THOUGHT
A JOURNAL OF PHILOSOPHY
THE NORTHERN INSTITUTE OF PHILOSOPHY

Thought ISSN 2161-2234

ORIGINAL ARTICLE

Moore's Paradox, Introspection and Doxastic Logic

Adam Rieger

University of Glasgow

An analysis of Moore's paradox is given in doxastic logic. Logics arising from formalizations of various introspective principles are compared; one logic, K5c, emerges as privileged in the sense that it is the weakest to avoid Moorean belief. Moreover it has other attractive properties, one of which is that it can be justified solely in terms of avoiding false belief. Introspection is therefore revealed as less relevant to the Moorean problem than first appears.

Keywords Moore's paradox; doxastic logic; introspection; formal epistemology; logic of belief

DOI:10.1002/tht3.181

'The sun is shining, but I don't believe that it is.' As G.E. Moore famously noted, there's something wrong with believing this.[1] It has, however, proved surprisingly difficult to say exactly *what* is wrong.[2] The proposition has a quality akin to inconsistency; yet, clearly, it could be true.

The problem lends itself, I shall argue, to an illuminating formal analysis. Apart from the pioneering work of Hintikka (1962), not a great deal has been done in this area. Sorensen (1988, p. 23) writes ' interest in doxastic logic has dwindled, so much so that it is not feasible to erect a convincing analysis [of Moore's paradox] on this kind of foundation.'[3] But I shall argue that, contrary to Sorensen's view, doxastic logic *is* a fruitful way to approach the problem raised by Moore. Below I investigate the logics obtained by formalizing various introspection principles. A particular system of logic, not previously studied, which I call K5c, will emerge as key. Moreover, this logic can be justified, on a plausible view of what is wrong with Moorean beliefs, in terms solely of the avoidance of false belief. Since this can be done without appealing to introspection, this last notion turns out to be less relevant to the Moorean problem than first appears.

1 Preliminaries

The language of the logics to be considered is a propositional language, together with an operator B. The intended interpretation of Bp is that p is believed by some (fixed) rational agent. Sometimes we will write $\Box$ instead of B when this is convenient—typically when considering the logic in the context of modal logics in general. It is sometimes useful to use the dual operator $\Diamond$; the dual of B may be written C ('doxastic possibility').

Correspondence to: E-mail: Adam.Rieger@glasgow.ac.uk

We assume that the logic includes as theorems all classical propositional tautologies. We also assume the axiom

(K) $$B(p \to q) \to (Bp \to Bq)$$

which states that belief is closed under believed material implication.

A further basic assumption is the rule of necessitation

(Nec) From $\vdash p$ derive $\vdash Bp$,

which here can be thought of as formalizing the principle that the agent believes all the (doxastic) logical truths.

From (Nec) and (K) it follows that belief is also closed under (formal) entailment: for if Bp and $p \vdash q$, then $\vdash p \to q$, whence $\vdash B(p \to q)$ by (Nec), and hence Bq by (K).

The logics considered will therefore all be extensions of the standard normal modal system K, and can be handled by standard Kripke models. There is already a significant degree of idealization encoded in these axioms; for example, they entail that the agent believes arbitrarily long propositional tautologies. Worries about the finite capacities of the human mind, however, seem misplaced. The project is to analyze in what respect someone with a Moorean belief is irrational, and the approach is through constructing a theory of rationality, thought of as the principles obeyed by an ideally rational agent. A human disbelieving a long tautology is committing a sin against rationality, a pardonable one but a sin nonetheless.

Another natural axiom for our doxastic logic expresses that the agent does not believe any contradictions:

$$Bp \to \neg B \neg p$$

that is,

$$Bp \to Cp$$

or in more familiar modal terms

(D) $$\Box p \to \Diamond p.$$

The logics we consider will, in fact, all be extensions of KD.[4]

As a further preliminary we make the now standard distinction (see, e.g., Green and Williams 2007, p. 5) between the *commissive* Moore paradox

The sun is shining and I believe that it is not shining

and the *omissive* Moore paradox

The sun is shining and it's not the case that I believe it is shining.

In the formalism these become take the form of the schemas

(CMP) $$a \wedge B \neg a$$

and

(OMP) $$a \wedge \neg Ba.$$

2 Introspection

It is intuitively apparent that an agent with sufficient introspective powers is immune from Moore's paradox.[5] Let us explore this more carefully.

It would be a mistake to expect even an ideal agent to obey the axiom

(T) $$Bp \to p;$$

while this axiom is standard in many applications of modal logic, here it would require that the agent is infallible, that is, they have no false beliefs. This is clearly too strong; even an ideally rational agent may, for example, be the victim of misleading evidence. (Still less would it be appropriate to require the converse axiom $p \to Bp$, which is the belief version of omniscience.)

There are some spheres, however, in which it might be thought an ideal agent should not err. In particular, such an agent might be thought to have special access to their own beliefs, in such a way that they believe something if, and only if, they actually do believe it. This suggests the axioms

(4) $$Bp \to BBp$$

and

(4c) $$BBp \to Bp.$$

(4) is sometimes known as *positive introspection*; it is of course a familiar axiom in modal logic. The converse (4c) is less familiar as it is a consequence, indeed an instance, of T; it might be called *positive belief infallibility*.[6]

There are introspective powers one could desire of an agent which are not captured by these axioms. For example, an agent can obey (4), yet *fail* to have a belief without realizing that they do so. This suggests the axiom of *negative introspection*

(5) $$\neg Bp \to B\neg Bp.$$

There is also its converse, *negative belief infallibility*

(5c) $$B\neg Bp \to \neg Bp.$$

In box-diamond form (5) is $\neg\Box p \to \Box\neg\Box p$ which on replacing p by $\neg p$ takes the familiar form

$$\Diamond p \to \Box\Diamond p.$$

We shall have more to say about (5c) below.[7]

We can show that an agent[8] who is positively introspective and infallible — that is, who satisfies (4) and (4c) — will be immune from Moore paradoxical beliefs.

For the omissive case, suppose (for reductio)

$$B\left(p \wedge \neg Bp\right).$$

Then

$$Bp \wedge B\neg Bp$$

by closure under entailment. By (4)

$$Bp \to BBp;$$

hence

$$BBp \wedge B\neg Bp$$

which violates (D).

For the commissive case: suppose

$$B\left(p \wedge B\neg p\right).$$

Then

$$Bp \wedge BB\neg p.$$

But by (4c),

$$BB\neg p \to B\neg p.$$

Thus

$$Bp \wedge B\neg p$$

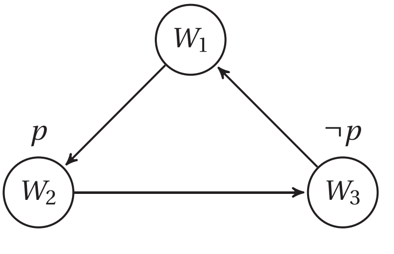
once again violating (D).

Can this simple observation be regarded as a solution of Moore's paradox? That is, should we simply make the diagnosis that what is wrong with a Moore-paradoxical believer is that they fail to be sufficiently introspective or are fallible about their own beliefs? This does not seem entirely satisfactory. For one thing, having a Moorean belief seems to be a failure of *rationality*. But it is unclear why a weakness in introspective powers should be irrational. A deficiency in this area seems, on the face of it, to be akin to having (say) bad eyesight rather than to believing a contradiction.

In addition, the proofs above leave it open that some weaker principles might suffice to avoid Moore, perhaps ones with a better claim to be rationality principles. That is what we shall now investigate.

3 Some systems of doxastic logic

How weak can a doxastic logic be and still guarantee immunity from Moore's paradoxes? We note first that KD is too weak; we can construct a KD-model in which both an omissive and a commissive Moorean belief is held. Consider the model below. In W_2 we have $p, \neg\Box p, \Box\neg p$ all true, and hence in W_1, we have that both $\Box(p \wedge \neg\Box p)$ (omissive Moorean belief) and $\Box(p \wedge \Box\neg p)$ (commissive Moorean belief) are true. Note that since R is serial this is a KD-model.



The considerations in the previous section show, in effect, that adding the axioms (4) and (4c) to KD—to form what Chellas (1980), p. 142 calls K4!—suffices to prevent the paradox. But might a weaker logic be enough?

A convenient way to approach this problem is as follows. One with an commissive Moorean paradoxical belief satisfies

$$B(a \wedge B\neg a)$$

and hence

$$Ba \wedge BB\neg a$$

for some *a*. We can simply outlaw such beliefs by adopting an axiom (schema)—call it 'no commissive Moore' or (Ncm) for short—of the form

$$Bp \to \neg BB\neg p.$$

With a little manipulation of negations this can be written in box-diamond form as

(Ncm) $\quad \Box p \to \Diamond\Diamond p.$

It is immediately apparent that the weakest strengthening of KD which is immune from commissive Moore will be the logic KDNcm obtained by simply adding this as an axiom schema; and in general a (KD) logic will avoid commissive Moore iff it contains KDNcm.

Similarly omissive Moore can be avoided by adding the axiom schema

$$Bp \to \neg B\neg Bp$$

or in box-diamond terms

(Nom) $\quad \Box p \to \Diamond\Box p.$

What of these axioms? As far as I know, no-one has previously given a specific name to the axiom I have labeled (Ncm).[9] (Nom), on the other hand, can be re-written in dual form

$$\Box\Diamond p \to \Diamond p$$

and is therefore equivalent to (5c) above, 'negative belief infallibility.' The key to avoiding the omissive Moore paradox, therefore, is not to be 'peculiar' in the sense of Smullyan.

We now have five principles, or axioms, on the table: (4), (4c), (5), (5c) and (Ncm). By adding one or a combination of them to KD, we obtain a number of alternative doxastic logics, which we can compare in strength. When doing so, it is convenient to know, for each axiom, a constraint on the accessibility relation *R* of a standard Kripke model such that the axiom is valid is the class of models satisfying that constraint. It is familiar that (4) corresponds to transitivity and (5) to Euclideanness (and (D) to serialness: $\forall x \exists y Rxy$). There are standard methods allowing one to determine a constraint on *R* for a large class of candidate axioms (see Lemmon 1977, p. 52, 67). Using these we obtain the following:

Axiom		Constraint on R
4	$\Box p \rightarrow \Box\Box p$	$\forall u \forall v \forall w((Ruv \wedge Rvw) \rightarrow Ruw)$ (transitivity)
4c	$\Box\Box p \rightarrow \Box p$	$\forall u \forall v(Ruv \rightarrow \exists w(Ruw \wedge Rwv))$ (density)
5	$\Diamond p \rightarrow \Box\Diamond p$	$\forall u \forall v \forall w((Ruv \wedge Ruw) \rightarrow Rvw)$ (Euclideanness)
5c = Nom	$\Box\Diamond p \rightarrow \Diamond p$	$\forall u \exists v(Ruv \wedge \forall w(Rvw \rightarrow Ruw))$
Ncm	$\Box p \rightarrow \Diamond\Diamond p$	$\forall u \exists v(Ruv \wedge \exists w(Ruw \wedge Rwv))$

We are now in a position to investigate the relationships between these axioms and the various logics obtained by adding some subset of them to KD.

For example, we can prove that (5) (with (D)) entails (4c). For an instance of (D) is $\Box\Box p \rightarrow \Diamond\Box p$. (5) (contraposed from the version above) is $\Diamond\Box p \rightarrow \Box p$; putting these together yields (4c): $\Box\Box p \rightarrow \Box p$.

Each of (4c) and (5c) entails (D) (in K). This is unsurprising given the constraints on R in the table; demonstrating it proof-theoretically can be done as follows. We will take the case of (5c): intuitively the idea is that if (D) fails, everything is necessary, while from (5c) one can derive that something is possible (and hence its negation is not necessary). Suppose for *reductio* some instance of (D) fails: then for some a, $\Box a$ and $\Box\neg a$. Hence

$$\Box(a \wedge \neg a).$$

Since

$$\vdash (a \wedge \neg a) \rightarrow \neg\Box(p \rightarrow p)$$

(where $\vdash$ denotes derivability in K), we have

$$\vdash \Box((a \wedge \neg a) \rightarrow \neg\Box(p \rightarrow p))$$

by (Nec), whence by (K)

$$\vdash \Box(a \wedge \neg a) \rightarrow \Box\neg\Box(p \rightarrow p).$$

Hence on our original assumption,

$$\Box\neg\Box(p \rightarrow p),$$

that is

$$(\star) \qquad \neg\Diamond\Box(p \rightarrow p).$$

But also

$$\vdash (p \rightarrow p),$$

so by (Nec)

$$\vdash \Box(p \rightarrow p).$$

An instance of (5c) (contraposed) is

$$\Box(p \rightarrow p) \rightarrow \Diamond\Box(p \rightarrow p);$$

hence

$$\Diamond\Box(p \to p)$$

by modus ponens, contradicting ($\star$). An exactly similar proof can be used to derive (D) from (4c).

Countermodels to demonstrate independence are fairly easy to construct, using the constraints on R as a guide. As illustration, to demonstrate that KD4 does not contain K4c we need a serial model which is transitive but not dense. One isomorphic to *less than* on the natural numbers will do: on letting p be true at ‘world 1’ of the model and false at worlds 2 onwards, we obtain $\Box\Box p$ true and $\Box p$ false at world 0, as required.

In a similar way, one can construct a model satisfying the constraint for (5c) but not full transitivity to demonstrate that (5c) does not entail (4).

I will not give full details of the proof, but it turns out that there are 10 distinct doxastic logics that can be obtained by adding a subset of the 5 axioms to KD, with the following diagram illustrating relative strength:

$$\begin{tikzcd}
 & \mathrm{K5!} \arrow[r] \arrow[dr] & \mathrm{KD5} \arrow[r] & \mathrm{K4c} \arrow[dr] & & \\
\mathrm{KD45} \arrow[ur] \arrow[dr] & & \mathrm{K4c5c} \arrow[ur] \arrow[dr] & & \mathrm{KDNcm} \arrow[r] & \mathrm{KD} \\
 & \mathrm{K4!} \arrow[r] \arrow[ur] & \mathrm{KD4} \arrow[r] & \mathrm{K5c} \arrow[ur] & &
\end{tikzcd}$$

Here we write, as above, (4!) for the conjunction of (4) and (4c) (and similarly with (5!)). Since, as noted above, each of (4c) and (5c) entail (D), logics are labeled as K4c rather than KD4c, etc, since the D would be redundant.

Once one has the diagram in place, a number of observations are possible.

1. K5c (that is, KDNom) contains KDNcm; thus avoiding an omissive Moore paradox guarantees avoiding a commissive one. This, indeed, should come as no surprise. A (KD) agent who has a commissive Moorean belief satisfies $B(p \wedge B\neg p)$, hence $BB\neg p$. Since the agent not only satisfies the D-axiom but believes that they do so (by (Nec)), we have $B(B\neg p \to \neg Bp)$ and hence $BB\neg p \to B\neg Bp$, and finally $B\neg Bp$ by modus ponens. Thus a commissive Moore gives rise to a corresponding omissive one, which is the contraposition of the result stated above.
2. The converse, however, does not hold; one may avoid all commissive Moorean paradoxes yet be liable to omissive ones.
3. The informal argument in the previous section showed that K4! sufficed to avoid the paradoxes; we now see that this can be weakened to KD4.[10]
4. As a particular application to a position that has appeared in the literature: Heal (1994, p. 21) attempted to solve Moore’s paradox using a version of the positive infallibility principle $BBp \to Bp$ (4c). Green and Williams (2007, p. 17) point out that this solves only the commissive paradox, leaving the omissive paradox unsolved.

This can be read off immediately from the diagram; the logic K4c contains the logic KDNcm, but not the logic KDNom = K5c.

More generally, we see that any of the logics KD45, K5!, K4!, K4c5c, KD4 and K5c suffice to avoid both Moore paradoxes, while KD5, K4c, KDNcm and KD do not. The fact that adding positive introspection—axiom (4)—to KD avoids the paradoxes might suggest that this is the appropriate rationality principle to adopt. However, the fact that a strictly weaker logic, K5c, suffices to solve the paradoxes throws this logic into the limelight. Somewhat surprisingly, it is negative infallibility—that if you believe you don't believe something, you really don't believe it—that emerges as the key rationality principle here.

4 K5c and avoiding false belief

The problem posed by Moore's paradox, as outlined above, is to explain what is wrong with believing a Moorean proposition, in the light of the fact that these propositions are consistent. I have so far held back from endorsing a particular solution to this, and the project of identifying K5c as the weakest logic that avoids Moorean belief is, I think, worthwhile whatever one's view. However, a plausible line of thought on this turns out to give further support to K5c.

The following idea has appeared several times in the literature and seems to me to be, at the very least, along the right lines (see, e.g., Deutscher 1967, p. 184; Williams 1994, p. 165). The suggestion is that, although Moorean propositions are consistent (can be true), they *cannot be true if believed*. It is quite easy to see that this applies to propositions of both the forms (OMP) and (CMP). For the omissive case, it is clear that $B(p \wedge \neg Bp))$, together with closure, yields Bp which is inconsistent with the believed proposition; and as observed above, one with a (CMP) belief must in any case also have the corresponding (OMP) belief. It is a simple matter, if desired, to formalize this in KD.

This seems to give an attractive answer as to why we should regard avoiding Moorean beliefs as a principle of *rationality*. For one who fails to avoid them is doomed as a matter of logic (given other weak assumptions about their rationality as encoded in KD) to possess at least one false belief.

In as far as K5c is (just) strong enough to guarantee their avoidance, therefore, this supports the logic as correctly embodying rationality principles. One might worry, however, whether (OMP) and (CMP) exhaust the possibilities here. Might there not be some other schema, distinct from both of these, which nevertheless has a similar effect, of producing inconsistency when combined with belief? And perhaps K5c is powerless to prevent such beliefs?

Happily we can prove that, were such alternative schemas to be proposed, K5c is already strong enough to avoid them.

Let us say a is *Moorean* if (according to the minimal logic KD) it is consistent, but cannot be true if believed. Formally, a is Moorean if

i. $\nvdash_{KD} \neg a$
ii. $\vdash_{KD} \neg (Ba \wedge a)$.

Then we have the following

Theorem 1 *Suppose* a *is Moorean. A (KD)-logic L satisfies* $\vdash_L \neg \mathrm{Ba}$ *iff L contains K5c.*

Proof. We first show that if L contains K5c and a is Moorean, then $\vdash_{\mathrm{L}} \neg \mathrm{Ba}$. For convenience, let us write B as $\Box$.

Suppose for *reductio* that $\Box a$. Since a is Moorean,

$$\vdash_{\mathrm{L}} \neg(\Box a \wedge a),$$

hence

$$\vdash_{\mathrm{L}} a \rightarrow \neg\Box a.$$

Hence, by (Nec),

$$\vdash_{\mathrm{L}} \Box(a \rightarrow \neg\Box a)$$

whence by (K),

$$\vdash_{\mathrm{L}} \Box a \rightarrow \Box\neg\Box a.$$

By modus ponens, we can therefore derive from our assumption $\Box a$

$$\Box\neg\Box a,$$

that is,

$$\Box\Diamond\neg a.$$

Since L contains K5c,

$$\vdash_{\mathrm{L}} \Box\Diamond\neg a \rightarrow \Diamond\neg a$$

since the formula is just an instance of the (5c) axiom. Hence

$$\Diamond\neg a$$

that is,

$$\neg\Box a$$

and hence

$$\Box a \wedge \neg\Box a.$$

We have thus reduced the assumption $\Box a$ to a contradiction, and can conclude by *reductio*

$$\vdash_{\mathrm{L}} \neg\Box a$$

as required.

For the opposite direction, suppose L does not contain (5c). Then there will be an L-model in which there is an omissive Moorean belief, that is, a model in which $B(p \wedge \neg Bp)$ holds. As proved above, OMP is Moorean. $\Box$

This theorem is really the central result of this paper: informally, K5c is exactly strong enough to outlaw beliefs which, though consistent, cannot be true if believed. How should we regard the axiom (5c)? It was introduced above as one of a family of introspection principles, but as I mentioned earlier, it seems dubious to regard these as rationality principles. The approach in this section suggests an alternative way to look at (5c): it is simply a self-standing principle of rationality, justified not by the introspective powers of our ideal agent but by the imperative to avoid false belief. Introspection has, after all, nothing to do with avoiding Moorean belief.

The line I am taking here is foreshadowed by Hintikka (1962). Hintikka's doxastic logic includes the (4) axiom, but he explicitly repudiates (pp. 53–7) using introspection to justify his rationality principles. His attempted justification, however, is problematic. The key axiom which yields the proof of (4) is his principle (A.CBB*) (p. 24), a special case of which is: if $a, Ba \vDash b$, then $Ba \vDash Bb$.[11] (4) follows easily by putting $b = Ba$.

Can (A.CBB*) be justified without appealing to introspection? Hintikka attempts (p. 25) to argue for it by citing an omissive Moore paradox case, and stating that it seems to him that to hold such beliefs is 'clearly inconsistent.' But even if we grant him this appeal to brute intuition, as is by now clear this will not motivate (4) (or (A.CBB*)) but only the weaker (5c).

5 Accuracy

In this section I show that K5c has yet further desirable properties. Milne (1993), pp. 501–2 introduces the following properties of a doxastic logic:

i. *Positive accuracy*: if $\vdash Ba$ then $\vdash a$. (Note we are guaranteed the converse of this since we're assuming the rule of necessitation.)
ii. *Negative accuracy*: if $\vdash \neg Ba$ then $\vdash \neg a$.

Intuitively, positive accuracy means that the only beliefs held as a matter of logic are ones which are already (doxastic) logic truths; a violation, as Milne puts it, 'commits a rational individual to believing propositions that are, in the logic's own terms, at best accidental' (p. 501). Similarly, a violation of negative accuracy means that the logic commits the agent to *disbelieving* a proposition that is consistent (by the lights of that very logic). Milne describes each of positive and negative accuracy as an 'adequacy condition,' noting (p. 517) that KD is both positively and negatively accurate but (p. 520) that KD45 is neither, and concludes that KD45 is 'much too strong.'

In view of the approach of this paper, and in particular the last section, it should come as no surprise that I do not regard negative accuracy as a virtue for a doxastic logic. The lesson of Moore's paradox, indeed, is exactly that it can sometimes be irrational to believe consistent propositions. The negative accuracy of KD precisely shows that it is too weak to avoid Moore-paradoxical belief. In fact all the Moore-avoiding logics will fail negative accuracy.[12]

Positive accuracy, however, does seem to be desirable, and it is unclear what could justify a violation. It is interesting to examine the proof that KD45 fails to have positive accuracy. The reason is that the formula

(U) $$B(Bp \to p)$$

is theorem of KD45 but of course the T axiom $Bp \to p$ is not. Smullyan (1988, p. 78) calls an agent who satisfies (U) *conceited*. Here as elsewhere, conceit is not an attractive feature:

a conceited agent believes herself to have no false beliefs, but the kind of ideal agent we are modeling is fallible, and optimally should be aware of this.

Somewhat surprisingly perhaps, conceit is a consequence of negative introspection: the (U) axiom follows from (5) alone. For a violation of (U) means that, for some a, $\diamond(\Box a \wedge \neg a)$; hence $\diamond \Box a \wedge \diamond \neg a$, that is, $\diamond \Box a \wedge \neg \Box a$, violating (5). In fact it is quite easy to show that (U) is strictly intermediate in strength between (5) and (4c) (see Chellas 1980, pp. 92, 140 – 1,167). It follows that the same proof can show any (KD) logic containing (5) but not (T) will fail to be positively accurate: thus KD5, K5! and KD45 are all ruled out as too strong on this basis.

Ideally, then, we would like our logic to be positively but not negatively accurate. Happily, K5c meets these conditions.

Theorem 2 *K5c is positively but not negatively accurate.*

Proof. First, K5c is not negatively accurate, as above. For if $a = (p \wedge \neg Bp)$, then $\vdash_{K5c} \neg Ba$ but $\nvdash \neg a$ (for $\neg a$ is equivalent to $p \rightarrow Bp$, which is not a theorem even of KD45, indeed not even of S5).

For the positive accuracy, we use the method of 'safe extensions' (see Lemmon 1977, p. 27 or Chellas 1980, p. 99). We will prove that if $\vDash_{K5c} \Box a$ then $\vDash_{K5c} a$; the result then follows by the standard soundness and completeness theorems.

The proof proceeds by contraposition. Suppose that it is not the case that $\vDash_{K5c} a$; thus there is a K5c model in which, at some world W, a is false. We use this to construct a model in which, at some world, $\Box a$ is also false. Enlarge the model by adding a new world W' and extending the accessibility relation so that it holds (i) between W' and W, and (ii) between W' and any world W_i such that RWW_i in the original model (and is otherwise exactly as before). By the 'safe extension theorem' Lemmon (1977, p. 27), a is still false at W in the new model. And $\Box a$ is therefore false at W'.

It remains to check that the new model is still a K5c model. We can show easily that R satisfies the constraint given above, $\forall u \exists v(Ruv \wedge \forall w(Rvw \rightarrow Ruw))$. For $u = W'$, taking v to be W will do; the construction of R ensures the condition holds. And if u is any other world W_i, the condition holds in the new model exactly as it did in the old one, since the accessibility relation between worlds other than W' is exactly as it was in the old model.

Hence, as required, not-$\vDash_{K5c} a$ implies not-$\vDash_{K5c} \Box a$; contraposing we have $\vDash_{K5c} \Box a \Rightarrow \vDash_{K5c} a$ as required. □

We can give a variant of this proof to show that some other logics, notably KD4, are on a par with K5c as far as accuracy is concerned; but as we saw earlier, these are strictly stronger than is needed to avoid Moorean propositions.

6 Conclusion

The last few sections suggest that the previously neglected logic K5c has quite a lot going for it: it is exactly strong enough to avoid not only the standard Moorean paradoxes, both omissive and commissive versions, but also any other propositions that share with these the feature that they cannot be true if believed. Thus, if avoiding Moore-paradoxical beliefs is a matter of rationality, K5c embodies exactly the rationality principles required. And it has exactly the accuracy properties that we should hope for from a doxastic logic.

Is K5c the One True Doxastic Logic? This may be to claim too much. One can reasonably worry both that it is too strong and that it is too weak. On the first, even the K-axiom may be doubted, even for an idealized agent, for reasons concerning the preface paradox and such matters: one might have good evidence for each of a large number of propositions, yet be reluctant to believe their conjunction, on the grounds that one is probably wrong about at least one of them.[13] On the other side, there are Moore-like phenomena, not falling under the definition given above, which K5c is too weak to avoid. For example, Sorensen (2000) discusses some iterated Moore-like propositions, such as 'It's raining, but I don't believe that I believe it's raining.' This is not Moorean in the sense defined above since it can be true even if believed, yet intuitively believing it seems to carry an irrationality akin to that of the original Moore.

An extension of the analysis given in this paper is an obvious avenue to explore for both these problems, but that will have to wait another day. I hope, though, that my investigations here might convince some that—contra Sorensen's pessimistic view cited in the introduction—doxastic logic does have a useful role to play.[14]

Notes

1 Or asserting it, but I shall not consider that here.

2 See, for example, the introduction to Green and Williams (2007) for a survey of attempts.

3 Of course, a great deal of work has been done on the logic of belief change, and some of that has touched on Moore's paradox: see for example Van Benthem (2004) and Segerberg (2006). But the focus of these papers is on the problems that Moore paradoxes can create when updating in *dynamic* logics. As far as I know, the simple approach suggested in the present paper has not previously been considered.

4 Milne (1993), p. 517 suggests KD as an 'appealing' minimal doxastic logic.

5 For a treatment of Moore which puts introspection centre-stage, see for example Shoemaker (1995).

6 The entertaining Smullyan (1988) (and the earlier paper Smullyan 1986) attaches memorable labels to some of these properties: agents obeying (4) and (4c) are called *normal* and *stable* respectively.

7 Smullyan (1988, p. 81) calls one who violates it *peculiar*.

8 As stated above, we suppose that the agent satisfies the axioms K and D.

9 It is, though, an instance of a well-studied *schema*: in the terminology of Lemmon (1977, p. 51) it is an instance of the schema G' with $m = p = 0, n = 1, q = 2$.

10 Indeed, Hintikka (1962) uses a logic equivalent to KD4 in giving his solution to Moore.

11 Hintikka puts his axioms in terms of consistency, but it is easier to see what is going on if one contraposes and puts them in terms of entailment.

12 Milne has told me (in conversation) that he no longer regards negative accuracy as an adequacy condition.

13 As an anonymous referee observes, the preface and lottery paradoxes also given reason to doubt the principle that one should not hold a set of beliefs if at least one is guaranteed to be false, and hence threaten to undermine the justification of K5c given in Section 4.

14 My thanks to David Bain, Bob Hale, Stephan Leuenberger, Peter Milne, Martin Smith and two anonymous referees, and apologies to anyone I've left off this list by mistake. I began thinking about the issues in this paper while I held an award (AH/H006230/1) from the AHRC under their Research Leave scheme in the Autumn of 2010.

References

Chellas, Brian F. *Modal Logic: An Introduction*. Cambridge; New York: Cambridge University Press, 1980.

Deutscher, Max. "Bonney on Saying and Disbelieving." *Analysis* **270**.6 (1967): 184 – 6.

Green, Mitchell and John N. Williams, (eds). *Moore's Paradox*. Oxford: Clarendon Press, 2007.

Heal, Jane. "Moore's Paradox: A Wittgensteinian Approach." *Mind* **1030**.409 (1994): 5 – 24.

Hintikka, Jaakko. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, New York: Cornell University Press, 1962.

Lemmon, E. J. *An Introduction to Modal Logic: The Lemmon Notes*. Monograph Series — American Philosophical Quarterly; no. 11. Oxford: B. Blackwell, 1977.

Milne, P. "Minimal Doxastic Logic: Probabilistic and Other Completeness Theorems." *Notre Dame Journal of Formal Logic* **4** (1993): 499 – 526.

Segerberg, Krister. "Moore Problems in Full Dynamic Doxastic Logic." *Poznan Studies in the Philosophy of the Sciences and the Humanities* **910**.1 (2006): 95 – 110.

Shoemaker, Sydney. "Moore's Paradox and Self-knowledge." *Philosophical Studies* **770**.2 (1995): 211 – 28.

Smullyan, Raymond M. "Logicians Who Reason About Themselves," in *Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning About Knowledge*, Monterey, California, 1986, 341 – 52. Morgan Kaufmann Publishers Inc.

Smullyan, Raymond M. *Forever Undecided: A Puzzle Guide to Gödel*. Oxford; New York: Oxford University Press, 1988.

Sorensen, Roy A. *Blindspots*. Clarendon Library of Logic and Philosophy. Oxford; New York: Clarendon Press; Oxford University Press, 1988.

Sorensen, Roy. "Moore's Problem with Iterated Belief." *The Philosophical Quarterly* **500**.198 (2000): 28 – 43.

Van Benthem, Johan. "What One May Come to Know." *Analysis* **640**.282 (2004): 95 – 105.

Williams, J. N. "Moorean Absurdity and the Intentional 'Structure' of Assertion." *Analysis* **540**.3 (1994): 160 – 6.