A CORRECT POLYNOMIAL TRANSLATION OF S4 INTO INTUITIONISTIC LOGIC

RAJEEV GORÉ AND JIMMY THOMSON
SUBMITTED 2013,
TO APPEAR IN JOURNAL OF SYMBOLIC LOGIC

Abstract. We show that the polynomial translation of the classical propositional normal modal logic S4 into the intuitionistic propositional logic Int from Fernández is incorrect. We give a modified translation and prove its correctness, and provide implementations of both translations to allow others to test our results.

§1. Introduction. It is well known that the validity and satisfiability problems for the classical propositional normal modal logic `S4` and the intuitionistic propositional logic `Int` are PSPACE-complete and thus there must exist a polynomial translation from each into the other. The Gödel translation [2] provides a translation from `Int` into `S4`, but the only published polynomial translation from `S4` into `Int` we could find is by Fernández [1]. Here, we first show that the translation is incorrect. By pinpointing the flaws, we give a correct polynomial translation from `S4` into `Int`.

The paper is structured as follows. In Section 2 we define the syntax and Kripke semantics of the propositional intuitionistic logic `Int` and of the propositional normal modal logic `S4`. In Section 3 we show that the original translation is incorrect. In Section 4, we give our solution and prove it correct.

§2. Semantic Preliminaries. We define `Int`-formulae from an infinite set *Prop* of propositional variables using the following BNF grammar where $p \in Prop$ and $\bot$ is the *falsum* constant:

$$\varphi = \bot \mid p \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \varphi \to \varphi$$

We also define $\neg\varphi = (\varphi \to \bot)$. We use rooted Kripke models of `Int` which are structures $\mathcal{M} = (W, R, L, r)$ where: W is a non-empty set of possible worlds; R is a reflexive, transitive and antisymmetric binary relation on W; the valuation $L : Prop \to 2^W$ obeys **persistence**: if $w \in L(p)$ and $R(w, v)$ then $v \in L(p)$; and $r \in W$ is a root world such that $\forall w \in W.R(r, w)$ holds. Since `Int` enjoys the finite model property, we can restrict ourselves to models where W is finite.

The semantics of `Int` are given in Figure 1. An `Int`-formula φ is `Int`-satisfiable if there exists some `Int`-model $\mathcal{M}$ and some world w in that `Int`-model such that $\mathcal{M}, w \Vdash \varphi$. An `Int`-formula is `Int`-valid if $\neg\varphi$ is not `Int`-satisfiable. That is, an `Int`-formula is `Int`-valid if every world w in every `Int`-model $\mathcal{M}$ obeys $\mathcal{M}, w \Vdash \varphi$.

$\mathcal{M}, w \not\Vdash \bot$	
$\mathcal{M}, w \Vdash p$	iff $w \in L(p)$
$\mathcal{M}, w \Vdash \varphi \wedge \psi$	iff $\mathcal{M}, w \Vdash \varphi$ and $\mathcal{M}, w \Vdash \psi$
$\mathcal{M}, w \Vdash \varphi \vee \psi$	iff $\mathcal{M}, w \Vdash \varphi$ or $\mathcal{M}, w \Vdash \psi$
$\mathcal{M}, w \Vdash \varphi \to \psi$	iff $\forall v$. if $R(w, v)$ then $\mathcal{M}, v \not\Vdash \varphi$ or $\mathcal{M}, v \Vdash \psi$

FIGURE 1. Kripke semantics for `Int`

We define `S4`-formulae over an infinite set *Prop* of propositional variables using the following BNF grammar where $p \in Prop$ and $\bot$ is the *falsum* constant:

$$\varphi = \bot \mid p \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \varphi \to \varphi \mid \Box\varphi$$

Again we define $\neg\varphi = (\varphi \to \bot)$. We can also define $\Diamond\varphi = (\neg\Box\neg\varphi)$. For `S4`, Kripke models are structures $\mathcal{M} = (W, R, L, r)$ where: W is a non-empty set of possible worlds; R is a reflexive and transitive binary relation on W; $L : Prop \to 2^W$ is a valuation; and $r \in W$ is a root world obeying $\forall w \in W.R(r, w)$.

The semantics of `S4` are given in Figure 2. An `S4`-formula φ is `S4`-satisfiable if there exists some `S4`-model $\mathcal{M}$ and some world w in that `S4`-model such that $\mathcal{M}, w \Vdash \varphi$. An `S4`-formula is `S4`-valid if $\neg\varphi$ is not `S4`-satisfiable. That is, an `S4`-formula is `S4`-valid if every world w in every `S4`-model $\mathcal{M}$ obeys $\mathcal{M}, w \Vdash \varphi$.

We can restrict this class further because `S4` is complete with respect to the class of binary, reflexive and transitive Kripke frames which are rooted finite trees of finite clusters of worlds where, within a cluster, all worlds are related to each other.

§3. Translating `S4`-formulae into `Int`-formulae. If N is the number of $\Box$-symbols that appear in an `S4` formula φ, then we can restrict ourselves to those frames with at most $N + 1$ distinct clusters along any branch, and each cluster has at most $N + 1$ worlds. In such a frame, we say that the *level* of a world is the number of clusters between the root and the cluster containing that world. If the world is in the root cluster, then it has level 0.

To represent such `S4`-frames, the translation of Fernández [1] creates multiple `Int`-propositions p^i_j for each `S4`-proposition p in the `S4`-formula, intended to represent the valuation of the `S4`-proposition p in a world with level i, using j to

$\mathcal{M}, w \not\Vdash \bot$	
$\mathcal{M}, w \Vdash p$	iff $w \in L(p)$
$\mathcal{M}, w \Vdash \varphi \wedge \psi$	iff $\mathcal{M}, w \Vdash \varphi$ and $\mathcal{M}, w \Vdash \psi$
$\mathcal{M}, w \Vdash \varphi \vee \psi$	iff $\mathcal{M}, w \Vdash \varphi$ or $\mathcal{M}, w \Vdash \psi$
$\mathcal{M}, w \Vdash \varphi \to \psi$	iff $\mathcal{M}, w \not\Vdash \varphi$ or $\mathcal{M}, w \Vdash \psi$
$\mathcal{M}, w \Vdash \Box\varphi$	iff $\forall v$. if $R(w, v)$ then $\mathcal{M}, v \Vdash \varphi$

FIGURE 2. Kripke semantics for `S4`

$$
\begin{aligned}
(p)^n_m &= p^n_m \\
(\bot)^n_m &= \bot \\
(\psi_1 \cdot \psi_2)^n_m &= (\psi_1)^n_m \cdot (\psi_2)^n_m \\
(\Box\psi)^n_m &= b^n_\psi
\end{aligned}
$$

FIGURE 3. Translation from Fernández [1] of an S4-formula ψ at the m^{th} world in a cluster at level n in an S4-model to an Int-formula $(\psi)^n_m$. The propositional variables b^n_ψ are disjoint from p^n_m and "$\cdot$" on the left/right hand side of the equal sign represents the same binary connective of S4/Int respectively.

distinguish between worlds within a cluster in the given S4-model. Figure 3 gives the Int -formula $(\psi)^n_m$ which represents the valuation of the S4-formula ψ at the m^{th} world in a cluster n clusters from the root in the S4-model. The branching of the Int-model allows for multiple clusters with level n in the S4-model.

The translation also makes use of new Int-propositions l^i intended to indicate the level of an S4-cluster, and new Int-propositions b^i_ψ to indicate when the S4 -formula $\Box\psi$ holds at a cluster of level i.

To determine S4-validity, Fernández [1] defines the translation φ^{Int}_o, written with an extra subscript o for "original", as shown in Figure 4. The claim is that φ^{Int}_o is Int-valid iff φ is S4-valid.

However, there are two errors in this encoding. First, consider a formula φ of S4 with no $\Box$-formulae. In this case, $N = 0$, and so $\text{Lev}(\varphi) = (l^0 \wedge \neg l^0)$, leading

$$\text{Lev}(\varphi) = l^0 \wedge \neg l^N \wedge \bigwedge_{k=0}^{N-1} (l^{k+1} \to l^k)$$

$$\text{Mid}^n(\varphi) = l^n \to \left(\bigwedge_{\Box\psi \in sub(\varphi)} \left(b^n_\psi \vee \neg b^n_\psi\right) \wedge \bigwedge_{p \in sub(\varphi)} \bigwedge_{0 \le m \le N-1} (p)^n_m \vee \neg (p)^n_m \right)$$

$$A^n_{o,\psi(\varphi)} = \bigwedge_{n \le k < N} \left(l^k \to l^{k+1} \vee \bigwedge_{m=0}^{N-1} (\psi)^k_m \right)$$

$$\text{Box}^n_{o,\psi(\varphi)} = l^n \to \left((b^n_\psi \to l^{n+1} \vee A^n_{o,\psi(\varphi)}) \wedge (A^n_{o,\psi(\varphi)} \to l^{n+1} \vee b^n_\psi) \right)$$

$$P(\varphi) = \text{Lev}(\varphi) \wedge \bigwedge_{0 \le n \le N-1} \text{Mid}^n(\varphi) \wedge \bigwedge_{0 \le n \le N-1} \bigwedge_{\Box\psi \in sub(\varphi)} \text{Box}^n_{o,\psi(\varphi)}$$

$$\varphi^{\text{Int}}_o = P(\varphi) \to (\varphi)^0_0$$

FIGURE 4. S4 to Int translation φ^{Int}_o of Fernández [1]

$$\varphi = (\Box p \to p) \qquad \text{modified } N = 2 \qquad sub(\varphi) = \{p, \Box p\}$$

$$\text{Lev}(\varphi) = l^0 \wedge \neg l^N \wedge \bigwedge_{k=0}^{N-1}(l^{k+1} \to l^k) = \underline{l^0} \wedge \underline{\neg l^2} \wedge (l^1 \to \underline{l^0}) \wedge (l^2 \to \underline{l^1})$$

$$\text{Mid}^n(\varphi) = l^n \to \left(\bigwedge_{\Box\psi \in sub(\varphi)} \left(\underline{b_\psi^n \vee \neg b_\psi^n}\right) \wedge \bigwedge_{p \in sub(\varphi)} \bigwedge_{0 \le m \le N-1} \underline{(p)_m^n \vee \neg (p)_m^n}\right)$$

$$\text{Box}^0_{o,p(\varphi)} = l^0 \to \left((b_\psi^0 \to \underline{l^1} \vee A^0_{o,p(\varphi)}) \wedge (A^0_{o,p(\varphi)} \to \underline{l^1} \vee b_\psi^0)\right)$$

$$A^1_{o,p(\varphi)} = \bigwedge_{1 \le k < 2}\left(l^k \to l^{k+1} \vee \bigwedge_{m=0}^{2-1}(p)_m^k\right) = l^1 \to l^{1+1} \vee \underline{(p_0^1 \wedge p_1^1)}$$

$$\text{Box}^1_{o,p(\varphi)} = l^1 \to \left((b_p^1 \to l^{1+1} \vee \underline{A^1_{o,p(\varphi)}}) \wedge (A^1_{o,p(\varphi)} \to l^{1+1} \vee \underline{b_p^1})\right)$$

$$P(\varphi) = \text{Lev}(\varphi) \wedge \bigwedge_{0 \le n \le 1} \text{Mid}^n(\varphi) \wedge \text{Box}^0_{o,p(\varphi)} \wedge \text{Box}^1_{o,p(\varphi)}$$

$$\varphi_o^{\text{Int}} = P(\varphi) \to (\varphi)_0^0$$

$$w \qquad \Vdash \qquad l^0, l^1, p_0^1, p_1^0, p_1^1, b_p^0, b_p^1$$

FIGURE 5. Computation of φ_o^{Int} using the Fernández translation from Example 1. Underlines indicate the formulae that are "true" at w in the given model and which directly influence the truth value of the larger formula.

to $\varphi_o^{\text{Int}} = \bot \to \varphi_0^0$, which is `Int`-valid regardless of φ. The obvious solution here is to modify N to be one more than the number of $\Box$ symbols in φ.

The second error is more subtle, and we demonstrate it via an example.

Example 1. Consider the `S4`-valid formula $\varphi = (\Box p \to p)$ which has only one $\Box$-symbol and thus a modified N of 2. The translation given in Figure 4 requires the following `Int`-formulae as new propositions: l^0, l^1, l^2 for level formulae, b_p^0, b_p^1 to represent the formula $\Box p$, and $p_0^0, p_1^0, p_0^1, p_1^1$ to represent the value of p in up to two worlds and up to two levels.

Now consider the `Int`-model $\mathcal{M} = (\{w\}, \{(w,w)\}, L, w)$ with a single reflexive world w, and $w \in L(\psi)$ for $\psi \in \{l^0, l^1, p_0^1, p_1^0, p_1^1, b_p^0, b_p^1\}$, and $w \notin L(\psi)$ for $\psi \in \{l^2, p_0^0\}$.

Referring to Figure 5, where underlines indicate the parts that are "true" at w and which directly affect the truth value of the larger formulae, we obviously have $\mathcal{M}, w \Vdash \text{Lev}(\varphi)$. We have $\mathcal{M}, w \Vdash \text{Mid}^n(\varphi)$ because in a single-world model, $\psi \vee \neg\psi$ is intuitionistically true for all ψ. Since $\mathcal{M}, w \Vdash l^1$, we have $\mathcal{M}, w \Vdash \text{Box}^0_{o,p(\varphi)}$, because the inner implications are made true by the "escape hatch" provided by $l^{n+1} = l^1$. We also have $\mathcal{M}, w \Vdash p_0^1 \wedge p_1^1$, thus $\mathcal{M}, w \Vdash A^1_{o,p(\varphi)}$, and since $\mathcal{M}, w \Vdash b_p^1$, we have $\mathcal{M}, w \Vdash \text{Box}^1_{o,p(\varphi)}$. Thus we have $\mathcal{M}, w \Vdash P(\varphi)$. However, $\mathcal{M}, w \nVdash \varphi_0^0 = b_p^0 \to p_0^0$, and so $\mathcal{M}$ is an `Int`-countermodel to φ_o^{Int}, despite φ being `S4`-valid.

The culprits are the "escape hatches" l^{n+1} in $\mathrm{Box}^n_{o,\psi(\varphi)}$ and l^{k+1} in $A^n_{o,\psi}(\varphi)$, which allow us to ignore constraints imposed by l^0 by jumping straight to l^1.

What happens if we keep the original definition of N and just drop the l^N part from $Lev(\varphi)$? Example 1 is no longer a counter-example, but using $\varphi = (\Box p \to p) \wedge (\Box q \to q)$ with exactly the same structure does give a counterexample.

§4. Solution. As mentioned before, the first step of the solution is to modify N to be one more than the number of $\Box$-formulae in the given `S4`-formula φ. This is probably what was intended, as no proofs need to change and there are no 0-standard models, as defined by Fernández [1]. To avoid confusion, we will retain N as the number of $\Box$-symbols, as used by Fernández [1], and use $M = N + 1$ for the modified value.

The second change is to modify the definition of $A^n_{o,\psi}$ and $\mathrm{Box}^n_{o,\psi}$ as follows:

$$A^n_\psi(\varphi) = \bigwedge_{n \le k < M} \left(l^k \to \bigwedge_{m=0}^{M-1} (\psi)^k_m \right)$$
$$\mathrm{Box}^n_\psi(\varphi) = l^n \to \big((b^n_\psi \to A^n_\psi(\varphi)) \wedge (A^n_\psi(\varphi) \to b^n_\psi) \big)$$

This removes the "escape hatches" in the formula in the case where a higher l proposition was true. All conditions imposed on formulae by some l^i must be met, regardless of whether other l^j formulae are true. For example we no longer have $\mathcal{M}, w \Vdash P(\varphi)$ in Example 1 because $\mathcal{M}, w \nVdash \mathrm{Box}^0_p(\varphi)$: that is, we have $\mathcal{M}, w \Vdash b^0_p$ and $\mathcal{M}, w \Vdash l^0$, but $\mathcal{M}, w \nVdash A^0_p(\varphi)$ because $\mathcal{M}, w \nVdash p^0_0$.

We write φ^{Int}_c for our "correct" translation (we cannot use n for "new" as it clashes with the integers used as subscripts). Note that our translation φ^{Int}_c is actually smaller than the translation φ^{Int}_o of Fernández [1] since all we have done is remove some disjunctions, and so the translation remains polynomial.

4.1. Converting `Int`-models to `S4`-models. We work with rooted and finite `Int`-models $\mathcal{M} = (W, R, L, r)$. We intend to show that the modified φ^{Int}_c has an `Int`-countermodel iff φ has an `S4`-countermodel.

First, we prove some lemmas about small modifications to `Int`-models.

DEFINITION 2. Given an `Int`-model $\mathcal{M} = (W, R, L, r)$, a world $u \in W$, and a finite set $\mathcal{L}$ of propositional variables such that $\forall w \in W, \forall p \in \mathcal{L}$, if $R(w, u)$ and $w \neq u$ then $w \notin L(p)$. Define $insert(\mathcal{L}, u, \mathcal{M}) = (W', R', L', r')$ as follows:

1. let v be a new world not in W
2. if $r = u$ then $r' = v$ otherwise $r' = r$
3. $W' = W \cup \{v\}$
4. $R' = R \cup \{(v, v)\} \cup \{(v, x) \mid (u, x) \in R\} \cup \{(y, v) \mid (y, u) \in R \,\&\, y \neq u\}$
5. for all p we have $L'(p) \cap W = L(p)$
6. for all $p \notin \mathcal{L}$ we have $v \in L'(p)$ iff $u \in L'(p)$
7. for all $p \in \mathcal{L}$ we have $v \notin L'(p)$.

That is, we insert a new world v as an immediate predecessor of u, where all proper predecessors y of u are made proper predecessors of v and all successors x of u including u itself are made successors of v.

LEMMA 3. *If $\mathcal{M} = (W, R, L, r)$ is an* `Int`*-model with a world u, and $\mathcal{L}$ is a set of propositional variables falsified at all y such that $R(y, u)$ and $y \neq u$, then $\mathcal{M}' = insert(\mathcal{L}, u, \mathcal{M})$ is an* `Int`*-model and for all* `Int`*-formulae ψ which do not include propositions from $\mathcal{L}$ and for all $w \in W$, we have $\mathcal{M}', w \Vdash \psi$ iff $\mathcal{M}, w \Vdash \psi$, and additionally we have $\mathcal{M}', v \Vdash \psi$ iff $\mathcal{M}, u \Vdash \psi$.*

PROOF. We first prove that $\mathcal{M}'$ is still an `Int` -model. That is, we have to prove that $\mathcal{M}'$ is transitive, reflexive, antisymmetry and persistence. Of these, we deal only with the non-trivial cases.

Transitivity still holds: the only case that could possibly fail is $R'(a, b)$ and $R'(b, v)$ but not $R'(a, v)$ for some $a \neq v$ and $b \neq v$. Since both a and b are in the original model, the edge $R'(a, b)$ is from the original model, hence $R(a, b)$. Since $R'(b, v)$, we must have $R(b, u)$ and $b \neq u$ by definition of R'. By the transitivity of R we must have $R(a, u)$, and thus by definition of R' we must have $R'(a, v)$ as required.

The valuation L' obeys the persistence property: because the original model had a persistent valuation, the only way for $\mathcal{M}'$ to not have a persistent valuation is if the introduction of v changed something. Suppose for a contradiction that for some proposition p and some world w we have $R'(v, w)$, $v \in L'(p)$ and $w \notin L'(p)$. Then $p \notin \mathcal{L}$, and $u \in L(p)$ by the definition of L'. Similarly, $w \notin L(p)$. Since $R(u, w)$, the original $\mathcal{M}$ does not satisfy persistence, giving a contradiction. Suppose then that $R'(w, v)$ and $w \in L'(p)$ and $v \notin L'(p)$. Then we must have $R(w, u)$ and $w \neq u$ by the definition of R'. If $p \in \mathcal{L}$ then $w \in W$ and $w \in L'(p)$ implies $w \in L(p)$, contradicting the definition of $\mathcal{L}$, hence $p \notin \mathcal{L}$. But then $v \notin L'(p)$ implies that $u \notin L(p)$, and the persistence of $\mathcal{M}$ implies that $w \notin L(p)$, and hence $w \notin L'(p)$: contradiction. Thus $\mathcal{M}'$ is an `Int`-model.

Now we prove by structural induction on ψ that we must have $\mathcal{M}, u \Vdash \psi$ iff $\mathcal{M}', v \Vdash \psi$, and $\mathcal{M}, w \Vdash \psi$ iff $\mathcal{M}', w \Vdash \psi$. First the base cases:

$\psi = p$: Since p appears in ψ, we must have $p \notin \mathcal{L}$ and so by Definition 2.6 $u \in L(p)$ iff $v \in L'(p)$. Thus $\mathcal{M}, u \Vdash p$ iff $\mathcal{M}', v \Vdash p$. Additionally, by Definition 2.5, we have $w \in L'(p)$ iff $w \in L(p)$, thus $\mathcal{M}, w \Vdash p$ iff $\mathcal{M}', w \Vdash p$.

$\psi = \bot$: Trivially, $\mathcal{M}, u \nVdash \bot$, $\mathcal{M}', v \nVdash \bot$, $\mathcal{M}, w \nVdash \bot$ and $\mathcal{M}', w \nVdash \bot$.

Now the step cases, using the following inductive hypotheses:

IH1: for all subformulae ϕ of ψ we have $\mathcal{M}, u \Vdash \phi$ iff $\mathcal{M}', v \Vdash \phi$,

IH2: for all subformulae ϕ of ψ and for all worlds $w \in W$ we have $\mathcal{M}, w \Vdash \phi$ iff $\mathcal{M}', w \Vdash \phi$.

$\psi = \psi_1 \wedge \psi_2$**:** Suppose that $\mathcal{M}, u \Vdash \psi_1 \wedge \psi_2$. Then $\mathcal{M}, u \Vdash \psi_1$ and $\mathcal{M}, u \Vdash \psi_2$, so by IH1 we have $\mathcal{M}', v \Vdash \psi_1$ and $\mathcal{M}', v \Vdash \psi_2$, and thus $\mathcal{M}', v \Vdash \psi_1 \wedge \psi_2$. Similarly, if $\mathcal{M}, u \nVdash \psi_1 \wedge \psi_2$ then $\mathcal{M}, u \nVdash \psi_i$ for some $i \in \{1, 2\}$, and so $\mathcal{M}', v \nVdash \psi_i$ and therefore $\mathcal{M}', v \nVdash \psi_1 \wedge \psi_2$.

Similarly $\mathcal{M}, w \Vdash \psi_1 \wedge \psi_2$ iff $\mathcal{M}, w \Vdash \psi_1$ and $\mathcal{M}, w \Vdash \psi_2$, which by IH2 holds iff $\mathcal{M}', w \Vdash \psi_1$ and $\mathcal{M}', w \Vdash \psi_2$ and thus $\mathcal{M}', w \Vdash \psi_1 \wedge \psi_2$ as required.

$\psi = \psi_1 \vee \psi_2$**:** Similar to the above.

$\psi = \psi_1 \to \psi_2$**:** Suppose that $\mathcal{M}, u \Vdash \psi_1 \to \psi_2$. Then for all $w \in W$, if $R(u, w)$ and $\mathcal{M}, w \Vdash \psi_1$ then $\mathcal{M}, w \Vdash \psi_2$. For these w (which does not include v) by IH2, if $\mathcal{M}', w \Vdash \psi_1$ then we also have $\mathcal{M}', w \Vdash \psi_2$. Finally it follows from IH1 that if $\mathcal{M}', v \Vdash \psi_1$ then $\mathcal{M}', v \Vdash \psi_2$ because the same held for u.

Suppose instead that $\mathcal{M}, u \not\Vdash \psi_1 \to \psi_2$. Then there must exist a witness $w \in W$ such that $R(u, w)$ and $\mathcal{M}, w \Vdash \psi_1$ and $\mathcal{M}, w \not\Vdash \psi_2$. But this same witness will also exist in $\mathcal{M}'$ by IH2, thus $\mathcal{M}', w \Vdash \psi_1$ and $\mathcal{M}', w \not\Vdash \psi_2$. Since w is reachable from u, and v is a predecessor of u, we must also have w reachable from v, and thus $\mathcal{M}', v \not\Vdash \psi_1 \to \psi_2$.

For any $w \in W$, suppose that $\mathcal{M}, w \Vdash \psi_1 \to \psi_2$. Then for all successors x of w, if $\mathcal{M}, x \Vdash \psi_1$ then $\mathcal{M}, x \Vdash \psi_2$. Thus by IH2, for all $x \neq v$ with $R(w, x)$, if $\mathcal{M}', x \Vdash \psi_1$ then $\mathcal{M}', x \Vdash \psi_2$. If v is a successor of w in $\mathcal{M}'$ then u must also be a successor of w in $\mathcal{M}'$, and so by IH1, if $\mathcal{M}', v \Vdash \psi_1$ then $\mathcal{M}', v \Vdash \psi_2$. Thus $\mathcal{M}', w \Vdash \psi_1 \to \psi_2$ as required.

If instead $\mathcal{M}, w \not\Vdash \psi_1 \to \psi_2$ then there is some successor x of w such that $\mathcal{M}, x \Vdash \psi_1$ and $\mathcal{M}, x \not\Vdash \psi_2$. By IH2, we have $\mathcal{M}', x \Vdash \psi_1$ and $\mathcal{M}', x \not\Vdash \psi_2$, and thus $\mathcal{M}', w \not\Vdash \psi_1 \to \psi_2$.

$\dashv$

Effectively Lemma 3 states that we can insert "copies" of worlds with minor changes to some atomic propositions $\mathcal{L}$ without changing the truth values of formulae which do not refer to those atomic propositions.

Next we prove that if our amended φ_c^{Int} has an `Int`-countermodel, then φ has an `S4`-countermodel.

Definition 4. If $\mathcal{M} = (W, R, L, r)$ is an `Int`-model such that $\mathcal{M}, r \Vdash P(\varphi)$, then for $w \in W$, let $Lv(w)$ be defined as the index i such that $w \in L(l^i)$ and $w \notin L(l^{i+1})$.

As long as $\mathcal{M}, w \Vdash \mathrm{Lev}(\varphi)$ then $Lv(w)$ has a unique definition because then we must have $\mathcal{M}, w \Vdash l^0$ and thus $Lv(w) \geq 0$, and we must have $\mathcal{M}, w \not\Vdash l^M$ and thus $Lv(w) < M$, and we must have that if $\mathcal{M}, w \Vdash l^k$ then $\mathcal{M}, w \Vdash l^j$ for all $j < k$.

Definition 5. A model $\mathcal{M} = (W, R, L, r)$ which falsifies φ_c^{Int} is *stratified* if:

1. $Lv(r) = 0$;
2. for any two worlds $w, v \in W$, if $R(w, v)$ and $Lv(v) > Lv(w)+1$ then there is another (necessarily different) world u such that $R(w, u)$ and $R(u, v)$ with $Lv(u) = Lv(w) + 1$; and
3. if for some $w, u \in W$ we have $R(w, u)$ and $Lv(w) = Lv(u)$ then $w = u$.

We now prove that there must be a stratified `Int`-countermodel to φ_c^{Int} if there is any `Int`-countermodel of φ_c^{Int}.

Lemma 6. *If a countermodel to φ_c^{Int} exists, then one satisfying condition 1 of Definition 5 exists.*

Proof. Let $\mathcal{M} = (W, R, L, r)$ be an `Int`-countermodel of φ_c^{Int}. Without loss of generality, assume $\mathcal{M}, r \Vdash P(\varphi)$ and $\mathcal{M}, r \not\Vdash (\varphi)_0^0$. If $Lv(r) = 0$ then the lemma holds immediately. Otherwise $Lv(r) \geq 1$ and so we have $\mathcal{M}, r \Vdash l^1$ and $\mathcal{M}, r \Vdash l^0$. Create a new `Int`-model $\mathcal{M}' = \mathit{insert}(\mathcal{L}, r, \mathcal{M}) = (W', R', L', r')$ according to Lemma 3 using $\mathcal{L} = \{l^i \mid 0 < i \leq M\}$.

The new model $\mathcal{M}'$ still falsifies φ_0^0 at the new root r' according to Lemma 3 because φ_0^0 does not refer to any proposition in $\mathcal{L}$. Note that $Lv(r') = 0$ by the definition of L' as required. It remains to show that $\mathcal{M}', r' \Vdash P(\varphi)$.

We obviously have $\mathcal{M}', r' \Vdash \mathrm{Lev}(\varphi)$. The successors of r are also successors of r', so the only way for r' to fail $\mathrm{Mid}^i(\varphi)$ would be to fail locally. Since $\mathcal{M}', r' \Vdash l^i$ only for $i = 0$, we have $\mathcal{M}', r' \Vdash \mathrm{Mid}^i(\varphi)$ for $i > 0$. For $i = 0$, $\mathrm{Mid}^0(\varphi)$ does not refer to any propositions in $\mathcal{L}$ and thus by Lemma 3 we must also have $\mathcal{M}', r' \Vdash \mathrm{Mid}^0(\varphi)$.

Finally, we show that r' satisfies $\mathrm{Box}^n_\psi(\varphi)$. For $n > 0$ it satisfies $\mathrm{Box}^n_\psi(\varphi)$ vacuously because $\mathcal{M}', r' \nVdash l^n$, and all strict successors of r' satisfy $\mathrm{Box}^n_\psi(\varphi)$ because they did in $\mathcal{M}$. For $n = 0$, we have $\mathcal{M}, r \Vdash b^0_\psi \leftrightarrow A^0_\psi(\varphi)$, and we want to show that $\mathcal{M}', r' \Vdash b^0_\psi \leftrightarrow A^0_\psi(\varphi)$. Because $b^0_\psi \notin \mathcal{L}$, we have $\mathcal{M}', r' \Vdash b^0_\psi$ iff $\mathcal{M}, r \Vdash b^0_\psi$, so it remains to show that $\mathcal{M}, r \Vdash A^0_\psi(\varphi)$ iff $\mathcal{M}', r' \Vdash A^0_\psi(\varphi)$.

Suppose that $\mathcal{M}, r \nVdash A^0_\psi(\varphi)$. Then there must be some successor which satisfies l^k and falsifies ψ^k_m for some k and m, and such a successor is also a successor of r' thus $\mathcal{M}', r' \nVdash A^0_\psi(\varphi)$.

Suppose instead that $\mathcal{M}, r \Vdash A^0_\psi(\varphi)$ and thus since $\mathcal{M}, r \Vdash l^0$ we must have $\mathcal{M}, r \Vdash (\psi)^0_m$ for all $0 \le m < M$. The only way that r' could fail to satisfy $A^0_\psi(\varphi)$ is to do so locally, and with $k = 0$. However, since $(\cdot)^n_m$ does not refer to any l^i, we must also have $\mathcal{M}', r' \Vdash (\psi)^k_m$ iff $\mathcal{M}, r \Vdash (\psi)^k_m$ using Lemma 3, so $\mathcal{M}', r' \Vdash (\psi)^0_m$ and thus $\mathcal{M}', r' \Vdash A^0_\psi(\varphi)$.

Thus $\mathcal{M}', r' \Vdash P(\varphi)$, and $\mathcal{M}', r' \nVdash \varphi^0_0$, hence $\mathcal{M}'$ is a countermodel to φ^{Int}_c with $Lv(r') = 0$ as required. ⊣

Note that Lemma 6 does not hold for the original specification of φ^{Int}_o from Fernández [1]: the counterexample we gave cannot be converted to one with $Lv(r) = 0$ while still satisfying the original $P(\varphi)$. In particular $\mathrm{Box}^0_{o,p}(\varphi)$ will fail to hold if l^1 is false at the root as required by $Lv(r) = 0$.

LEMMA 7. *If an* `Int`*-countermodel of* φ^{Int}_c *exists, then one satisfying conditions 1 and 2 of Definition 5 exists.*

PROOF. Let $\mathcal{M} = (W, R, L, r)$ be an `Int`-countermodel of φ^{Int}_c after applying Lemma 6, with $w, v \in W$ such that $Lv(w) = j$, $Lv(v) > j + 1$, $R(w, v)$. Thus we have $\mathcal{M}, w \Vdash l^j$ and $\mathcal{M}, w \nVdash l^{j+1}$, and $\mathcal{M}, v \Vdash l^{j+2}$. Suppose that there is no u such that $R(w, u)$, $R(u, v)$ and $Lv(u) = j + 1$, and thus condition 2 does not hold. Let $\mathcal{L} = \{l^i \mid j + 1 < i \le Lv(v)\}$, and consider $\mathcal{M}' = insert(\mathcal{L}, v, \mathcal{M})$ where the newly introduced world is u.

That is, u is a copy of v, added between w and v with the valuation only differing on the level variables in $\mathcal{L}$. Note that $Lv(u) = j + 1$ because $l^{j+1} \notin \mathcal{L}$ and so $\mathcal{M}', u \Vdash l^{j+1}$, but $l^{j+2} \in \mathcal{L}$ so $\mathcal{M}', u \nVdash l^{j+2}$.

A similar argument to Lemma 6 applies, again using Lemma 3. The structure $\mathcal{M}'$ is an `Int`-model, the truth of formulae which do not refer to $l^k \in \mathcal{L}$ does not change between $\mathcal{M}$ and $\mathcal{M}'$, and the truth of the formulae which do refer to $l^k \in \mathcal{L}$ is preserved because the l^k are falsified on the left of an implication.

Let the "gap" between a world x and one of its immediate successors y be defined as $Lv(y) - Lv(x) - 1$ if $Lv(y) > Lv(x)$, and 0 if $Lv(y) = Lv(x)$. The sum of these gaps is unchanged between $\mathcal{M}$ and $\mathcal{M}'$ except that for the gaps between v and the immediate predecessors of v. The gap between w and u is 0, while the gaps between u and the previous immediate successors of w is decreased by 1, so the total sum of the gaps decreases through this process. Since our

Int-models are finite we repeat the process until condition 2 holds. Note that because the original model satisfies Condition 1, and because we do not change the root (we add a world in between two other existing worlds) the model $\mathcal{M}'$ must still satisfy Condition 1. ⊣

Note that this may break Condition 3, since the world v may already have a predecessor x with level $j+1$, but x is not a successor of w. When we introduce the new world u we make u a successor of x, which causes Condition 3 to fail.

LEMMA 8. *If an* Int*-countermodel to* φ_c^{Int} *exists, then one satisfying all three conditions of Definition 5 exists.*

PROOF. Let $\mathcal{M} = (W, R, L, r)$ be an Int-countermodel of φ_c^{Int} satisfying conditions 1 and 2 after applying Lemma 7, with worlds $a, b \in W$ such that $Lv(a) = Lv(b)$, $R(a, b)$ and $a \neq b$, thus breaking condition 3.

There must be a pair of "adjacent" worlds w and u such that $Lv(w) = Lv(u)$, $R(w, u)$, $w \neq u$ and there is no distinct v such that $R(w, v)$ and $R(v, u)$. We show that we get closer to satisfying condition 3 by removing the edge $R(w, u)$. Let $\mathcal{M}' = (W, R', L, r)$ where $R' = R \setminus \{(w, u)\}$.

The relation R' is still transitive because R was, and there is no "intermediate" world v that could require the removed edge. Reflexivity and antisymmetry are also preserved.

Suppose that $\mathcal{M}, r \Vdash P(\varphi)$, but $\mathcal{M}', r \nVdash P(\varphi)$. The only change is the removal of $R(w, u)$, so it is simple to see that $\mathcal{M}' \Vdash Lev(\varphi)$ and $\mathcal{M}' \Vdash Mid^n(\varphi)$. Therefore we must have $\mathcal{M}' \nVdash \mathrm{Box}_\psi^n(\varphi)$. Thus there must be some world x such that $\mathcal{M}', x \Vdash l^n$ and $\mathcal{M}', x \nVdash b_\psi^n \to A_\psi^n(\varphi)$ or $\mathcal{M}', x \nVdash A_\psi^n(\varphi) \to b_\psi^n$. We consider each case to obtain a contradiction.

Suppose that $\mathcal{M}', x \nVdash b_\psi^n \to A_\psi^n(\varphi)$. Expanding the semantics, there must therefore be some indices k and m and some world y such that $R(x, y)$ and $\mathcal{M}', y \Vdash b_\psi^n$ and $\mathcal{M}', y \Vdash l^k$ and $\mathcal{M}', y \nVdash (\psi)_m^k$. All propositional variables referred to by $(\psi)_m^k$ will have superscript k, and since $\mathcal{M}', y \Vdash \mathrm{Mid}^k(\varphi)$ we must have $\mathcal{M}', y \Vdash \phi^k$ or $\mathcal{M}', y \Vdash \phi^k \to \bot$ for all propositional variables ϕ^k, thus the valuations are fixed in all successors. The valuations are common between $\mathcal{M}$ and $\mathcal{M}'$, thus $\mathcal{M}, y \nVdash (\psi)_m^k$ as well, and so $\mathcal{M} \nVdash \mathrm{Box}_\psi^n$, a contradiction

Suppose instead that $\mathcal{M}', x \nVdash A_\psi^n(\varphi) \to b_\psi^n$. There must therefore be a world y such that $R(x, y)$ and $\mathcal{M}', y \Vdash A_\psi^n(\varphi)$ and $\mathcal{M}', y \nVdash b_\psi^n$. Because $\mathcal{M} \Vdash \mathrm{Box}_\psi^n(\varphi)$ and $\mathcal{M}, y \Vdash l^n$, we must have $\mathcal{M}, y \Vdash A_\psi^n(\varphi) \to b_\psi^n$, and because $\mathcal{M}, y \nVdash b_\psi^n$ we must have $\mathcal{M}, y \nVdash A_\psi^n(\varphi)$. Thus the witness falsifying $A_\psi^n(\varphi)$ must be u, and y must be w (otherwise the witness would still exist in $\mathcal{M}'$); that is $\mathcal{M}, u \Vdash l^k$ and $\mathcal{M}, u \nVdash \psi_m^k$ for some k and m. However, this means that $Lv(u) \geq k$, and thus $Lv(w) \geq k$. Since $\mathcal{M}, w \Vdash \mathrm{Mid}^k(\varphi)$ we have $\mathcal{M}, w \Vdash \phi^k$ or $\mathcal{M}, w \Vdash \phi^k \to \bot$, and since $R(w, u)$ we must have $\mathcal{M}, w \Vdash \phi^k$ iff $\mathcal{M}, u \Vdash \phi^k$. Thus we must have $\mathcal{M}', w \nVdash A_\psi^n(\varphi)$, a contradiction.

Thus $\mathcal{M}', r \Vdash P(\varphi)$, and $\mathcal{M}', r \nVdash \varphi_0^0$, and so $\mathcal{M}'$ is a countermodel with at least one fewer instance of Condition 3 failing. Since Int has the finite model property we can begin with a finite model (and a finite number of failures of conditon 3) and repeat the process until condition 3 holds. Since we only remove edges between worlds with the same level, we do not break either Condition 1 or Condition 2 if they hold initially. ⊣

COROLLARY 9. *If there is some* $\mathtt{Int}$*-countermodel to* φ_c^{Int} *then there is a stratified* $\mathtt{Int}$*-countermodel to* φ_c^{Int}.

PROOF. Given an arbitrary finite rooted $\mathtt{Int}$-countermodel to φ_c^{Int}, apply Lemma 6 to obtain a model satisfying Condition 1, then Lemma 7 to introduce worlds to satisfy Condition 2 without destroying Condition 1. Finally we use Lemma 8 to combine worlds with the same level to satisfy condition 3 without breaking condition 2 or condition 1. ⊣

We now show how $\mathtt{Int}$-countermodels of φ_c^{Int} correspond to $\mathtt{S4}$-countermodels of φ following Fernández [1] but being mindful of our modifications.

DEFINITION 10. Let $\mathcal{M}^{\mathtt{Int}} = (W^{\mathtt{Int}}, R^{\mathtt{Int}}, L^{\mathtt{Int}}, r^{\mathtt{Int}})$ be a stratified $\mathtt{Int}$ countermodel for φ_c^{Int}, such that $\mathcal{M}^{\mathtt{Int}}, r^{\mathtt{Int}} \nVdash (\varphi)_0^0$ and $\mathcal{M}^{\mathtt{Int}}, r^{\mathtt{Int}} \Vdash P(\varphi)$.

For each $x \in W^{\mathtt{Int}}$, let $\overline{x} = \{x_0, \dots, x_{M-1}\}$ be a set of M distinct worlds, and let $W^{\mathtt{Int}\mapsto\mathtt{S4}}$ be the disjoint union of all $\overline{x}$. Let $R^{\mathtt{Int}\mapsto\mathtt{S4}} = \{(x_m, y_n) \mid R^{\mathtt{Int}}(x, y)\}$, and $x_m \in L^{\mathtt{Int}\mapsto\mathtt{S4}}(p)$ iff $x \in L^{\mathtt{Int}}(p_m^{Lv(x)})$.

Define $\mathcal{M}^{\mathtt{Int}\mapsto\mathtt{S4}} = (W^{\mathtt{Int}\mapsto\mathtt{S4}}, R^{\mathtt{Int}\mapsto\mathtt{S4}}, L^{\mathtt{Int}\mapsto\mathtt{S4}}, r_0^{\mathtt{Int}})$.

LEMMA 11. *If* ψ *is a subformula of* φ*, then* $\mathcal{M}^{\mathtt{Int}\mapsto\mathtt{S4}}, x_m \Vdash \psi$ *iff* $\mathcal{M}^{\mathtt{Int}}, x \Vdash \psi_m^{Lv(x)}$.

PROOF. We proceed by induction on the structure of ψ. First the base cases:

$\psi = \bot$**:** Trivially true.

$\psi = p$**:** By the definition of $L^{\mathtt{Int}\mapsto\mathtt{S4}}$ the lemma holds.

Now the step cases, using the inductive hypothesis that for all formulae smaller than ψ the property already holds.

$\psi = \psi_1 \wedge \psi_2$**:** By definition, we have $\mathcal{M}^{\mathtt{Int}\mapsto\mathtt{S4}}, x_m \Vdash \psi_1 \wedge \psi_2$ iff $\mathcal{M}^{\mathtt{Int}\mapsto\mathtt{S4}}, x_m \Vdash \psi_i$ for all $i \in \{1, 2\}$. By the induction hypothesis, $\mathcal{M}^{\mathtt{Int}\mapsto\mathtt{S4}}, x_m \Vdash \psi_i$ iff $\mathcal{M}^{\mathtt{Int}}, x \Vdash (\psi_i)_m^{Lv(x)}$, and thus $\mathcal{M}^{\mathtt{Int}}, x \Vdash (\psi_1 \wedge \psi_2)_m^{Lv(x)}$ as required.

$\psi = \psi_1 \vee \psi_2$**:** As above.

$\psi = \psi_1 \to \psi_2$**:** If $\mathcal{M}^{\mathtt{Int}\mapsto\mathtt{S4}}, x_m \Vdash \psi_1 \to \psi_2$ then x_m either satisfies ψ_2 or falsifies ψ_1. By induction this translates to $\mathcal{M}^{\mathtt{Int}}$, thus $\mathcal{M}^{\mathtt{Int}}, x \nVdash (\psi_1)_m^{Lv(x)}$ or $\mathcal{M}^{\mathtt{Int}}, x \Vdash (\psi_2)_m^{Lv(x)}$. Both of these formulae refer to only propositional atoms indexed by m, and so because $\mathrm{Mid}^{Lv(x)}(\varphi)$ holds, all successors of x will give the same valuation, and thus either satisfy $(\psi_2)_m^{Lv(x)}$ or falsify $(\psi_1)_m^{Lv(x)}$, and so $\mathcal{M}^{\mathtt{Int}}, x \Vdash (\psi_1 \to \psi_2)_m^{Lv(x)}$

If instead $\mathcal{M}^{\mathtt{Int}}, x \Vdash (\psi_1 \to \psi_2)_m^{Lv(x)}$, then because $R^{\mathtt{Int}}$ is reflexive we must have $\mathcal{M}^{\mathtt{Int}}, x \nVdash (\psi_1)_m^{Lv(x)}$ or $\mathcal{M}^{\mathtt{Int}}, x \Vdash (\psi_2)_m^{Lv(x)}$. Using the inductive hypothesis, we thus have $\mathcal{M}^{\mathtt{Int}\mapsto\mathtt{S4}}, x_m \Vdash \psi_1 \to \psi_2$ as required.

$\psi = \Box\psi_1$**:** Because $\mathrm{Box}_{\psi_1}^{Lv(x)}(\varphi)$ holds, we have $\mathcal{M}^{\mathtt{Int}}, x \Vdash b_{\psi_1}^{Lv(x)}$ iff $\forall y. R^{\mathtt{Int}}(x, y)$ implies $\forall k. \mathcal{M}^{\mathtt{Int}}, y \Vdash (\psi_1)_k^{Lv(y)}$. By induction, for each of these worlds y we have $\mathcal{M}^{\mathtt{Int}\mapsto\mathtt{S4}}, y_k \Vdash \psi_1$. By the definition of $R^{\mathtt{Int}\mapsto\mathtt{S4}}$, these y_k are exactly the worlds such that $R^{\mathtt{Int}\mapsto\mathtt{S4}}(x_m, y_k)$, thus we have $\mathcal{M}^{\mathtt{Int}}, x \Vdash b_{\psi_1}^{Lv(x)}$ iff

$\forall y_k.R(x_m, y_k)$ implies $\mathcal{M}^{\texttt{Int}\mapsto\texttt{S4}}, y_k \Vdash \psi_1$. This is exactly the definition of $\mathcal{M}^{\texttt{Int}\mapsto\texttt{S4}}, x_m \Vdash \Box\psi_1$.

⊣

COROLLARY 12. *If there is an* Int*-countermodel to* φ_c^{Int} *then there is an* S4*-countermodel to* φ*. Equivalently, if* φ *is* S4*-valid, then* φ_c^{Int} *is* Int*-valid.*

PROOF. By Corollary 9 if there is an Int-countermodel to φ_c^{Int} then there must be a stratified Int-countermodel $\mathcal{M}^{\texttt{Int}}$ as well. Construct $\mathcal{M}^{\texttt{Int}\mapsto\texttt{S4}}$ as described in Definition 10. Applying Lemma 11 to $\mathcal{M}^{\texttt{Int}\mapsto\texttt{S4}}$ and choosing $\psi = \varphi$, we find that because $\mathcal{M}^{\texttt{Int}}, r^{\texttt{Int}} \nVdash (\varphi)_0^0$ and $Lv(r^{\texttt{Int}}) = 0$, we must have $\mathcal{M}^{\texttt{Int}\mapsto\texttt{S4}}, r_0^{\texttt{Int}} \nVdash \varphi$, as required. ⊣

4.2. Converting S4-models to Int-models. It remains to show that the converse holds, that if there is an S4-countermodel to φ then there is an Int-countermodel to φ_c^{Int}.

We will use the same notion of N-standard frames as Fernández [1], though we refer to it as M-standard to avoid confusion between the N used by Fernández [1] and the $M = N+1$ that we use. If $K = (W, R)$ is an S4-frame, then let $\overline{x}$ denote the R-equivalence class of worlds $\{y \mid (x, y) \in R \text{ and } (y, x) \in R\}$. The quotient W/R with induced relation $\overline{R}$ forms a partial order since R is transitive and reflexive, and taking the quotient ensures that it is antisymmetric as well.

DEFINITION 13. An S4 Kripke frame $K = (W, R)$ is *M-standard* if:

1. Any strictly ascending chain in $\overline{R}$ has length shorter than M;
2. For all $x \in W$, $\overline{x}$ has exactly M elements, $\{x_0, \ldots, x_{M-1}\}$;
3. $(W/R, \overline{R})$ forms a tree.

Fernández [1] proves the following theorem:

THEOREM 14 (Theorem 5.1 of [1]). *If* $\mathcal{M} = (W, R, L, r)$ *is an* S4*-model, and* φ *is a formula of* S4*, then there is an M-standard model* $\mathcal{M}^\varphi$*, such that for all subformulae* ψ *of* φ*, we have* $\mathcal{M}^\varphi, r^\varphi \Vdash \psi$ *iff* $\mathcal{M}, r \Vdash \psi$*.*

Thus if there is a countermodel to φ, then there is an M-standard countermodel to φ. Let $\mathcal{M}^{\texttt{S4}} = (W^{\texttt{S4}}, R^{\texttt{S4}}, L^{\texttt{S4}}, r^{\texttt{S4}})$ be such a model. Let $Lv(\overline{x})$ be the length of the shortest chain $\overline{R}(\overline{r^{\texttt{S4}}}, w_1), \overline{R}(w_1, w_2), \ldots, \overline{R}(w_{n-1}, \overline{x})$ where each w_i is distinct, and there is no intermediate such that $\overline{R}(w_i, u)$ and $\overline{R}(u, w_{i+1})$. We now define an Int-model which is a countermodel to φ_c^{Int}.

DEFINITION 15. Define $\mathcal{M}^{\texttt{S4}\mapsto\texttt{Int}} = (W^{\texttt{S4}\mapsto\texttt{Int}}, R^{\texttt{S4}\mapsto\texttt{Int}}, L^{\texttt{S4}\mapsto\texttt{Int}}, r^{\texttt{S4}\mapsto\texttt{Int}})$, where

- $W^{\texttt{S4}\mapsto\texttt{Int}} = W^{\texttt{S4}}/R^{\texttt{S4}}$
- $R^{\texttt{S4}\mapsto\texttt{Int}} = \overline{R^{\texttt{S4}}}$
- $r^{\texttt{S4}\mapsto\texttt{Int}} = \overline{r^{\texttt{S4}}}$
- $\overline{w} \in L^{\texttt{S4}\mapsto\texttt{Int}}(l^i)$ iff $Lv(\overline{w}) \geq i$
- $\overline{w} \in L^{\texttt{S4}\mapsto\texttt{Int}}(p_m^i)$ iff $Lv(\overline{w}) = i$ and $w_m \in L^{\texttt{S4}}(p)$, or $Lv(\overline{w}) > i$ and the immediate predecessor of $\overline{w}$ in $R^{\texttt{S4}\mapsto\texttt{Int}}$ is $\overline{v}$ with $\overline{v} \in L^{\texttt{S4}\mapsto\texttt{Int}}(p_m^i)$
- $\overline{w} \in L^{\texttt{S4}\mapsto\texttt{Int}}(b_\psi^i)$ iff $Lv(\overline{w}) = i$ and $\mathcal{M}^{\texttt{S4}}, w_0 \Vdash \Box\psi$, or $Lv(\overline{w}) > i$ and the immediate predecessor of $\overline{w}$ in $R^{\texttt{S4}\mapsto\texttt{Int}}$ is $\overline{v}$ with $\overline{v} \in L^{\texttt{S4}\mapsto\texttt{Int}}(p_m^i)$.

Now we prove that $\mathcal{M}^{\texttt{S4}\mapsto\texttt{Int}}$ is in fact an Int-model, $\mathcal{M}^{\texttt{S4}\mapsto\texttt{Int}}, r^{\texttt{S4}\mapsto\texttt{Int}} \Vdash P(\varphi)$, and $\mathcal{M}^{\texttt{S4}\mapsto\texttt{Int}}, r^{\texttt{S4}\mapsto\texttt{Int}} \nVdash \varphi_0^0$.

Lemma 16. *$\mathcal{M}^{S4\mapsto Int}$ is an Int-model.*

Proof. First, $R^{S4\mapsto Int}$ is transitive, reflexive, because R^{S4} was, and it is antisymmetric because clusters have been collapsed to their equivalence class. We must show that $L^{S4\mapsto Int}$ is persistent.

If $R^{S4\mapsto Int}(\overline{w},\overline{v})$, then $Lv(\overline{w}) \leq \overline{v}$ from the definition of Lv. Thus if $\overline{w} \in L^{S4\mapsto Int}(l^i)$, then $Lv(\overline{v}) \geq i$ and so $\overline{v} \in L^{S4\mapsto Int}(l^i)$ as required.

For the other propositions, the truth is defined inductively based on the truth at predecessors, so if $\overline{w} \in L^{S4\mapsto Int}(p^n_m)$ then any successor $\overline{v}$ will also be in $L^{S4\mapsto Int}(p^n_m)$, as required. ⊣

Lemma 17. *For all subformulae ψ of φ, $\mathcal{M}^{S4}, w_m \Vdash \psi$ iff $\mathcal{M}^{S4\mapsto Int}, \overline{w_m} \Vdash (\psi)^{Lv(w_m)}_m$.*

Proof. Much of the proof is the same as for Lemma 11. The only difference is for $\Box$-formulae.

By the definition of $L^{S4\mapsto Int}(b^n_{\psi_1})$ we have $\overline{w_m} \in L(b^{Lv(w_m)}_{\psi_1})$ iff $\mathcal{M}^{S4}, w_0 \Vdash \Box\psi_1$, and thus $\mathcal{M}^{S4}, w_m \Vdash \Box\psi_1$ since w_0 and w_m must have the same set of successors. Therefore $\mathcal{M}^{S4}, w_m \Vdash \Box\psi_1$ iff $\mathcal{M}^{S4\mapsto Int}, \overline{w_m} \Vdash (\Box\psi_1)^{Lv(w_m)}_m$, as required. ⊣

Lemma 18. *We have $\mathcal{M}^{S4\mapsto Int}, r^{S4\mapsto Int} \Vdash P(\varphi)$ in the constructed intuitionistic model.*

Proof. From the definition of $L^{S4\mapsto Int}$ we obviously have $\mathcal{M}^{S4\mapsto Int}, \overline{w} \Vdash l^{i+1} \to l^i$. We also have $\mathcal{M}^{S4\mapsto Int}, \overline{w} \Vdash l^0$, since $Lv(\overline{w}) \geq 0$. Also, because the models are M-standard, the maximum chain length is $M-1$, thus $Lv(\overline{w}) < M$ and so $\mathcal{M}^{S4\mapsto Int}, \overline{w} \vdash \neg l^M$. Thus $\mathcal{M}^{S4\mapsto Int} \Vdash \mathrm{Lev}(\varphi)$.

Next, if $Lv(\overline{w}) = i$ then $\overline{w} \in L^{S4\mapsto Int}(p^i_m)$ iff $w_m \in L^{S4}(p)$ for all atomic propositions p. All successors $\overline{v}$ of $\overline{w}$ must have $Lv(\overline{v}) > i$ and thus if $\overline{w} \notin L^{S4\mapsto Int}(p^i_m)$ then $\mathcal{M}^{S4\mapsto Int}, \overline{w} \Vdash \neg(p^i_m)$. Thus we have $\mathcal{M}^{S4\mapsto Int} \Vdash l^n \to p^n_m \vee \neg p^n_m$ for all n, m and p. A similar argument applies to b^i_ψ. Thus we have $\mathcal{M}^{S4\mapsto Int} \Vdash \mathrm{Mid}^n(\varphi)$ for all n.

The base case of the definition of $\overline{w} \in L^{S4\mapsto Int}(b^i_\psi)$ requires that $\mathcal{M}^{S4}, w_0 \Vdash \Box\psi$ which is exactly when all R^{S4} successors v_m of w_0 satisfy ψ. Any such v_m will correspond to a $\overline{v_m}$ with $Lv(\overline{v_m}) \geq Lv(\overline{w})$, and it will satisfy $\psi^{Lv(\overline{v_m})}_m$ due to Lemma 17. Thus if $\mathcal{M}^{S4\mapsto Int}, \overline{w} \Vdash b^n_\psi$, then all successors $\overline{v}$ will satisfy $l^k \to (\psi^k_m)$ for any $k \geq n$ and any m. Similarly, if $\overline{w} \notin L^{S4\mapsto Int}(b^{Lv(w)}_\psi$ then there must be some successor v_m of w_0 such that $\mathcal{M}^{S4}, v_m \nVdash \psi$ and thus $\mathcal{M}^{S4\mapsto Int}, \overline{v_m} \nVdash (\psi)^k_m$ for $k = Lv(v_m)$. Thus $\mathcal{M}^{S4\mapsto Int} \Vdash \mathrm{Box}^n_\psi(\varphi)$ as required. ⊣

Corollary 19. *If $\mathcal{M}^{S4}, r^{S4} \nVdash \varphi$ then we have $\mathcal{M}^{S4\mapsto Int}, r^{S4\mapsto Int} \nVdash \varphi^{\mathrm{Int}}_c$.*

Proof. Lemma 18 gives us $\mathcal{M}^{S4\mapsto Int}, r^{S4\mapsto Int} \Vdash P(\varphi)$, and then because $\mathcal{M}^{S4}, r^{S4} \nVdash \varphi$ and $Lv(\overline{r^{S4}}) = 0$, using Lemma 17 we have $\mathcal{M}^{S4\mapsto Int}, r^{S4\mapsto Int} \nVdash (\varphi)^0_0$. Thus we have $\mathcal{M}^{S4\mapsto Int}, r^{S4\mapsto Int} \nVdash \varphi^{\mathrm{Int}}_c$ as required. ⊣

Main theorem We have shown that there is an S4-countermodel to φ if and only if there is an Int-countermodel to φ^{Int}_c, and thus φ^{Int}_c is a faithful translation from S4 to Int.

Remark 20. We might ask where Fernández [1] goes awry; where does the purported proof fail? The theorems and lemmas presented there appear to be correct, and yet Example 1 demonstrates that the original translation is wrong.

The problem is with his application of his Theorem 4.1, which is our Lemma 11. The theorem states that worlds in the constructed `S4`-model satisfy the same formulae as the worlds in the original `Int`-model of the same level. The assumption that Fernández [1] makes is effectively that the original `Int`-models are stratified, and in particular that the root of the `Int`-countermodel has level 0. If this is the case, then applying his Theorem 4.1 will indeed result in an `S4`-model where the root falsifies φ, because the `Int`-model falsifies $(\varphi)^0_0$. What we illustrated with Example 1 was that this assumption does not always hold for the original definition of φ_o^{Int}, and indeed because his Theorem 4.1 is correct the example cannot be "fixed" into a stratified model.

By changing the translation as we have, we are able to prove that all models of the modified translation can be converted into stratified models according to Corollary 9, and then Fernández's original proofs only require slight changes to account for the changed translation to prove that this new translation is in fact correct. Our Lemmas 6 to 8 which we use to prove Corollary 9 are the bulk of the new work here, and they do not hold for the original translation.

An implementation of our translation is available at the URL below:

`http://users.cecs.anu.edu.au/~rpg/S4ToInt/`

There are also options to apply the original translation φ_o^{Int} of Fernández [1], as well as that translation using our M instead of N. Thus the reader can test that: our translation φ_c^{Int} is correct; the original translation φ_o^{Int} is incorrect; and that even changing N to M in the original translation is still incorrect.

Acknowledgements: We thank the reviewer for the detailed comments on our previous submission(s) which have improved the paper immensely.

REFERENCES

[1] David Fernández, *A polynomial translation of s4 into intuitionistic logic*, ***The Journal of Symbolic Logic***, vol. 71 (2006), no. 3, pp. 989–1001.

[2] Kurt Gödel, ***Kurt Gödel: Collected Works: Volume I: Publications 1929-1936***, vol. 1, Oxford university press, 1986.