbf{T}_n$, the smallest normal modal logic containing $\Diamond\mathbf{T}_n$, for $n \in \mathit{Nat}$, and $\mathbf{K}\Diamond\mathbf{T}$, the smallest normal modal logic containing each $\Diamond\mathbf{T}_n$.[1] Evidently

$$\mathbf{K}\Diamond\mathbf{T}_1 \subseteq \mathbf{K}\Diamond\mathbf{T}_2 \subseteq \ldots \subseteq \mathbf{K}\Diamond\mathbf{T}_n \subseteq \ldots \subseteq \mathbf{K}\Diamond\mathbf{T}$$

and Hughes shows (Theorem 7 of [3]) that all these inclusions are strict. Theorem 1 of [3] establishes, by a canonical model argument, that $\mathbf{K}\Diamond\mathbf{T}$ is determined by (i.e., is sound and complete w.r.t.) the class of all frames $\langle W, R\rangle$ in which each point has a reflexive successor – that is, frames satisfying $\forall x\exists y(Rxy\,\&\,Ryy)$, and that many frames for $\mathbf{K}\Diamond\mathbf{T}$ lie outside this class. He mentions among his examples the frame $\langle \mathit{Nat}, <\rangle$, which is suggestive

[1] In [3] the letter 'M' replaces what we have written here as $\Diamond$; we make this replacement to avoid confusion with McKinsey's axiom, also often known as M (or **M** or *M*). On the latter usage **KMT** (or **KTM**) would be the smallest normal modal logic containing **M** ($= \Box\Diamond p \to \Diamond\Box p$) and **T** ($= \Box p \to p$), a very different matter from $\mathbf{K}\Diamond\mathbf{T}$ (called *KMT* in [3]). In fact the completeness result from [3] mentioned next has as its prototype the treatment in Lemmon and Scott [5], p. 74*f.*, of a logic with axioms $\mathbf{M}_n$ related to **M** in the same way that the formulas $\mathbf{T}_n$ are related to **T**. Here we write p (and sometimes below q) for the first (and second) in the list of all propositional variables (sentence letters) $p_1, p_2, \ldots, p_n, \ldots$, and in modal formulas notate conjunction and material implication by means of "$\land$" and "$\to$"; in (non-modal) first and second order formulas below we use instead "&" and "$\Rightarrow$". Note that on the present understanding of a normal modal logic we require closure under uniform substitution (of arbitrary formulas for such variables).

of the following result, not mentioned by Hughes but already (then) well known, having been proved by R. Goldblatt and S. Thomason in 1974.[2] We give a simpler proof of this fact than those in the literature, which makes use of the fact that $\mathbf{K\Diamond T} \subseteq \mathbf{KD4}$.[3]

Theorem 1.1 *The class of frames in which every point has a reflexive successor is not modally definable.*

Proof. Suppose, for a contradiction, that there is some set Γ of modal formulas valid on precisely the frames in question. Since $\mathbf{K\Diamond T} \subseteq \mathbf{KD4}$ Hughes's proof of Theorem 1 in [3] shows that in the canonical frame for **KD4**, every point has a reflexive successor. By our supposition, them, alll formulas in Γ are on all valid on this frame and are therefore provable in **KD4**. Since **KD4** also has frames not meeting the condition – such as Hughes's example $\langle Nat, < \rangle$ – the soundness of **KD4** w.r.t. the class of serial transitive frames means that the formulas in Γ, being **KD4**-provable are also valid on such frames, contradicting our supposition. □

Hughes also shows that $\mathbf{K\Diamond T}$ has the finite model property ([3], Theorem 2) and is not finitely axiomatizable ([3], Theorem 10), in the sense of this phrase in which the rule of necessitation is not held (as in, for example, Lemmon [4]) to count against an axiomatization's being finite.[4] Of concern to us in what follows are two other results and one remark in Hughes. The first result ([3], Theorem 11) is that no closed first order formula defines the class of frames for $\mathbf{K\Diamond T}$. (The other result is Theorem 1.2 below, and the remark in question follows $\Diamond\mathbf{t}_1$ below.) The argument here uses the compactness theorem of first order logic, and, as Hughes writes:[5]

> What Theorem 11 shows is that there can be no single wff of first-order logic which defines the class of all frames for $\mathbf{K\Diamond T}$. It does not show that there cannot be an infinite set of such wff which does this. In particular, the possibility is left open that for each $\mathbf{K\Diamond T}_n$ the class of all its frames is definable by a class of first-order wff (or even by a single first-order wff), and that the frames for $\mathbf{K\Diamond T}$ are definable by the union of these classes.

[2] See [1], p. 35.

[3] That **KD4** proves $\mathbf{K\Diamond T}_n$ for arbitrary n follows from the completeness **KD4** w.r.t. the class of transitive serial frames and the fact that no such frame is finitely colorable in the sense introduced two paragraphs before Theorem 1.2 below, together with that theorem itself.

[4] Hughes also claims – Theorem 4 in [3] – to show that $\mathbf{K\Diamond T}$ is decidable, which may the case, though it is not easy to see how the proof he gives differs from a general argument that any recursively axiomatizable logic with the finite model property must be decidable, which is something shown to be false in [6].

[5] This passage appears on p. 179 of [3], and in quoting it Hughes's nomenclature "*KMT*" has been replaced by "$\mathbf{K\Diamond T}$", as in another quotation from Hughes, below.

For the case of $n = 1$, the question raised here about $\mathbf{K}\Diamond\mathbf{T}_1$, has, as Hughes remarks, a simple affirmative answer, in view of the equivalence in any normal modal logic of $\Diamond\mathbf{T}_1$ and the formula $\Box\Box p \to \Diamond p$: namely in having for its class of frames precisely those satisfying the first order condition

$\Diamond\mathbf{t}_1$ $\quad \forall x \exists y(Rxy \,\&\, \exists z(Rxz \,\&\, Rzy))$,

"(B)ut," Hughes writes ([3], p. 179), "I do not know whether analogous results can be obtained for any or all of the other $\mathbf{K}\Diamond\mathbf{T}_n$ systems." We shall see an argument in the following section (the proof of Theorem 2.1) which would seek to extend this result to the $n = 2$ case.

Finally, by way of summary of points from [3], we have the notion of an n-colouring of a frame $\langle W, R\rangle$, adapted from graph theory: this is an assignment of n colours to the elements of W subject to the condition that whenever Rxy for $x, y \in W$, x and y are assigned different colours. A frame for which such a colouring exists for a given n is *n-colorable*, and a frame which is n-colorable for some $n \in Nat$ is *finitely colorable*.[6]

Using the simple observation that in any model $\langle W, R, V\rangle$ no formula of the form $\Box A \to A$ can be false at both x and y when Rxy, Hughes proves a striking result – recently recalled in the more heavily graph-theoretical [2] – to state which we use the notation $\langle R(x), R_x\rangle$, for a certain kind of 'local' subframe of a frame $\langle W, R\rangle$ with $x \in W$. Here $R(x)$ is $\{y \in W \mid Rxy\}$ and R_x is the restriction of R to $R(x)$.[7] What Hughes shows ([3], Theorem 12) is that the class of frames for $\mathbf{K}\Diamond\mathbf{T}$ consists precisely of those frames $\langle W, R\rangle$ for which none of the subframes $\langle R(x), R_x\rangle$, for $x \in W$, is finitely colorable. We prefer a 'semantically local' formulation of this result (see van Benthem [1], esp. Chapter 7), using the individual axioms $\Diamond\mathbf{T}_n$, and candidate validating points one by one, which is what the proof of Hughes's Theorem 12 gives. Where $x \in W$ we say A is *valid at x* in the frame $\langle W, R\rangle$ – notated $\langle W, R\rangle \models_x A$ – when A is true at x (notation: $\langle W, R, V\rangle \models_x A$) in every model $\langle W, R, V\rangle$ on that frame.

Theorem 1.2 *For all $n \in Nat$, all frames $\langle W, R\rangle$, and all $x \in W$, we have $\langle W, R\rangle \models_x \Diamond\mathbf{T}_n$ if and only if the frame $\langle R(x), R_x\rangle$ is not n-colorable.*

2 A (Faulty) Treatment of the Case of $\Diamond\mathbf{T}_2$

It is easy to see that $\Diamond\mathbf{t}_1$ gives a necessary and sufficient condition for the validity of $\Diamond\mathbf{T}_1$ on a frame, and, with its first universal quantifier removed, of

[6] The use of the '-our' spelling for 'colour' not does require the insertion of a 'u' in 'colorable'. Compare 'coloration'; 'colorable' can be thought of as a back-formation from this – *cf.* 'soluble' (as opposed to 'solvable').

[7] Similarly, below, for the notation $R^n(x)$, in which R^n is the n-fold relative product of R with itself.

giving such a condition for the validity of this formula at a point x in a frame. There are several ways of reformulating the condition so as to suggestion a possible extension to cover the case of $\Diamond\mathbf{T}_n$ for $n > 1$. For example, sticking to the latter, local version of $\Diamond\mathbf{t}_1$, we can see it as demanding if a point x that $R(x) \cap R^2(x) \neq \varnothing$: so perhaps an analogous condition for $\Diamond\mathbf{T}_2$ might be $R^2(x) \cap R^3(x) \neq \varnothing$, or again $R(x) \cap R^2(x) \cap R^3(x) \neq \varnothing$. What we need instead, however, is the following idea. An *n-chain* in a frame $\langle W, R\rangle$ is a sequence $x_1, \ldots, x_n$ of elements of W with Rx_1x_2, Rx_2x_3,...,$Rx_{n-1}x_n$; we say that x_1 *starts* this n-chain. (Note that we do not require $x_i \neq x_{i+1}$.) Such a sequence is a *transitive* n-chain if the restriction of R to $\{x_1, \ldots, x_n\}$ is a transitive relation; equivalently: if Rx_ix_j for $i, j \in \{1, \ldots, n\}$ whenever $i < j$. Thus 2-chains are automatically transitive. The condition $\Diamond\mathbf{t}_1$ can be thought of as saying that every point starts a transitive 3-chain. Local formulation: with its initial universal quantifier removed, $\Diamond\mathbf{t}_1$ says that x starts a transitive 3-chain. Thus we can say that in any frame the points at which $\Diamond\mathbf{T}_1$ is valid are precisely those starting some transitive 3-chain. To give a simple answer to Hughes's question as "whether analogous results can be obtained for any or all of the other $\mathbf{K}\Diamond\mathbf{T}_n$ systems", we turn to the case of $\Diamond\mathbf{T}_2$. (N.B.: Although we are calling this a theorem, we will be retracting it in the following section.)

Theorem 2.1 *For all frames $\langle W, R\rangle$ and all $x \in W$, $\langle W, R\rangle \models_x \Diamond\mathbf{T}_2$ if and only if x starts some transitive 4-chain in $\langle W, R\rangle$.*

Proof. Let us recall that $\Diamond\mathbf{T}_2$ is the formula $\Diamond((\Box p \to p) \wedge (\Box q \to q))$. The 'if' direction of the theorem involves a straightforward verification of the kind involved in soundness proofs, and we omit the details here. (The relevant soundness result in the present instance is the soundness of $\mathbf{K}\Diamond\mathbf{T}_2$ w.r.t. the class of all frames in which every point starts a transitive 4-chain.) We turn to the 'only if' direction. Suppose $\Diamond\mathbf{T}_2$ is valid at x in $\langle W, R\rangle$. Thus for any choice of V, there is some $y \in R(x)$ for which $\langle W, R, V\rangle \models_y \Box p \to p$ and $\langle W, R, V\rangle \models_y \Box q \to q$. In particular this holds when V is chosen with $V(p) = R^2(x)$. (We will specify $V(q)$ presently.) For such a choice of V, since $y \in R(x)$, this forces $R(y) \subseteq V(p)$, so the antecedent of $\Box p \to p$, a conditional we have seen to be true at y in any any $\langle W, R, V\rangle$ with $V(p)$ as here specified, is itself true at y. So the consequent of this conditional is also true at y. This consequent is the formula p, so $y \in R^2(x)$. Thus for some $z_1 \in W$, Rxz_1 and Rz_1y. Fix on such a z_1 and now set $V(q) = \{w \in W \mid \exists u\,\mathbf{.}\, Rz_1u \,\&\, Rxu \,\&\, Ruw\}$. Consider any $v \in R(y)$. v satisfies the condition just imposed on the elements of $V(q)$, namely that for some u, Rz_1u, Rxu, and Ruv, because we may take the point y as our u. Thus any such successor of y lies in $V(q)$, so $\Box q$ is true at y. As remarked above, $\Box q \to q$ is true at y, and so q itself is true at y, which means, given the way $V(q)$ was defined, that for some point – call it z_2 – we have Rz_1z_2,

Rxz_2 and Rz_2y. As we already have Rxz_1, Rz_1y and Rxy, this means that x starts the transitive 4-chain x, z_1, z_2, y. □

Corollary 2.2 *The frames for* $\mathbf{K}\Diamond\mathbf{T}_2$ *are precisely those in which every element starts some transitive 4-chain.*

As the (putative) proof of Theorem 2.1 makes manifest, the condition on frames in Corollary 2.2 is (or can be rewritten as) a first-order condition, so we have answered the question raised by Hughes's discussion as to whether $\mathbf{K}\Diamond\mathbf{T}_1$ was unique in this respect: it is not. We could proceed in terms of the machinery of n-chains to extend this line of enquiry to $\mathbf{K}\Diamond\mathbf{T}_n$ for $n > 2$, but, as foreshadowed above, we have already gone wrong, and it is time to see how.

3 Counterexample and Diagnosis

Theorem 2.1, as we have called it, is false, as is Corollary 2.2. Specifically, its 'only if' direction fails,[8] as we can see from the following counterexample. Consider the 6-element frame we get by taking x with $R(x) = \{y_1, \ldots, y_5\}$, where $y_1, \ldots, y_5$ form a pentagon in which each y_i bears the relation to the points immediately adjacent to it. (The easiest way to diagram this, should the reader care to, would be to draw the pentagon and place x in its center, with directed edges going to each of the y_i.) Because of the absence of (loops and) 'internal' edges amongst the y_i, no point in $R(x)$ starts a transitive 3-chain, so x itself starts no transitive 4-chain; according to the 'only if' direction of Theorem 2.1, then, $\Diamond\mathbf{T}_2$ is not valid at x in the frame described. But this contradicts Hughes's Theorem 1.2 above, since the 'local' subframe $\langle R(x), R_x\rangle$ is not 2-colorable,[9] which suffices, according to that result, for $\Diamond\mathbf{T}_2$ to valid at x in the original 6-element frame.

To see where the proof of Theorem 2.1 goes wrong, let us recast the hypothesis that $\Diamond\mathbf{T}_2$ is valid at x in the frame $\langle W, R\rangle$ as an explicit second order condition on x: capital P, Q, are monadic second-order variables corresponding to the propositional variables p, q:

[8] There is no problem about the 'if' direction, which we can in any case derive from (Hughes's) Theorem 1.2 thus. If $x \in W$ starts a transitive 4-chain x, y, z, w, say, in $\langle W, R\rangle$ then $\{y, z, w\} \subseteq R(x)$ and since these points form a triangle (a "3-gon" in the language of note 9 below) $R(x)$ has no 2-colouring, so we get the result from the 'if' direction of Theorem 1.2. As we are about to see, however, while x's starting a transitive four change is sufficient for $\langle R(x), R_x\rangle$ to fail to be 2-colourable, it is not necessary.

[9] The point is familiar from graph theory: no n-gon, for n odd, is 2-colorable. The fact that we are working with directed graphs makes no difference here, and while we have made our pentagon's edges doubly directed, any other way of orienting these edges (of turning the graph – often called P_5 in the literature – into a digraph, that is) would have served equally well for the counterexample.

$$\forall P \forall Q \exists y \in R(x)[(R(y) \subseteq P \Rightarrow y \in P) \,\&\, (R(y) \subseteq Q \Rightarrow y \in Q)].$$

Corresponding to the part of the above argument (i.e., in the proof of Theorem 2.1) which ran "In particular this holds when V is chosen with $V(p) = R^2(x)$" we have a step of $\forall$ elimination (or 'instantiation'), removing the initial $\forall P$:

$$\forall Q \exists y \in R(x)[(R(y) \subseteq R^2(x) \Rightarrow y \in R^2(x)) \,\&\, (R(y) \subseteq Q \Rightarrow y \in Q)].$$

This we can simplify to:

$$\forall Q \exists y \in R^2(x)[R(y) \subseteq Q \Rightarrow y \in Q].$$

Spelling out the shorthand, and using "z_1" rather than just "z", in order to mimic the above proof:

$$\forall Q \exists y \exists z_1 [Rxy \,\&\, Rxz_1 \,\&\, Ryz_1 \,\&\, [R(y) \subseteq Q \Rightarrow y \in Q)].$$

The earlier proof (of Theorem 2.1) continued thus: "Fix on such a z_1 and now set $V(q) = \{w \in W \mid \exists u \,\textbf{.}\, Rz_1u \,\&\, Rxu \,\&\, Ruw\}$." With the explicit "$\forall Q$" at the front, we can see that this continuation is illegimate. For each choice of Q there is some way of choosing (y and) z_1 so as to satisfy a certain condition, and now we are trying to specify a particular Q by using the z_1 that itself depends on how Q is chosen.

The illegimate move is somewhat masked rhetorically by the way the proof of Theorem 2.1 is set out. Even at the stage of saying what $V(p)$ is to be, there is a slightly obscuring move, in that the existence of a "$y \in R(x)$ for which $\langle W, R, V \rangle \models_y \Box p \to p$ and $\langle W, R, V \rangle \models_y \Box q \to q$" is mentioned *before* $V(p)$ is specified, with the words "In particular this holds when V is chosen with $V(p) = R^2(x)$". But even though y as above has already been introduced, we see that $V(p)$ is not defined in such a way as to make it depend on y. Thus the stipulation regarding $V(p)$ is actually legitimate, even though a clearer way of proceeding would have been to do this first, before mentioning the existence of a point y satisfying the above conditions. There is then the parenthetical comment "We will specify $V(q)$ presently" as though there a similar legitimacy will similarly accrue to any such stipulation concerning $V(q)$, since after all the validity of our formula at x requires its truth at x *regardless* of how $V(p)$ and $V(q)$ are chosen: validity at a point is after all truth at that point in *all* models. As the preceding paragraph makes clear, however, the universal quantification here means only that we consider all ways of specifying $V(p)$ and $V(q)$ which themselves depend only on x.

Autobiographical Remark and Acknowledgment. The author did not (initially) set out deliberately to use the rhetorical devices just commented on in order to produce a sophistical argument, but was thoroughly taken in by them and thought of the proof of Theorem 2.1 as successful, only to be

shown otherwise – in pursuit of a more general lemma to the effect that if there are no transitive n-chains in a digraph, then it has an n-colouring – by his son, Bryn Humberstone, whose 7-vertex counterexample to this would-be lemma revealed the confusion (and the clash with Hughes's result, given as Theorem 1.2 here).

References

[1] J. van Benthem, *Modal Logic and Classical Logic*, Bibliopolis, Naples 1985.

[2] R. Goldblatt, I. Hodkinson, and Y. Venema, 'Erdős Graphs Resolve Fine's Canonicity Problem', *Bulletin of Symbolic Logic* **10** (2004), 186–208.

[3] G. E. Hughes, 'Every World Can See a Reflexive World', *Studia Logica* **49** (1990), 174–181.

[4] E. J. Lemmon, 'Some Results on Finite Axiomatizability in Modal Logic', *Notre Dame Journal of Formal Logic* **6** (1965), 301–308.

[5] E. J. Lemmon and D. S. Scott, *An Introduction to Modal Logic*, ed. K. Segerberg, American Philosophical Quarterly Monograph Series, Basil Blackwell, Oxford 1977. (Originally circulated 1966.)

[6] A. Urquhart, 'Decidability and the Finite Model Property', *Journal of Philosophical Logic* **10** (1981), 367–370.