Notre Dame Journal of Formal Logic
Volume 52, Number 3, 2011

Being Wrong: Logics for False Belief

Christopher Steinsvold

Abstract We introduce an operator to represent the simple notion of being wrong. Read Wp to mean: the agent is wrong about p. Being wrong about p means believing p though p is false. We add this operator to the language of propositional logic and study it. We introduce a canonical model for logics of being wrong, show completeness for the minimal logic of being wrong and various other systems. En route we examine the expressiveness of the language. In conclusion, we discuss an open question regarding K4.

1 Introduction

If agent 1 claims that p is true, and agent 2 responds with "You're wrong about p," or more simply "You're wrong," agent 2 means that agent 1 believes p and p is false. Given the language of modal logic, we can define *being wrong about* φ with,

$$W\varphi \stackrel{\text{def}}{=} (\Box\varphi \wedge \neg\varphi).$$

Our interest is in studying the logic of $W\varphi$ in isolation from the basic modal language. This is best compared to the study of logics of contingency, where an operator representing $\Diamond\varphi \wedge \Diamond\neg\varphi$ is studied. In similar fashion we take W as primitive and define the language of being wrong, $\mathfrak{L}^W$, with,

$$\mathfrak{L}^W = \{W, \bot, \rightarrow, (\,,), p_1, p_2, \ldots\}.$$

We define the set of formulas, $\mathfrak{F}^W$, with the following BNF,

$$\varphi \in \mathfrak{F}^W := p \,|\, \bot \,|\, (\varphi_1 \rightarrow \varphi_2) \,|\, W\varphi.$$

In comparison, the modal logic KT can be viewed as the minimal logic of being correct. That is, we may view KT as the logic of an operator representing $\Box\varphi \wedge \varphi$. With this in mind, we are simply inspecting the other side of a well-known coin.

Received November 9, 2009; accepted April 18, 2011; printed July 25, 2011
2010 Mathematics Subject Classification: Primary, 03B42
Keywords: modal logic, wrong, incorrect, false, belief
 10.1215/00294527-1435438

Definition 1.1 For any normal modal logic L, let L_W be the set of all theorems of L expressible only in terms of the Boolean connectives and W (where W represents $\Box\varphi \wedge \neg\varphi$).

The focus here is on completeness results for various L_W. Given the doxastic nature of being wrong, our focus will be on those L_W where L contains a common axiom for belief. We present a simple canonical model to show completeness for the minimal logic of being wrong, among others. Though we were able to show completeness for a number of L_W where L contains the 4 axiom (e.g., $K45_W$), we were unable to find a complete set of axioms for $K4_W$ itself. We conclude with a discussion of this problem. Along the way, we examine what is expressible in the language of being wrong (in comparison with the usual modal language). One simple result is a completeness proof for KT_W. Epistemically speaking this is awkward (it is the logic of being wrong for an agent who is never wrong), but technically it tells us much information about the expressiveness of our language.

Thematically, the study of logics of being wrong is related to the study of logics of ignorance (see Lomuscio and van der Hoek [7], Steinsvold [12], where an operator representing as $\neg K\varphi \wedge \neg K\neg\varphi$ is studied). Though the motivation differs, the study of logics of ignorance is formally the same as the study of contingency logics, where contingency represents $\Diamond\varphi \wedge \Diamond\neg\varphi$ (viz., Aristotle—see Brogan [2], Cresswell [3], Humberstone [4], Kuhn [5], Montgomery and Routley [9], Mortensen [10], Zolin [14]). There is also the study of logics of essence and accident (see Marcos [8], Steinsvold [11], [13], Kushido [6]), where an operator symbolizing $\varphi \wedge \Diamond\neg\varphi$ is studied. Though there is no obvious alethic interpretation for the operator studied here, the study of logics of being wrong seems like a natural next step in such studies.

2 Possible World Semantics

A *frame* $F = \langle W, R\rangle$ is a nonempty set W together with $R \subseteq W \times W$. Members of W are *worlds* or *points*. A *valuation* V is a function from the set of propositional variables into the power set of W. $M = \langle W, R, V\rangle$ is a *model*. We define *truth in a model at a world* as follows:

$$M, w \models p \text{ iff } w \in V(p),$$
$$M, w \models \bot \text{ iff } 0 = 1,$$
$$M, w \models \varphi \to \psi \text{ iff if } M, w \models \varphi \text{ then } M, w \models \psi,$$
$$M, w \models W\varphi \text{ iff } M, w \not\models \varphi \text{ and } (\forall x)(\text{if } wRx \text{ then } M, x \models \varphi).$$

φ is *valid in model* M if and only if φ is true at every point in M. φ is *valid in a frame* F if and only if φ is valid in every M based on F.

Worlds which relate to themselves and only to themselves turn out to be surprisingly useful.

Definition 2.1 Call w *narcissistic* if and only if w relates to itself and only to itself. Call a frame *narcissistic* if and only if all the worlds are narcissistic; that is,

$$(\forall x)(xRx \wedge (\forall y)(xRy \to x = y)).$$

The following will be useful in showing completeness for $K45_W$.

Lemma 2.2 *Let $M = \langle W, R, V\rangle$, let $W^r = \{x \in W | xRx\}$, and let $R^+ \subseteq W^r \times W$. Where $M^+ = \langle W, R \cup R^+, V\rangle$, and $w \in W$,*

$$M, w \models \varphi \text{ iff } M^+, w \models \varphi.$$

Proof The nonmodal cases are straightforward. First, observe that for all $x \notin W^r$ and $y \in W$ we have

$$\langle x, y\rangle \in R \text{ iff } \langle x, y\rangle \in R \cup R^+.$$

Clearly, $R \subseteq R \cup R^+$. And if x is not reflexive and $\langle x, y\rangle \in R \cup R^+$, then $\langle x, y\rangle \in R$. This fact makes our proof simple.

Assume $M, w \models W\varphi$, by the definition of truth, $M, w \not\models \varphi$; thus by induction hypothesis (IH), $M^+, w \not\models \varphi$. Clearly, w can't bear R to itself; thus $w \notin W^r$. Thus $\langle w, y\rangle \in R$ if and only if $\langle w, y\rangle \in R \cup R^+$, for all $y \in W$. And if wRz, then $M, z \models \varphi$; so by induction hypothesis $M^+, w \models W\varphi$. The converse is similar. □

3 The System S^W

Call the following axiom system S^W.

A1 $W\varphi \to \neg\varphi$
A2 $(W\varphi \wedge W\psi) \to W(\varphi \wedge \psi)$
R1 $\vdash \varphi \to \psi \Rightarrow \vdash (W\varphi \wedge \neg\psi) \to W\psi$,
with MP, all propositional tautologies, and uniform substitution.

In the next section we show that $S^W = K_W$; that is, S^W is the minimal logic of being wrong. We leave the soundness of S^W for the reader. To denote the smallest extension of S^W which includes an axiom A and is closed under all the rules of S^W, we write $S^W \oplus$ A. We now prove some theorems.

From the following we can derive substitution of equivalents in S^W.

Theorem 3.1 $\vdash \varphi \leftrightarrow \psi \Rightarrow \vdash W\varphi \leftrightarrow W\psi$.

Proof

(1) $\vdash \varphi \leftrightarrow \psi$, assumption.
(2) $\vdash \neg\varphi \leftrightarrow \neg\psi$, from (1).
(3) $\vdash \varphi \to \psi$, from (1).
(4) $\vdash \psi \to \varphi$, from (1).
(5) $\vdash (W\varphi \wedge \neg\psi) \to W\psi$, from (3) and R1.
(6) $\vdash ((W\varphi \wedge \neg\psi) \to W\psi) \to [(\neg\psi \leftrightarrow \neg\varphi) \to ((W\varphi \wedge \neg\varphi) \to W\psi)]$, instance of the tautology,

$$((A \wedge B) \to C) \to [(B \leftrightarrow D) \to ((A \wedge D) \to C)].$$

(7) $\vdash (\neg\varphi \leftrightarrow \neg\psi) \to ((W\varphi \wedge \neg\varphi) \to W\psi)$, MP, (5), and (6).
(8) $\vdash (W\varphi \wedge \neg\varphi) \to W\psi$, MP, (2), and (7).
(9) $\vdash W\varphi \to \neg\varphi$, A1.
(10) $\vdash W\varphi \to (W\varphi \wedge \neg\varphi)$, from (9).
(11) $\vdash W\varphi \to W\psi$, from (8) and (10).
And, just as we derived line (11), we can give an analogous proof for
(12) $\vdash W\psi \to W\varphi$.
(13) $\vdash W\psi \leftrightarrow W\varphi$, from (11) and (12).

□

Theorem 3.2 $\vdash (W(\varphi \wedge \psi) \wedge \neg\psi) \to W\psi$.

Proof

(1) $\vdash (\varphi \wedge \psi) \to \psi$, PC.
(2) $\vdash (W(\varphi \wedge \psi) \wedge \neg\psi) \to W\psi$, from (1) and R1. □

Theorem 3.3 $\vdash (W(\varphi \wedge \psi) \wedge \neg\varphi \wedge \neg\psi) \leftrightarrow (W\varphi \wedge W\psi)$.

Proof

(1) $\vdash W\varphi \to \neg\varphi$, A1.
(2) $\vdash W\psi \to \neg\psi$, A1.
(3) $\vdash (W\varphi \wedge W\psi) \to W(\varphi \wedge \psi)$, A2.
(4) $\vdash (W\varphi \wedge W\psi) \to (W(\varphi \wedge \psi) \wedge \neg\varphi \wedge \neg\psi)$, from (1), (2), and (3).
(5) $\vdash (W(\varphi \wedge \psi) \wedge \neg\psi) \to W\psi$, Theorem 3.2.
(6) $\vdash (W(\varphi \wedge \psi) \wedge \neg\varphi) \to W\varphi$, variation of Theorem 3.2.
(7) $\vdash (W(\varphi \wedge \psi) \wedge \neg\varphi \wedge \neg\psi) \to (W\varphi \wedge W\psi)$, from (5) and (6).
(8) $\vdash (W(\varphi \wedge \psi) \wedge \neg\varphi \wedge \neg\psi) \leftrightarrow (W\varphi \wedge W\psi)$, from (4) and (7). □

Presumably there is a shorter proof of the following.

Theorem 3.4 $\vdash W(\varphi \wedge \psi) \to (W\varphi \vee W\psi)$.

Proof

(1) $\vdash (W(\varphi \wedge \psi) \wedge \neg\psi) \to W\psi$, Theorem 3.2.
(2) $\vdash (W(\varphi \wedge \psi) \wedge \neg W\psi) \to \psi$, from (1).
(3) $\vdash W(\varphi \wedge \psi) \to \neg(\varphi \wedge \psi)$, instance of A1.
(4) $\vdash W(\varphi \wedge \psi) \to (\psi \to \neg\varphi)$, from (3).
(5) $\vdash (W(\varphi \wedge \psi) \wedge \neg W\psi) \to \neg\varphi$, from (2) and (4).
(6) $\vdash (W(\varphi \wedge \psi) \wedge \varphi) \to W\psi$, from (5).
And just as we derived line (6), we can similarly derive
(7) $\vdash (W(\varphi \wedge \psi) \wedge \psi) \to W\varphi$.
(8) $\vdash (W(\varphi \wedge \psi) \wedge (\varphi \vee \psi)) \to (W\varphi \vee W\psi)$, from (6) and (7).
(9) $\vdash (W(\varphi \wedge \psi) \wedge \neg\varphi \wedge \neg\psi) \to (W\varphi \wedge W\psi)$, from Theorem 3.3.
(10) $\vdash (W(\varphi \wedge \psi) \wedge \neg(\varphi \vee \psi)) \to (W\varphi \wedge W\psi)$, from (9), DeMorgan.
(11) $\vdash (W(\varphi \wedge \psi) \wedge \neg(\varphi \vee \psi)) \to (W\varphi \vee W\psi)$, weakening consequent of (10).
(12) $\vdash (W(\varphi \wedge \psi) \to (W\varphi \vee W\psi)$, from (8) and (11). □

4 The Canonical Model

Let W^W be the set of all maximally consistent sets of the logic S^W.

Definition 4.1 Where $w \in W^W$, call w a w^r world if and only if $\neg(\exists W\varphi)(W\varphi \in w)$, and call w a $w^{\neg r}$ world if and only if $(\exists W\varphi)(W\varphi \in w)$.

Clearly, every world in W^W is either a w^r world or a $w^{\neg r}$ world.

Definition 4.2 Where $w, x \in W^W$, we define R^W on $W^W \times W^W$ with

1. if w is a w^r world, let wR^Wx if and only if $w = x$, and
2. if w is a $w^{\neg r}$ world, let wR^Wx if and only if $(\forall W\varphi)$(if $W\varphi \in w$ then $\varphi \in x$).

Each w^r is narcissistic. Each $w^{\neg r}$ is nonreflexive (by axiom A1). Let $w \in V^W(p)$ if and only if $p \in w$; thus we have $M^W = \langle W^W R^W V^W \rangle$.

Lemma 4.3 *For all $w \in W^W$,*

$$M^W, w \models \psi \text{ iff } \psi \in w.$$

Proof The nonmodal cases are straightforward. Assume $W\varphi \in w$. By axiom A1, $\neg\varphi \in w$. By IH, $M^W, w \not\models \varphi$. Since $W\varphi \in w$, we know w is a $w^{\neg r}$ world. Thus if $wR^W y$, $\varphi \in y$, all y. By IH, if $wR^W y$, then $M^W, y \models \varphi$, all y. By definition of truth, $M^W, w \models W\varphi$.

Conversely, assume $W\varphi \notin w$. Let $\lambda(w) = \{\psi | W\psi \in w\}$. If $\lambda(w)$ is empty, then w is a w^r world, and since all w^r are narcissistic, $M^W, w \not\models W\varphi$. If $\lambda(w)$ is nonempty, then w is a $w^{\neg r}$ world.

To get a contradiction, assume $M^W, w \models W\varphi$. By the definition of truth, $M^W, w \not\models \varphi$, and by IH, $\neg\varphi \in w$. If $\lambda(w) \cup \{\neg\varphi\}$ is consistent, we are done (by Lindenbaum's lemma and IH).

Assume $\lambda(w) \cup \{\neg\varphi\}$ is inconsistent. Thus for some $\psi_1, \ldots, \psi_n \in \lambda(w)$,

$$\mathrm{S}^W \vdash (\psi_1 \wedge \cdots \wedge \psi_n) \rightarrow \varphi.$$

Since $\psi_1, \ldots, \psi_n \in \lambda(w)$, by repeated application of Axiom 2, we have

$$W(\psi_1 \wedge \cdots \wedge \psi_n) \in w.$$

We know $\neg\varphi \in w$; thus $W(\psi_1 \wedge \cdots \wedge \psi_n) \wedge \neg\varphi \in w$. We use rule R1 to conclude $W\varphi \in w$. Contradiction. $\square$

Theorem 4.4 $\mathrm{S}^W = \mathrm{K}_W$.

Proof K is the minimal modal logic. By definition, K_W is the minimal logic of being wrong. Since S^W is sound and complete, $\mathrm{K}_W = \mathrm{S}^W$. $\square$

Comment 4.5 We comment on an alternative definition of R^W. In our model, all the reflexive worlds are narcissistic, and this may seem unnatural. It may seem more natural to simply define R^W with

$$wR^W x \text{ iff } (\forall W\varphi)(\text{if } W\varphi \in w \text{ then } \varphi \in x),$$

for all w, x. One can show completeness for S^W with this definition, but this has a strange consequence: every w^r world will relate to every world.

Let A^T be $\neg W\varphi$.

Theorem 4.6 *$\neg W\varphi$ is valid in F if and only if R is reflexive.*

Proof For the reader. $\square$

The logic $\mathrm{K}\oplus\varphi \leftrightarrow \Box\varphi$, known as Triv, is logic of narcissistic frames.

Theorem 4.7 $\mathrm{S}^W \oplus \mathrm{A}^T = \mathrm{KT}_W = \mathrm{S4}_W = \mathrm{S5}_W = \mathrm{Triv}_W$.

Proof Consider the canonical model for $\mathrm{S}^W \oplus \mathrm{A}^T$. Since no world contains $W\varphi$ for any φ, each world is a w^r world, and so each world is narcissistic (thus $\mathrm{S}^W \oplus \mathrm{A}^T = \mathrm{Triv}_W$). Since $\mathrm{S}^W \oplus \mathrm{A}^T$ is sound and complete for reflexive models, $\mathrm{S}^W \oplus \mathrm{A}^T = \mathrm{KT}_W$. Since R^W is transitive as well, $\mathrm{S}^W \oplus \mathrm{A}^T = \mathrm{S4}_W$. Since R^W is also Euclidean, $\mathrm{S}^W \oplus \mathrm{A}^T = \mathrm{S5}_W$. $\square$

The proof of Theorem 4.7 yields the following semantic information.

Theorem 4.8 *There is no $\varphi \in \mathfrak{F}^W$ such that the validity of φ corresponds to any of the following properties of a frame:*

(1) *transitivity,*
(2) *Euclidean,*
(3) *symmetry,*
(4) *weakly connected* $(\forall x)(\forall y)(\forall z)((xRy \wedge xRz) \rightarrow (yRz \vee y = z \vee zRy))$,
(5) *weakly directed* $(\forall x)(\forall y)(\forall z)((xRy \wedge xRz) \rightarrow (\exists v)(yRv \wedge zRv))$,
(6) *partially functional* $(\forall x)(\forall y)(\forall z)((xRy \wedge xRz) \rightarrow z = y)$,
(7) *narcissistic* $(\forall x)(xRx \wedge (\forall y)(xRy \rightarrow x = y))$,
(8) *partially narcissistic* $(\forall x)(\forall y)(xRy \rightarrow x = y)$.

Proof To get a contradiction assume there is some $\varphi \in \mathfrak{F}^W$ and the validity of φ corresponds to one of the properties (1) through (8). Inspection of the canonical model of KT_W in the proof of Theorem 4.7 shows the canonical model for KT_W (trivially) has properties (1) through (8) (since each world is narcissistic). Thus, our presumed sentence φ is valid in the canonical frame for KT_W. Since KT_W is complete, φ is a theorem of KT_W. This means reflexivity implies the property which φ corresponds to. But reflexivity does not imply any of these properties. □

Theorem 4.7 was useful for showing Theorem 4.8, yet we can show something stronger than Theorem 4.7 syntactically (thanks to an anonymous reviewer): KT_W (i.e., $\mathrm{S}^W \oplus \mathrm{A}^T$) is a maximal logic.

Theorem 4.9 *For all φ,*

$$\varphi \notin \mathrm{KT}_W \text{ iff } \mathrm{KT}_W \oplus \varphi \vdash \bot.$$

Proof The direction from right to left is straightforward (KT_W is consistent). Assume $\varphi \notin \mathrm{KT}_W$. Note that $W\psi \leftrightarrow \bot \in \mathrm{KT}_W$, for all ψ. Let φ' be the result of replacing every occurrence of $W\psi$ in φ with $\bot$. Using Theorem 3.1, we have that $\mathrm{KT}_W \vdash \varphi \leftrightarrow \varphi'$. Thus $\varphi' \notin \mathrm{KT}_W$. Furthermore, φ' is a sentence of the propositional calculus, and so we know it is not a theorem of PC either. But by a classic result for PC we know

$$\varphi' \notin \mathrm{PC} \text{ iff } \mathrm{PC} \oplus \varphi' \vdash \bot.$$

Thus, $\mathrm{PC} \oplus \varphi' \vdash \bot$; moreover, $\mathrm{KT}_W \oplus \varphi' \vdash \bot$. And since φ and φ' are equivalent in KT_W, $\mathrm{KT}_W \oplus \varphi \vdash \bot$. □

As a consequence we have the following corollary.

Corollary 4.10 *If* $\mathrm{KT} \subseteq \mathrm{L}$ *and* L *is consistent,*

$$\mathrm{L}_W = \mathrm{KT}_W.$$

Proof If $\mathrm{KT} \subseteq \mathrm{L}$, then $\mathrm{KT}_W \subseteq \mathrm{L}_W$. Assume $\varphi \in \mathrm{L}_W$ but $\varphi \notin \mathrm{KT}_W$. By Theorem 4.9, $\mathrm{KT}_W \oplus \varphi$ is inconsistent, and so L_W is inconsistent. □

Besides completeness for the minimal logic of being wrong and KT_W, we have focused on negative results and the lack of expressiveness in the language. This lack of expressiveness largely revolves around the fact that our language becomes rather mute at reflexive worlds. We now focus on the positive. Let A^D be $\neg W\bot$.

Theorem 4.11 *$\neg W\bot$ is valid in F if and only if R is serial.*

Proof For the reader. □

Theorem 4.12 $\mathrm{S}^W \oplus \mathrm{A}^D = \mathrm{KD}_W$.

Proof Since KD is the modal logic of serial frames, KD_W is the logic of being wrong for serial frames. Consider the canonical model for $\mathrm{S}^W \oplus \mathrm{A}^D$. All the w^r worlds relate to something, namely, themselves. Suppose some $w^{\neg r}$ related to no world, and let $\lambda(w^{\neg r}) = \{\psi \mid W\psi \in w^{\neg r}\}$. $\lambda(w^{\neg r})$ must be inconsistent.

Thus for some $\psi_1, \ldots, \psi_n \in \lambda(w^{\neg r})$, $\mathrm{S}^W \oplus \mathrm{A}^D \vdash (\psi_1 \wedge \cdots \wedge \psi_n) \to \bot$. By Axiom 2, $W(\psi_1 \wedge \cdots \wedge \psi_n) \in w^{\neg r}$. Since $W(\psi_1 \wedge \cdots \wedge \psi_n) \wedge \neg\bot \in w$, we use rule R1 to conclude $W\bot \in w$. Contradiction. □

Let A^Q be $W\varphi \to W(\neg W\psi \wedge \varphi)$, and let Q be $\Box(\Box\varphi \to \varphi)$.

Theorem 4.13 *$W\varphi \to W(\neg W\psi \wedge \varphi)$ is valid in F if and only if*

$$(\forall x)(\forall y)(\textit{if } (xRy \wedge x\not R x) \textit{ then } yRy).$$

Proof For the reader. □

Theorem 4.14 $\mathrm{S}^W \oplus \mathrm{A}^Q = \mathrm{KQ}_W = \mathrm{K4Q}_W$.

Proof KQ is the logic of secondarily reflexive frames; that is,

$$Q \text{ is valid in } F \text{ iff } (\forall x)(\forall y)(\text{if } xRy \text{ then } yRy).$$

We show that the canonical model for $\mathrm{S}^W \oplus \mathrm{A}^Q$ is secondarily reflexive, which will imply the validity of axiom A^Q. Being narcissistic, each w^r world is secondarily reflexive. Thus considering any $w^{\neg r}$ world, we know $W\varphi \in w^{\neg r}$, for some $W\varphi$. Suppose $w^{\neg r} R^W y$. If y is a w^r world, done. If not then some $W\psi \in y$. We know that $W\varphi \to W(\neg W\psi \wedge \varphi) \in w^{\neg r}$. So $W(\neg W\psi \wedge \varphi) \in w^{\neg r}$, and since $w^{\neg r} R^W y$, $\neg W\psi \wedge \varphi \in y$; thus $\neg W\psi \in y$. Contradiction.

Thus $\mathrm{S}^W \oplus \mathrm{A}^Q = \mathrm{KQ}_W$. To see that $\mathrm{S}^W \oplus \mathrm{A}^Q = \mathrm{K4Q}_W$, note that we have already shown our frame is secondarily reflexive, which implies that if $wR^W y$ then y is narcissistic. Thus R^W is transitive. □

We now show $\mathrm{S}^W \oplus \mathrm{A}^Q = \mathrm{K45}_W$. The canonical frame is not Euclidean, but we can straightforwardly manipulate the frame and make it so.

Theorem 4.15 $\mathrm{S}^W \oplus \mathrm{A}^Q = \mathrm{K5}_W = \mathrm{K45}_W$.

Proof K45 is the modal logic of transitive and Euclidean frames. Consider any world w in the canonical model for $\mathrm{S}^W \oplus \mathrm{A}^Q$ and let $M = \langle W, R, V\rangle$ be the submodel generated by w. Note that our language is modal, so we don't need to prove a separate generated submodel theorem for it. Thus we know, for all $x \in W$,

$$M, x \models \varphi \text{ iff } M^W, x \models \varphi.$$

Let $Z(w) = \{x \in W | wRx\}$. As shown in the proof of Theorem 4.14, the canonical model for $\mathrm{S}^W \oplus \mathrm{A}^Q$ is secondarily reflexive, and so all the worlds in $Z(w)$ are reflexive. Create a new model $M' = \langle W, R', V\rangle$, where $R' = R \cup Z(w) \times Z(w)$. By Lemma 2.2,

$$M, x \models \varphi \text{ iff } M', x \models \varphi.$$

M' is clearly Euclidean, and we know Q is a theorem of K5. Thus since any nontheorem of $\mathrm{S}^W \oplus \mathrm{A}^Q$ fails in a Euclidean model, $\mathrm{S}^W \oplus \mathrm{A}^Q = \mathrm{K5}_W$. Since M' is also transitive, $\mathrm{S}^W \oplus \mathrm{A}^Q = \mathrm{K45}_W$. □

Putting previous results together we have $S^W \oplus A^Q \oplus A^D = KD45_W$.

5 K4$_W$?

As mentioned in the Introduction, a complete set of axioms for K4$_W$ has eluded us. In conclusion we informally enumerate various ideas which may help resolve this. As shown in Theorem 4.8, the property of transitivity is not definable in our language. This, however, is not the main obstacle.

The simpler difficulty is that iterations of W do not mimic iterations of $\Box$. The 4 axiom is $\Box\varphi \to \Box\Box\varphi$, but we do not have $W\varphi \to WW\varphi$ in K4$_W$ (in fact, $S^W \vdash W\varphi \to \neg WW\varphi$). Besides being technically recalcitrant, $WW\varphi$ is an epistemic bogey monster. To be wrong that you are wrong about φ implies $B(B\varphi \wedge \neg\varphi)$ (a Moorean paradox).

Aiming to show completeness for K4D$_W$ might be easier than aiming for completeness of K4$_W$. The following sentence is a member of K4D$_W$ and may be helpful to this end:

$$W\varphi \to W(\neg W\neg\varphi \wedge \varphi).$$

As an extension of K4, the logic GL (K$\oplus\Box(\Box\varphi \to \varphi) \to \Box\varphi$) is of interest here (see Boolos [1]). Again, perhaps aiming for a complete set of axioms for GL$_W$ may help resolve our question about K4$_W$. The following theorem of GL$_W$ may be of use to this end:

$$(W(\neg W\varphi \wedge \psi) \wedge \neg\varphi) \to W\varphi.$$

Note that we showed in Theorem 4.14 that $S^W \oplus A^Q = K4Q_W$; this tells us that $K4_W \subseteq S^W \oplus A^Q$, which narrows the search for our missing axiom(s) down somewhat. We end with some observations regarding K4$_W$ due to an anonymous reviewer. Call the following sentence $A^{3.9}$:

$$(W\varphi \wedge \neg W\bot) \to \neg W(W\neg\varphi \wedge \neg W\bot).$$

We have the following two results.

Theorem 5.1 $A^{3.9} \in K4_W$ *and* $A^{3.9} \notin K_W$.

Proof To see $A^{3.9} \in K4_W$, assume $A^{3.9}$ fails in some transitive model. Thus,

$$M, w \models W\varphi \wedge \neg W\bot \wedge W(W\neg\varphi \wedge \neg W\bot).$$

Since $M, w \models W\varphi \wedge \neg W\bot$, there is some x, wRx and $M, x \models \varphi$. We also know that $M, x \models W\neg\varphi \wedge \neg W\bot$, and so for some y, xRy and $M, y \models \neg\varphi$. By transitivity, wRy, thus $M, y \models \varphi$, contradiction. To see that $A^{3.9} \notin K_W$, consider a three-world model $W = \{a, b, c\}$, $R = \{\langle a, b\rangle, \langle b, c\rangle\}$, where $V(p) = \{b\}$. $A^{3.9}$ fails at a. $\Box$

Theorem 5.2 *$A^{3.9}$ is valid in a frame if and only if, for all x, either $\neg(\exists y)xRy$, or $(\exists y)(xRy \wedge \neg(\exists z)yRz)$, or $(\exists y)(\exists z)(xRy \wedge yRz \wedge xRz)$.*

Proof For the reader. $\Box$

References

[1] Boolos, G., *The Logic of Provability*, Cambridge University Press, Cambridge, 1993. Zbl 0891.03004. MR 1260008. 252

[2] Brogan, A. P., "Aristotle's logic of statements about contingency," *Mind*, vol. 76 (1967), pp. 46–81. 246

[3] Cresswell, M. J., "Necessity and contingency," *Studia Logica*, vol. 47 (1988), pp. 145–49. Zbl 0666.03015. MR 999771. 246

[4] Humberstone, I. L., "The logic of non-contingency," *Notre Dame Journal of Formal Logic*, vol. 36 (1995), pp. 214–29. Zbl 0833.03004. MR 1345745. 246

[5] Kuhn, S. T., "Minimal non-contingency logic," *Notre Dame Journal of Formal Logic*, vol. 36 (1995), pp. 230–34. Zbl 0833.03005. MR 1345746. 246

[6] Kushida, H., "The modal logic of Gödel sentences," *Journal of Philosophical Logic*, vol. 39 (2010), pp. 577–90. Zbl 1205.03066. MR 2721775. 246

[7] Lomuscio, A., and W. van der Hoek, "A logic for ignorance," *Electronic Notes in Theoretical Computer Science*, vol. 85 (2004). 246

[8] Marcos, J., "Logics of essence and accident," *Bulletin of the Section of Logic*, vol. 34 (2005), pp. 43–56. Zbl 1117.03305. MR 2186213. 246

[9] Montgomery, H., and R. Routley, "Non-contingency axioms for S4 and S5," *Logique et Analyse. Nouvelle Série*, vol. 11 (1968), pp. 422–24. Zbl 0169.30003. MR 0235990. 246

[10] Mortensen, C., "A sequence of normal modal systems with non-contingency bases," *Logique et Analyse. Nouvelle Série*, vol. 19 (1976), pp. 341–44. Zbl 0347.02014. MR 0497945. 246

[11] Steinsvold, C., "Completeness for various logics of essence and accident," *Bulletin of the Section of Logic*, vol. 37 (2008), pp. 93–101. MR 2460598. 246

[12] Steinsvold, C., "A note on logics of ignorance and borders," *Notre Dame Journal of Formal Logic*, vol. 49 (2008), pp. 385–92. Zbl 1180.03017. MR 2456654. 246

[13] Steinsvold, C., "The boxdot conjecture and the language of essence and accident," *The Australasian Journal of Logic*, vol. 10 (2011), pp. 18–35. 246

[14] Zolin, E. E., "Completeness and definability in the logic of noncontingency," *Notre Dame Journal of Formal Logic*, vol. 40 (1999), pp. 533–47. Zbl 0989.03019. MR 1858241. 246

Acknowledgments

I thank an anonymous reviewer for a constructive and thoughtful response.

Department of Philosophy
Brooklyn College
City University of New York
2900 Bedford Avenue
Brooklyn NY 11210
USA
steinsvold1@verizon.net