Logic and Logical Philosophy
Volume 15 (2006), 325–354
DOI: 10.12775/LLP.2006.019

Mark Jago

HINTIKKA AND CRESSWELL ON LOGICAL OMNISCIENCE*

Abstract. I discuss three ways of responding to the logical omniscience problems faced by traditional 'possible worlds' epistemic logics. Two of these responses were put forward by Hintikka and the third by Cresswell; all three have been influential in the literature on epistemic logic. I show that both of Hintikka's responses fail and present some problems for Cresswell's. Although Cresswell's approach can be amended to avoid certain unpalatable consequences, the resulting formal framework collapses to a sentential model of knowledge, which defenders of the 'possible worlds' approach are frequently critical of.

Keywords: Logical Omniscience, Epistemic Logic, Nonclassical Logics

1. Introduction

In this paper, I will discuss Hintikka's influential approach to epistemic logic and what has been perceived as its major flaw, the logical omniscience problems. I will focus on three responses to this family of problems, the first two originating with Hintikka himself, the third put forward by Max Cresswell. Hintikka develops what has become the predominant contemporary approach to epistemic logic in [12]. In a later paper, he summarizes his approach as holding that "in order to speak of what a certain person a knows and does not know, we have to assume a class ('space') of possibilities" [16,

*I would like to thank an anonymous referee for *Logic and Logical Philosophy*, who suggested a number of improvements to Section 2.

Received December 4, 2006; *Revised* February 1, 2007

ISSN: 1425-3305

p. 19]. These possibilities, which Hintikka calls 'model sets' in [12] and 'scenarios' in [16], are logically structured so that, for example, a conjunction holds in one such scenario only if both of its conjuncts hold there.[1] The idea is roughly as follows. If I do not know whether it is raining outside then, for all I know, my situation could be one in which it is currently raining outside or one in which it is not. Both situations are *epistemically possible* for me. When I see the rain outside my window, I gain knowledge and cease to entertain the situation in which it is not raining as a possibility. A state of knowledge is a restriction on the situations that are epistemically possible for the agent in question.

In Hintikka's framework, models of knowledge and belief are relational structures. The domain of the model is a set of scenarios S. In contemporary accounts, the sentences which hold in a scenario are closed under some logical consequence relation (say, that of classical propositional or first-order logic) and scenarios are frequently termed 'possible worlds'. The epistemic accessibility relation R ($\subseteq S \times S$) holds between scenarios in the domain of the model; a model of the knowledge of multiple agents will have a distinct accessibility relation R_a for each agent a. The satisfaction relation $\Vdash$ holding between a scenario s and a sentence (in a model $\mathcal{M}$, which I treat as implicit) is defined recursively in the usual way, with sentences '$\mathsf{K}_a\phi$' ('agent a knows that ϕ') dealt with as follows:

$$s \Vdash \mathsf{K}_a\phi \text{ iff, for all } s' \in S,\ R_a ss' \text{ only if } s' \Vdash \phi$$

Hintikka's framework contains some uncomfortable implications. Given the 'possible scenarios' (or 'possible worlds') semantics presented above, any sentence true in all scenarios in the model must be known by all agents in all of these scenarios. If what holds in a scenario is closed under logical consequence, then all logical consequences of what an agent knows must be known by that agent. As a particular instance of this phenomenon, all tautologies are known by all agents, regardless of the length of the tautology or the cognitive capacity (reasoning abilities, available memory, time in which to reason) of the agent. The term 'logical omniscience' is used to refer to a class of closure conditions on what an agent is modelled as knowing (see e.g. [33, p. 140]), including:

[1] Hintikka complains that "philosophers typically call [scenarios] possible worlds. This usage is a symptom of intellectual megalomania" [16, p. 19] but appeared to approve of the terminology 'possible worlds' in [15] (as its title 'Impossible Possible Worlds Vindicated' suggests).

1. *Knowledge of all valid sentences*: If ϕ is valid, then any agent knows that ϕ. This is often called the problem of *irrelevant knowledge*.
2. *Closure under logical entailment*: If ϕ entails ψ, then any agent who knows that ϕ must also know that ψ.
3. *Closure under logical equivalence*: If ϕ is logically equivalent to ψ, then any agent who knows that ϕ must also know that ψ.
4. *Closure under valid implication*: If $\phi \rightarrow \psi$ is valid, then any agent who knows that ϕ must also know that ψ.
5. *Closure under known implication*: Any agent who knows both that ϕ and that $\phi \rightarrow \psi$ must also know that ψ.
6. *Closure under conjunction*: Any agent who knows that $\phi \wedge \psi$ must also know that ϕ and that ψ.
7. *Closure under disjunction*: Any agent who knows that ϕ must also know that $\phi \vee \psi$, for any ψ.

Of these, closure under entailment is the strongest and entails all the other conditions.[2] (1) and (4) entail (5); and (2) is equivalent to (4), so long as '$\phi \rightarrow \psi$' is valid iff ϕ entails ψ.

In what follows, I will consider three responses to the logical omniscience problems. The first is Hintikka's reaction in [12]. After presenting his epistemic logic, Hintikka notes the logical omniscience problem (although not under that name) and gives a response. This response is important, despite its flaws, in that it seeks to justify the possible scenarios semantics in light of the logical omniscience problems without changing the logical details. Justification of an idealized concept of knowledge has been influential in the computer science literature, which tends to abstract from the computational restrictions of real agents. There are applications of Hintikka's epistemic logic in which such assumptions are harmless [10, p. 41]. Conceptually, we can think of Hintikka's logic as modelling the knowledge of idealized, superhuman reasoners or the knowledge that is *implicit* in what real agents know; see [31, 32] for a discussion. As Vardi notes, even if this conceptual approach can be justified, we remain in want of a logic of explicit knowledge [34, p. 294].

The second and third responses to logical omniscience that I consider, due to Hintikka [15] and Cresswell [6], take a different approach. They maintain

[2] Provided, that is, that the '$\vee$' introduction rule and the elimination rules for '$\rightarrow$' and '$\wedge$' are sound.

Hintikka's original analysis of knowledge in terms of possible situations but modify the logic of certain situations. This approach has also been influential in the literature on epistemic logic and logical omniscience, including [22, 20, 34, 8, 10, 35, 21] and, more recently, [27]. I discuss Hintikka's approaches in sections 2 and 3 and Cresswell's in sections 4 and 5. Finally, I discuss what I think is required to rescue Cresswell's approach in section 6, using ideas from [27] and [18].

2. Defensibility

One response to the logical omniscience problems is that the concept modelled in the logic differs from our everyday concept of knowledge (but that the two are sufficiently interrelated to allow us to call the resulting logic 'epistemic logic'). This is Hintikka's strategy in [12, §2.5], where he holds that the logical omniscience problem "does not go to show that our rules [governing the 'K' operator] are incorrect" [12, p. 31]. What it does show is that the statements:

It is possible, for all I know, that $\neg p$

I know that p_1

I know that p_2

$\vdots$

I know that p_n.

are consistent with one another, even when $p_1, \ldots, p_n$ entail p. Hintikka takes this to show that "the notion which [the rules of epistemic logic] define is unlike inconsistency ... and should be carefully distinguished from it" [12, p. 31].

Hintikka's inference rules give us a perfectly standard notion of consistency. What they do not give us is a perfectly standard notion of *knowledge*. Entailments from a set of sentences '$\mathsf{K}_a p_1$', ..., '$\mathsf{K}_a p_k$' to '$\mathsf{K}_a q$' should in "typical cases" be viewed as asserting that agent a is immune to certain kinds of criticism [12, p. 31]. If an agent can be shown "by means of some argument which he would be willing to accept" that q follows logically from what he says he knows, then it would be irrational of him and so *indefensible* to "persist in saying that he does not know whether q is the case" [12, p. 31].[3]

[3] The intuitive idea is that "he would have come to know that q all by himself if he had followed far enough the consequences of what he already knew" [12, p. 31]. As a special case, Hintikka proposes to call a sentence '$\mathsf{K}_a q$' *self-sustaining* when 'q' is valid.

Hintikka's proposal is thus to accept logical omniscience for an idealized notion of knowledge, cashed out in terms of indefensibility.[4] This is best viewed as an attempt to provide normative reasons for accepting the closure principles. To avoid the charge of circularity, Hintikka needs to cash out 'defensibility' without appealing to what entails what in his logical system. He holds [12, pp. 30–32] that:

(D1) '$\mathsf{K}_a p_1 \wedge \cdots \wedge \mathsf{K}_a p_n \wedge \neg \mathsf{K}_a q$' is indefensible for a iff there is an argument that a would accept that establishes that q, given 'p_1' to 'p_n' as premises.

As Chisholm has pointed out [4], 'indefensible' is a bizarre choice of terminology. Ordinarily, we would not say that it is the statement of a's lack of knowledge that is indefensible but rather "what it is that I am describing—namely, your neglect to draw all of the consequences of what you know, or your acceptance of something that is logically false" [4, p. 781]. Statements that Hintikka calls 'indefensible' may well be true, given that there are such cognitively bounded agents and so the terminology is "misleading" to say the least [4, p. 781].[5] But let us accept Hintikka's use of 'indefensible' for the time being as an analysis of entailments from '$\mathsf{K}_a p_1$' through '$\mathsf{K}_a p_n$' to '$\mathsf{K}_a q$' in the logical system. Then:

(D2) '$\mathsf{K}_a p_1$' through to '$\mathsf{K}_a p_n$' together entail '$\mathsf{K}_a q$' (in Hintikka's logic) iff '$\mathsf{K}_a p_1 \wedge \cdots \wedge \mathsf{K}_a p_n \wedge \neg \mathsf{K}_a q$' is indefensible for a.

From (D1) and (D3), we obtain:

(D3) '$\mathsf{K}_a p_1$' through to '$\mathsf{K}_a p_n$' together entail '$\mathsf{K}_a q$' (in Hintikka's logic) iff there is an argument that a would accept that establishes that q, given 'p_1' to 'p_n' as premises.

---

[4] He claims that the notion of indefensibility is "a notion important enough to deserve serious study" [12, p. 32]; yet he makes no comment as to whether the resulting notion does indeed provide us with an analysis of the way the verb 'to know' is actually used. If logical omniscience is taken to be a problem on empirical grounds, then one might find fault with Hintikka's notion of defensibility, as applied to knowledge. What empirical evidence is there, an objection might run, that there is any connection between defensibility and the way that 'to know' is actually used?

[5] Chisholm continues: "'Shocking', 'disappointing', or 'epistemically scandalous' might be less misleading" [4, p. 781]. But even this is not correct, as Hocutt points out: "if a does not know all the logical consequences of what he knows, and if I say so, it is still not my statement but a's stupidity which is "disappointing", "shocking", "scandalous", or whatever" [17, p. 444]. Stating that an agent lacks knowledge might, if correct, entail that the agent is not a perfect reasoner but that is clearly no defect of the statement itself.

This is false in both directions as indefensibility, so defined, is neither a necessary nor a sufficient condition for the entailments between ascriptions of knowledge in Hintikka's logic. It is not sufficient because just about any agent requires little or no argument to accept the sentence 'nothing that is red all over can, at the same time and in the same way, be green all over' as true. Call this sentence 'q'. '$\mathsf{K}_a q$' is not valid (i.e. not 'self-sustaining') in Hintikka's logic, hence the right-to-left direction of (D3) does not hold.[6]

It is not necessary, for there are sentences 'p_1' to 'p_n' and 'q' and agents a such that '$\mathsf{K}_a p_1$' to '$\mathsf{K}_a p_k$' entail '$\mathsf{K}_a q$' in Hintikka's logic, yet a would not accept any argument from 'p_1' through to 'p_n' as establishing 'q'. One example is when one or more of these sentences is so long that no physically realized agent has the cognitive capacity to parse it, let alone understand or accept it. Another is when a is a signed-up intuitionist or relevant logician and the argument requires intuitionistically or relevantly inadmissible reasoning.[7] To avoid these problems, the requirement that the argument in question must be accepted (or even be capable of being accepted) by the agent must go and there must be a universal standard of what counts as an argument. This modified definition is necessary but still not sufficient, as the colour exclusion example shows. The final amendment is to insist that the argument mentioned must be a wholly logical one, so as to exclude any argument whose conclusion is a conceptual but not a logical truth:

(D1′) '$\mathsf{K}_a p_1 \wedge \cdots \wedge \mathsf{K}_a p_n \wedge \neg\mathsf{K}_a q$' is indefensible for a iff there is a proof of 'q' from 'p_1' to 'p_n'.

which, together with (D2), gives us the true biconditional:

(D3′) '$\mathsf{K}_a p_1$' through to '$\mathsf{K}_a p_n$ together entail '$\mathsf{K}_a q$' (in Hintikka's logic) iff there is a proof of 'q' from 'p_1' to 'p_n'.

The problem is that (D3′) is utterly trivial. The purpose of this definition of indefensibility was to provide a notion that would both make the rules of Hintikka's logic (and in particular the necessitation and distribution rules) acceptable and explain why '$\mathsf{K}_a p$' should entail '$\mathsf{K}_a q$' whenever 'p' entails 'q'. (D1′) tells us that this is so because 'q' can be proved from 'p'. This amounts

[6] Of course we can add non-logical axioms to account for such examples but this cannot support (D3), precisely because such additional axioms are non-logical.

[7] Williamson [36] discusses a bizarre conspiracy theorist who denies that all foxes are foxes because (i) she believes that $\forall x(Fx \rightarrow Fx)$ holds only if there exist Fs and (ii) denies that any foxes exist. Such an agent would not accept the (trivial) argument that foxes are foxes, yet $\neg K_a \forall x(Fx \rightarrow Fx)$ is indefensible.

to no more than the claim that the proof theory in question is sound and complete with respect to classical (propositional or first-order) semantics, which of course it is. It in no way explains what needs to be explained. The circularity of Hintikka's discussion was noted first by Chisholm [4] and discussed by Hocutt [17], who notes:

> Indefensibility thus understood is clear, and so therefore is epistemic logic. What is not clear is why anyone would suppose epistemic logic, thus understood, to be, in any significant sense, epistemic. It is merely logic applied to the propositional contents of what happen to be knowledge claims. [17, p. 443]

Hintikka's analysis does not provide us with any reason to believe that his rules are applicable to our concept of knowledge if we did not already believe that they are to begin with.

3. Hintikka's Impossible Possible Worlds

By the time of his 1975 paper 'Impossible Possible Worlds Vindicated' [15], Hintikka's attitude appears to have shifted. He casts the problem of logical omniscience as follows (which I will put in terms of worlds, rather than situations, in accordance with Hintikka's terminology in [15]):

(1) 'a knows that ϕ' is true at w iff ϕ is true at every world epistemically accessible from w;

(2) There are a, 'ϕ', 'ψ' such that a knows that ϕ, 'ϕ' logically implies 'ψ' and yet a does not know that ψ;

(3) A sentence is logically true iff it is true at every possible world;

(4) Every epistemically possible world (and so every world epistemically accessible from any world) is logically possible.

(1–4) are clearly inconsistent. I will call this *Hintikka's problem*. Hintikka immediately argues that (2) is not the culprit [15, p. 476]: there really are such sentences, so related. The argument that (2) is false might run as follows. The conclusions of any valid argument are already contained in its premises; the process of logical deduction is not one that is capable of providing one with new information. Hence, the logical consequences of one's knowledge should be included in what one says one knows, such that no one could know that ϕ without also knowing that ψ when 'ϕ' entails 'ψ'.

But new information *is* gained through the process of logical deduction: one can hardly claim that upon finding, say, the first proof of Fermat's Last Theorem, the worldwide mathematical community did not learn anything new.[8] This is why some logical results are surprising.

Hintikka proposes to reject (4) and claim that not all epistemically possible worlds are logically possible: "the source of the trouble is obviously the last assumption (4) which is usually made tacitly, maybe even unwittingly. It is what prejudices the case in favour of logical omniscience" [15, p. 476]. Hintikka's reason for supposing that epistemically possible worlds need not be logically possible is as follows.

> Just because people ... may fail to follow the logical consequences of what they know *ad infinitum*, they may have to keep a logical eye on options which only *look* possible but which contain hidden contradictions. [15, p. 476]

The worlds that are epistemically possible for agent a from a world w should not be thought of as giving us the possibilities left open by what a knows at w; instead, they should give us the *apparent* possibilities (apparent, that is, given a's ability to "follow the logical consequences" of what she knows). Hintikka devotes the remainder of his article [15, pp. 477–483] to describing such 'impossible' possible worlds.

Impossible possible worlds are "worlds so subtly inconsistent that the inconsistency could not be known (perceived) by an everyday logician, however competent" [15, p. 478]. To explain how this is so, a short detour must be taken. Following Hintikka [13] and [14], quantified sentences may be interpreted in a game theoretic way.[9] The world is viewed as an urn from which individuals are drawn by the two players, called '$\forall$' and '$\exists$'. In a game G of the form $G[\forall x\phi(x)]$, player $\forall$ must pick an individual from the urn satisfying 'ϕ'; if she picks individual a, the game continues as $G[\phi(a)]$. Similarly, the game $G[\exists x\phi(x)]$ requires $\exists$ to pick an individual a satisfying 'ϕ' and continues as $G[\phi(a)]$. Analogously, $\forall$ decides whether $G[\phi \wedge \psi]$ proceeds as $G[\phi]$ or as $G[\psi]$ whereas $\exists$ decides how $G[\phi \vee \psi]$ should proceed. The game $G[\neg\phi]$ proceeds as the inverse game $\overline{G[\phi]}$, in which the players swap roles.

[8] I discuss a number of cases in which mathematical and logical reasoning provides an agent with genuinely new information in [18].

[9] Hintikka develops ideas found in [11] and [25]. The 'urn' terminology is taken from probability theory.

In this way, nested quantifiers represent constraints on sequences of draws from the urn. Just as in elementary probability theory, individuals can but need not be replaced after being drawn from the urn. Models in which all individuals are replaced immediately after being drawn are the *invariant* models; all others are *changing* models. Invariant models correspond to classical first-order semantics whereas draws without replacement correspond to what Hintikka terms the *exclusive* interpretation of the quantifiers.

From a classical point of view, all and only the invariant models count as genuine possible worlds. Hinkikka's idea is that, given a sentence of certain game-theoretic complexity, there is a set of variant models "which vary so subtly as to be indistinguishable from invariant ones at a certain level of logical analysis" [15, p. 483]. Hintikka's notion of 'logical analysis' is related to the maximum number of nested quantifiers found in the sentence to be evaluated in playing the game: this is the *depth* d of the sentence. For any finite sequence of d draws from the domain, models that are in fact variant will behave as invariant models with respect to sentences of depth d. They will agree with the classical models on the truth value of sentences that require no more than d draws to fully evaluate their truth value:

> in [a sentence] p [of depth d] we are considering at most d successive draws of individuals from the model that is supposed to make p true or false. Hence the question as to whether a person a who knows that p *has to know* also a certain logical consequence q of p is naturally discussed by reference to ... sequences of at most d draws of individuals from the domain. This many draws he will have to consider in spelling out to himself what p means, whereas there is no logically binding reason why he should consider sequences of draws of any greater length.
>
> [15, p. 482]

In order to investigate these notions in more detail, a few of the technical details are needed. An important notion is that of an *urn model*, introduced by Rantala in [29].

DEFINITION 1 (Urn sequence). Let $\mathcal{D}$ be a domain of individuals. An urn sequence Δ is a countable sequence $\langle D_i \mid i \in \mathbb{N}\rangle$ where $D_1 = \mathcal{D}$ and, for $i \geq 1$, $D_i \subseteq \mathcal{D}^i$ (the ith Cartesian power of the domain $\mathcal{D}$) such that, for some $a' \in \mathcal{D}$:

$$\langle a_1 \cdots a_i\rangle \in D_i \text{ only if } \langle a_1 \cdots a_i a'\rangle \in D_{i+1}$$

and, for all $a' \in \mathcal{D}$:

$$\langle a_1 \cdots a_i\rangle \in D_i \text{ if } \langle a_1 \cdots a_i a'\rangle \in D_{i+1}$$

DEFINITION 2 (Urn model). Let $\mathcal{L}$ be a first-order language and $\mathcal{M}$ be a first-order structure whose domain is $\mathcal{D}$, assigning an element of $\mathcal{D}$ to each constant and a set of n-tuples to each n-ary relation letter of $\mathcal{L}$. Then an urn model $\mathfrak{M}$ is a pair $\langle \mathcal{M}, \Delta \rangle$, where Δ is an urn sequence $\langle D_1 D_2 \cdots \rangle$.

Each D_i in Δ is a set of sequences of length i.[10] D_{i+1} can then be built from D_i as follows. For each sequence $\sigma \in D_i$, choose an individual $a \in \mathcal{D}$ and add the sequence σ appended by a to D_{i+1}. D_{i+1} is then the smallest set constructed in this way. For example, if $\mathcal{D} = \{a, b\}$, then each D_i will contain sequences of as and bs of length i. The first three elements of two possible urn sequences Δ and Δ' are shown in figure 1. In the first, the ith element of Δ is just $\mathcal{D}^i$, the ith Cartesian power of the domain $\mathcal{D}$; but in the second, D'_3 is a proper subset of $\mathcal{D}^3$. Each D_i can be thought of as containing the individuals that could have been drawn from the urn, in order, in the first i draws. Since D'_3 (in Δ') is a proper subset of D_3 (in Δ) in the example, Δ' represents an urn with less choices after the second draw than Δ. Intuitively, less individuals are available in the Δ' urn at the third draw than in the Δ urn.

$$\Delta = \quad \{a, b\} \quad \left\{ \begin{array}{c} \langle aa \rangle, \langle ab \rangle \\ \langle ba \rangle, \langle bb \rangle \end{array} \right\} \quad \left\{ \begin{array}{c} \langle aaa \rangle, \langle aba \rangle, \langle aab \rangle, \langle abb \rangle \\ \langle baa \rangle, \langle bab \rangle, \langle bba \rangle, \langle bbb \rangle \end{array} \right\} \quad \cdots$$

$$\Delta' = \quad \{a, b\} \quad \left\{ \begin{array}{c} \langle aa \rangle, \langle ab \rangle \\ \langle ba \rangle, \langle bb \rangle \end{array} \right\} \quad \left\{ \begin{array}{c} \langle aaa \rangle, \langle aba \rangle \\ \langle baa \rangle, \langle bba \rangle \end{array} \right\} \quad \cdots$$

Figure 1. Possible urn sequences

The notion of an individual being available at a particular draw i can be made precise as follows. If $i = 1$, then all individuals in the domain are available; otherwise, an individual a_i is available at i iff there is a sequence $\sigma = \langle a_1 \cdots a_{i-1} a_i \rangle \in D_i$. For $i > 1$, the set δ_i of individuals from which to choose from at draw i in an urn model $\mathfrak{M} = \langle \mathcal{M}, \Delta \rangle$ is:

$$\{a_i \mid \exists a_1 \cdots \exists a_{i-1} \langle a_1 \cdots a_{i-1} a_i \rangle \in D_i\}$$

Note that each $\delta_i \subseteq \mathcal{D}$ and $\delta_1 = \mathcal{D}$, as expected. An invariant model is one in which $\delta_i = \delta_{i+1}$, for each $i \in \mathbb{N}$:

[10] This is only strictly true for $i > 1$; when $i = 1$, D_i contains individuals, not sequences. But it does not hurt to identify sequences of length 1 with the individual they contain if the type signature of the sequences is not vital.

Definition 3 (Invariant models). Let $\mathcal{M}$ be a first-order structure with domain $\mathcal{D}$ and $\mathfrak{M} = \langle \mathcal{M}, \langle D_1 D_2 \cdots \rangle\rangle$ be an urn model. $\mathfrak{M}$ is said to be invariant iff, for every $i \in \mathbb{N}$, $\delta_i = \mathcal{D}$. Otherwise, $\mathfrak{M}$ is a changing model.

It follows immediately that there can be only one invariant urn model $\mathfrak{M}$ over each first-order structure $\mathcal{M}$. In the case of a changing model, imagine that the urn has a mechanism that can remove a number of individuals in between each draw and can replace some or all of these individuals later on. The first draw is always made with the full stock of individuals and the mechanism can never remove all of them.[11] In the example in figure 1, Δ' is a changing model: a is available for the first two draws but cannot be drawn on the third.

The interesting feature of changing models is that, up to a certain number of draws, they behave just like invariant models. Suppose an urn model $\mathfrak{M}$ over $\mathcal{M}$ and $\mathcal{D}$ is a changing model, so for some i, $\delta_i \subset \mathcal{D}$. Take the least such i; then $D_1 \cdots D_{i-1}$ is identical to the initial segment of length $i-1$ of the invariant urn model over $\mathcal{M}$. This means that $\mathfrak{M}$ will agree with the corresponding invariant model on the truth value of all sentences of depth $< i$.

An urn model $\mathfrak{M}$ satisfies a sentence 'ϕ', written '$\mathfrak{M} \approx\!\!\!\!| \ \phi$', when $\exists$ has a winning strategy in the game $G[\phi]$ played in $\mathfrak{M}$.[12] Let us take as our example the sentence 'someone knows everyone', formalized as '$\exists x \forall y Rxy$'. Let a model $\mathcal{M}$ with domain $\{a, b\}$ assign $R^{\mathcal{M}} = \{\langle aa\rangle, \langle ab\rangle, \langle bb\rangle\}$ and let $\mathfrak{M} = \langle \mathcal{M}, \Delta\rangle$ and $\mathfrak{M}' = \langle \mathcal{M}, \Delta'\rangle$ where Δ and Δ' are as in figure 1. It is clear that whether the game is played in $\mathfrak{M}$ or $\mathfrak{M}'$, $\exists$ has a winning strategy in picking a. Thus, both $\mathfrak{M} \models\!\!\approx \exists x \forall y Rxy$ and $\mathfrak{M}' \models\!\!\approx \exists x \forall y Rxy$.

Now change the example to 'everyone knows someone who knows everyone', i.e. '$\forall x \exists y \forall z (Rxy \wedge Ryz)$'. $\exists$ has a winning strategy in $\mathfrak{M}'$ only, for here $\forall$ is forced to pick a on the third draw. In $\mathfrak{M}$, $\forall$ has a winning strategy in drawing b first of all, for b does not know anyone who knows a. Hence, $\mathfrak{M}' \models\!\!\approx \forall x \exists y \forall z (Rxy \wedge Ryz)$ but $\mathfrak{M} \not\models\!\!\approx \forall x \exists y \forall z (Rxy \wedge Ryz)$. The reason that

[11] The existential condition on sequences in D_{i+1} ensures that each δ_i is nonempty.

[12] A standard recursive definition of '$\models\!\!\approx$' could be given as follows. First, define satisfaction-at-draw-i:

$$\mathfrak{M} \models\!\!\approx^i R^n(c_1, \ldots, c_n) \text{ iff } \langle c_1^{\mathfrak{M}} \cdots c_n^{\mathfrak{M}}\rangle \in R^{\mathfrak{M}} \text{ and } \{c_1^{\mathfrak{M}}, \ldots, c_n^{\mathfrak{M}}\} \subseteq \delta_i$$

i.e. the standard base clause with the proviso that the required individuals are available for selection at draw i. The clauses for Booleans are then perfectly standard. Next, $\mathfrak{M} \models\!\!\approx^i \exists x \phi(x)$ iff there is an individual $a \in \delta_i$ such that $\mathfrak{M}_a^c \models\!\!\approx^{i+1} \phi(c)$. Finally, $\mathfrak{M} \models\!\!\approx \phi$ iff $\mathfrak{M} \models\!\!\approx^i \phi$ for some i.

both urn models agree on the former but disagree on the latter sentence is that the former has depth 2, the latter depth 3 and $\delta_2 \supset \delta_3$ in $\mathfrak{M}'$.

It is in some ways surprising that there can be a model of '$\forall x \exists y \forall z(Rxy \wedge Ryz)$' in which $\langle ba \rangle \notin R^{\mathfrak{M}}$. This is the motivation behind calling changing urn models *impossible* possible worlds in the modal setting. Which of these worlds will count as epistemically possible for a given agent from a certain world w? Here Hintikka's notion of logical competence comes into play. It is only when considering sentences of depth ≥ 3 that $\mathfrak{M}'$ can be seen to be a changing model. If an agent's competence does not extend to sentences of depth 3 or greater then urn models such as $\mathfrak{M}'$ can appear possible and so be epistemic possibilities for that agent. In general, epistemic possibility is dependent on the notion of d-invariance, where d represents the number of draws from the urn required to evaluate a sentence.

DEFINITION 4 (d-invariant urn models). Let $\mathfrak{M} = \langle \mathcal{M}, \langle D_1 D_2 \cdots \rangle \rangle$ be an urn model as above such that $D_1 = \delta_1 = \cdots = \delta_d$. Then $\mathfrak{M}$ is said to be d-invariant.

Clearly, d-invariant urn models are d'-invariant for all $d' \leq d$ and all invariant models are d-invariant for all $d \in \mathbb{N}$. Each changing model must be d-invariant for some d, the lower limit being $d = 1$. d-invariant models behave as invariant models for all sentences of depth $\leq d$. All classically valid sentences are satisfied by all invariant worlds but a valid sentence of depth d need not be satisfied by a d'-invariant model, for any $d' < d$.

The accessible worlds for an agent are those worlds that look invariant to that agent. Agents who consider changing worlds to be invariant take some impossible worlds to be possible and so are not modelled as knowing all classically valid sentences or all classical consequences of what they know. Agents are instead modelled as knowing all d-consequences of what they know, where d is the measure of that agent's logical competence ('ψ' is a d-consequence of 'ϕ' iff all d-invariant urn models that satisfy 'ϕ' also satisfy 'ψ'). In this way, Hintikka's solution to what I have termed *Hintikka's problem* is to replace (4) by:

(4$'$) All epistemically possible worlds are urn models.[13]

There are two problems to be addressed here. Firstly, what happens in cases of quantifier-free sentences? and secondly, does closure of knowledge

[13] Or are isomorphic to urn models. This wording is preferred by those who think that worlds in general are not mathematical structures but can be described in terms of them.

under d-consequence fare better than the principle of closure under classical consequence?

Urn models must be defined relative to a nonempty domain in a first-order structure, so how can their use as a tool in propositional epistemic logic be judged? Given the methodology of urn models, the intuitive answer should be that any agent who knows anything at all should know all tautologies automatically. This is indeed a formal consequence of Rantala's semantics: $\approx\!\!\!\models \phi$ whenever 'ϕ' is a propositional tautology [29, p. 466, theorem 1]. Hence, '$\mathsf{K}_a\phi$' is valid in the urn models semantics whenever 'ϕ' is a propositional tautology and so every agent is modelled as knowing all propositional tautologies. Moreover, if 'ψ' is a consequence of 'ϕ' in classical propositional logic, then $\phi \mathrel{|\!\!\approx} \psi$ and so any agent that knows that ϕ must also know that ψ. Knowledge is modelled as being closed under propositional logical consequence. These are both forms of logical omniscience that clearly have not been avoided.

Secondly, it is not clear that talking of an agent's logical competence in terms of the quantifier depth of sentences is accurate or helpful in a discussion of knowledge. Consider a mathematician going through the process of proving some theorem, expressed as a sentence 'ϕ' of depth n, say. To begin with, she does not know that ϕ or that ψ (of depth $n-1$) even though, let us suppose, 'ψ' follows from what she does know. She then proves and thus comes to know that ϕ. What are we say of her logical competency? If she must make at least n draws from the domain in order to understand what 'ϕ' means, as Hintikka suggests, then she must have a logical competency of at least n. If so, all worlds epistemically accessible to her are n-invariant worlds and hence she will be modelled as knowing that ψ as well. 'ϕ' and 'ψ' might be completely unrelated sentences and so we have no reason to think that coming to know that the former is true must result in knowledge of the truth of the latter.

On the other hand, if there are worlds accessible to her that are m-invariant but not n-invariant, for some $m < n$, then she need not be modelled as knowing that ψ. But then in what sense does the 'n' parameter relate to her logical competence? She proved that 'ϕ' (of depth n) is true, which should clearly reflect on what we take her logical competency to be. In sum, either there are unwanted closure properties modulo the agent's logical competency or else there is no conceptual link between the agent's competency and the 'd' parameter of the d-invariant worlds.

4. Cresswell's Nonclassical Worlds

Max Cresswell [6] develops an account that has proved popular in the subsequent literature on propositional attitudes and logical omniscience.[14] Cresswell accepts the motivation behind calling logical omniscience a *problem* for theories of belief and knowledge, writing that:

> there is no reason why someone should not take a different propositional attitude (belief, say) to two propositions that are logically equivalent. And when a mathematician discovers the truth of a mathematical principle he does not thereby discover the truth of all mathematical principles. [6, p. 40]

His suggestion is the partitioning of worlds into two disjoint classes: the classical and the nonclassical.[15] Nonclassical worlds are worlds that do not obey all the usual logical laws so that not all truths of classical logic are true at such worlds. Part of Cresswell's concern is to be able to distinguish between the content of logically equivalent propositions. We intuitively want to say that there is *some* difference between the proposition that all transitive, irreflexive binary relations are asymmetrical, for example and a trivial proposition such as that expressed by '$a = a$'. Cresswell thinks of propositions in terms of sets of possible worlds. If all of the available worlds are classical then each of these propositions is identical to the entire domain of worlds. But if nonclassical wolds are added to the domain, the situation changes as a nonclassical world can belong to one proposition and not the other. Nonclassical worlds thus allow us to distinguish between logically equivalent propositions.

The nonclassical worlds approach can be used whether we think of propositions as sets of worlds or not. Within Hintikka's framework, we can allow that some of the epistemically possible worlds are nonclassical. If a nonclassical world, at which some classically valid sentences are not true, is epistemically accessible to agent a, then a will not be modelled as knowing all classically valid sentences. It is common within this approach to restrict the definition of validity to the classical worlds only, so that the valid sentences within this logic coincide with the classically valid sentences. Then agents are not modelled as knowing all the valid sentences of the epistemic logic and so a form of logical omniscience has been avoided.

---

[14] See [8, 10, 20, 21, 22, 27, 34, 35].

[15] Cresswell coins the term 'nonclassical' at [5, p. 354] but acknowledges a similar notion due to Richard Montague, under the label 'designated points of reference' [24, p. 382].

Allowing nonclassical worlds to function as epistemically possible worlds is essentially the move that Hintikka made in introducing his impossible possible worlds. Both Hintikka and Cresswell deny proposition (4) of what I termed Hintikka's problem (which states that every epistemically possible world is logically possible). The advantage of this manoeuvre is that the satisfaction clause for sentences of the form '$\mathsf{K}_a\phi$' remains unchanged from Hintikka's original proposal, thus retaining the intuition that gaining knowledge amounts to a restriction on the situations that the agent considers possible. There are several questions raised by this approach. Firstly, just what are nonclassical worlds? and secondly, are they the right logical tool to overcome the problem of logical omniscience? We shall deal with each of these questions in turn, beginning with the nature of nonclassical worlds.

Carnap introduces the notion of a possible world as a *state description* [3, p. 9], a set of atomic sentences each of which has a definite truth value independently of any other atomic sentence.[16] Worlds constructed from sentences are often called *ersatz worlds*. It is easy to see how this conception could be modified to accommodate nonclassical worlds: we simply modify the rules for the assignments of truth values to atomic sentences. Since worlds are viewed as sets of sentences, classical and nonclassical worlds are ontologically on a par.

Cresswell believes that the nonclassical worlds approach can be made to work even if one denies that possible worlds are sets of sentences:

> if possible worlds are taken as primitive then there is nothing to stop us from taking a subset and saying that these are the ones that are genuinely possible worlds; the others are in some sense impossible.
>
> [6, p. 40]

Following the work of Kripke and others on the semantics of modal logic in the 1960s, the tendency now (at least amongst modal logicians) is to view possible worlds as primitive elements of the theory. The question is, how can a primitive entity, thought of as a world, be at the same time "in some sense impossible"? Cresswell's answer is that the nonclassical worlds are not worlds at which the impossible happens: they are not genuinely impossible worlds. Rather, they are worlds at which the connectives have a nonstandard meaning. There may be worlds at which, for example, '¬' has

[16] Carnap is appealing to Wittgenstein's notion of *elementary propositions* in the *Tractatus* [37] at §5.3: "Every proposition is the result of truth operations on elementary propositions" and §5.134: "One elementary proposition cannot be deduced from another" [37].

an intuitionistic or paraconsistent meaning. A sentence which would be a contradiction under the classical interpretation of its constants can be true at worlds in which those connectives receive a different interpretation.[17]

Cresswell talks of the meaning of the connectives in terms of *meaning rules* which link a connective to its denotation, a mathematical function. The idea is that the denotation of the connectives differs from world to world. The classical worlds are the worlds at which the connectives denote the usual truth-functions, given by the truth-tables. At this point, one may reiterate Wittgenstein's "fundamental thought" from the *Tractatus* that "the "logical constants" do not represent" [37, 4.0312], i.e. that we should not talk in terms of logical constants denoting at all, let alone denoting different functions at different worlds. But we will not take this line, as it closes down the discussion immediately.

Let us instead assume that the logical constants do denote mathematical functions. What sense can we make of them denoting different constants at different worlds? One option, which we should reject, is that logical constants denote different functions at different worlds because different mathematical functions exist at different worlds, in the same way that the domain of individuals need not be constant across worlds. Mathematical function are not world-bound entities; they do not exist *at* worlds at all. Worlds, so far as they have any being at all, are maximal spatio-temporal states of affairs (and to maintain as Cresswell does that there are no worlds at which the impossible happens, they must be maximal in the classical sense). Mathematical functions, on the other hand, are abstract world-independent entities. This is part of what Wittgenstein meant when he claimed that the logical constants do not represent or denote: there is nothing in the world (or in any world) for them to represent.[18]

On this (broadly Platonist) conception of mathematical functions as necessary, world-independent entities, we can easily accommodate a wealth of logical entities, e.g. classical, intuitionistic, paraconsistent and relevant entailment relations. The question that arises is, what is it about a particular world that makes a logical symbol denote one of these functions rather than another? The answer that immediately presents itself is that it is facts about the way that a symbol is used, perhaps within some specified sub-community

---

[17] Intuitionistic negation is Cresswell's example, supposed to explain why double negation can fail at nonclassical worlds.

[18] Famously, Wittgenstein went further than this in the *Tractatus*, claiming that those things outside the limit of the world cannot be talked about meaningfully at all.

of logical experts, that determines its meaning. Any theory of meaning has to factor in use of the term in question to some degree or other. Had that symbol been used in certain other ways, it would have had a different meaning.

A comment of Cresswell's goes some way to supporting this reading. He notes that:

> The fact that we can reinterpret $\wedge$ and $\neg$ so that $\phi \wedge \neg\phi$ is true in a possible world no more shews us how a contradiction could ever be true than calling birds 'pigs' shews us how pigs could fly.[19] [6, p. 41]

If the facts that fix which function a particular symbol denotes at a world are facts about how that symbol is used, then it is clear that the facts that fix what '$\wedge$' and '$\neg$' mean could have been different without in the least suggesting that contradictions could have been true. This view takes non-classical worlds to be worlds that differ from our own only in the usage of logical symbols and not worlds at which impossible events occur. Cresswell's comment provides support for this notion of non-classical worlds. His analogy of calling birds 'pigs' invokes facts about usage of the word 'pigs' (given that there is nothing intrinsically *piggy* about 'pigs').

The problem is that this does not give us the right explanation of an agent's knowledge in terms of epistemic possibility. When an agent comes to know something, according to Hintikka's model, she is ruling out certain scenarios that she previously considered to be possible. If nonclassical worlds are worlds in which the usage of logical constants is nonclassical, then agents who find nonclassical worlds possible (i.e. agents for whom nonclassical worlds are epistemically accessible) should be unable to tell these nonclassical uses of the connectives form their classical use. Yet first-year students, brought up on classical truth-tables, can easily spot whether a particular symbol corresponds to a particular truth-table. They can tell whether a particular way of using '$\wedge$' is classical or nonclassical. They are able to distinguish between classical and nonclassical uses of the logical constants and so could discriminate the classical from the nonclassical worlds. Since they believe that the actual world is classical (at least, they take the meaning of the logical constants to be given by the classical truth-tables), they can always discriminate between the actual world and any nonclassical world. On Cresswell's model, therefore, they should know all classical tautologies. Experience with first-year logic classes suggests otherwise; hence this cannot be the correct explanation of what makes nonclassical worlds nonclassical.

[19] Cresswell uses '&' and '$\sim$' for conjunction and negation (I have substituted '$\wedge$ and '$\neg$' for consistency with the current notation).

There is a tension in (what I have taken to be) Cresswell's account of knowledge and the standard possible worlds analysis of other intensional notions, such as necessity and possibility. A sentence 'ϕ' is necessarily true not when an every utterance of 'ϕ' is true at all worlds, for this would make 'I'm here now' a necessary truth, which it certainly is not. 'ϕ' is necessarily true when it is a fact that ϕ and could not have been otherwise. In modal logic, we are tempted to read '$\Box\phi$' as meaning that 'ϕ' is true at every possible world without clarifying just *what* we take to be true or false at a world. Take w to be a world at which 'pigs' means *birds* but is otherwise just like the actual world. This is a world at which birds fly and pigs do not, even though 'pigs fly' can truthfully be uttered by w-landers. 'Pigs fly' is true-in-w-ish but not true-in-English at w, for what *we* mean by 'pigs fly' does not describe the facts at w. Hence, w is not a truthmaker for 'pigs could fly'.

David Kaplan explains the evaluative procedure in terms of direct reference theory [19]. First, an uttered sentence expresses a *Russellian proposition*, a structured entity containing the denotations of the uttered referring terms. In the case of sentences containing indexical or demonstrative elements, context determines the contribution of that term to the proposition. At this stage, temporal or modal operators are ignored. This proposition is then evaluated at a world or group of worlds and a timepoint or timepoints, depending on any modal or temporal operators in the original sentence, producing a determinate truth value. Thus, '$\phi \wedge \neg\phi$' expresses a proposition that is logically structured by containing whichever functions *we* mean by '$\wedge$' and '$\neg$'. For that proposition to be evaluated as true at some world, that world must genuinely be a world where the impossible happens, where contradictions are genuinely true.

In order for nonclassical worlds to do the job we require of them, they must be genuinely impossible worlds: not merely worlds at which pigs fly but worlds at which pigs both can and cannot fly. Zalta [39] provides a theory of genuine (as opposed to ersatz) impossible worlds, based on his theory of abstract objects [38]. Within this theory, no contradiction is true; however, contradictions can be true at impossible worlds. Priest also accepts impossible words into his ontology [27]. He argues for *noneism*, a variant on Meinongianism put forward by Richard Routley [30], according to which some objects (including impossible worlds) do not exist.[20] Here is not the place to

[20] Such non-existent objects do not *subsist* either; they have no being at all, according to Priest [27, pp. 13–15]. Although Meinong held that abstract entities such as numbers subsist (but do not exist), he also held that merely possible objects (and impossible objects)

evaluate these theories. I will assume for the sake of argument that there is at least one defensible notion of impossible (i.e. nonclassical) worlds.[21] In the next section, I will evaluate the extent to which the inclusion of nonclassical worlds is an adequate response to the problems discussed above.

5. Evaluating the Nonclassical Approach

I am going to take nonclassical worlds to be worlds at which certain contradictions can be true, i.e. worlds governed by paraconsistent rather than classical logic. To keep the presentation simple, I will concentrate on the propositional case only, in which both a proposition p and its negation $\neg p$ can be true simultaneously. In a paraconsistent setting, a valuation $\mathcal{V}$ assigns each primitive proposition p a *subset* of $\{true, false\}$ rather than one of its members (as a classical valuation does), so that p is true under $\mathcal{V}$ iff $true \in \mathcal{V}(p)$ and false under $\mathcal{V}$ iff $false \in \mathcal{V}(p)$. Primitives can be both true and false under a valuation (when $\mathcal{V}(p) = \{true, false\}$); they can also fail to be either (when $\mathcal{V}(p) = \varnothing$).

The truth conditions for complex propositions are just as in the classical case: '$\neg\phi$' is true under a valuation iff 'ϕ' is false and false iff 'ϕ' is true under that valuation (the second clause is necessary, as 'ϕ' not being false is not sufficient for the falsity of '$\neg\phi$'). '$\phi \wedge \psi$' is true under a valuation iff both 'ϕ' and 'ψ' are and false iff either 'ϕ' or 'ψ' are. A proposition 'ϕ' is a paraconsistent consequence of a set of propositions Γ iff 'ϕ' is true under every valuation under which every member of Γ is also true.

Nonclassical worlds are (or are isomorphic to) paraconsistent models. In a modal setting, '$w \Vdash_{\mathrm{t}} p$' means that 'p' is true at w and '$w \Vdash_{\mathrm{f}} p$' means that 'p' is false at w (in a model, which is implicit). $\mathcal{V}$ now assigns each primitive proposition-world pair a subset of $\{true, false\}$. We can formalize the above as follows:

$$
\begin{aligned}
&w \Vdash_{\mathrm{t}} p \text{ iff } true \in \mathcal{V}(p, w)\\
&w \Vdash_{\mathrm{f}} p \text{ iff } false \in \mathcal{V}(p, w)\\
&w \Vdash_{\mathrm{t}} \neg\phi \text{ iff } w \Vdash_{\mathrm{f}} \phi\\
&w \Vdash_{\mathrm{f}} \neg\phi \text{ iff } w \Vdash_{\mathrm{t}} \phi\\
&w \Vdash_{\mathrm{t}} \phi \wedge \psi \text{ iff } w \Vdash_{\mathrm{t}} \phi \text{ and } w \Vdash_{\mathrm{t}} \psi\\
&w \Vdash_{\mathrm{f}} \phi \wedge \psi \text{ iff } w \Vdash_{\mathrm{f}} \phi \text{ or } w \Vdash_{\mathrm{f}} \psi
\end{aligned}
$$

---

neither exist nor subsist [23].

[21] Priest defends the *dialethist* view that certain contradictions can be true in actuality [26]. From this point of view, the logical laws governing the actual world are nonclassical.

Classical worlds are worlds w for which $V(p, w)$ is a singleton, for every primitive proposition p; all other worlds are nonclassical. The definition of knowledge is then as before:

$w \Vdash_{\mathrm{t}} \mathsf{K}_a\phi$ iff $w' \Vdash_{\mathrm{t}} \phi$ for all worlds w' such that $R_a ww'$
$w \Vdash_{\mathrm{f}} \mathsf{K}_a\phi$ iff it is not the case that $w \Vdash_{\mathrm{t}} \mathsf{K}_a\phi$

Because both '$w \Vdash_{\mathrm{t}} \mathsf{K}_a\phi$' and '$w \Vdash_{\mathrm{f}} \mathsf{K}_a\phi$' are defined in terms of '$\Vdash_{\mathrm{t}}$', sentences of the form '$\mathsf{K}_a\phi$' behave classically. Does this move overcome Hintikka's problem? It does allow for models of agents that do not know all propositional tautologies and do not know all consequences of what they know. A version of Hintikka's problem can, however, be sentenceted as follows, by concentrating on *paraconsistent* rather than classical consequence. Each world w is closed under paraconsistent consequence: if 'ψ' is a paraconsistent consequence of 'ϕ' and $w \Vdash_{\mathrm{t}} \phi$, then $w \Vdash_{\mathrm{t}} \psi$. This applies to classical as well as nonclassical worlds, for all paraconsistent consequences of Γ are also classical consequences of Γ. We can then sentencete a version of Hintikka's problem as follows:

(P1) 'a knows that ϕ' is true at w iff ϕ is true at every world epistemically accessible from w;

(P2) There are a, 'ϕ', 'ψ' such that a knows that ϕ, 'ψ' is a paraconsistent consequence of 'ϕ' and yet a does not know that ψ;

(P3) Every world is either a classical or a nonclassical world.

From (P3), the truths at any accessible world are closed under paraconsistent consequence; from (P1) and (P2), 'ϕ' holds at all worlds accessible to a and therefore 'ψ' holds at all those worlds too. From (P1), a knows that ψ; but this contradicts (P2). It should be clear that (P2) is true: agents do not know all paraconsistent consequences of what they know. There are paraconsistent approaches to mathematics (often termed *inconsistent mathematics*), such as da Costa's [7], Brady's [2] and Priest, Routley and Norman's [28].[22] Yet no one knows all of the consequences of these axiomatic systems.

Let us turn from the version of the logical omniscience problem which focuses on closure under valid implication and consider closure under known

[22]This approach to mathematics retains Frege's original abstraction principle (every predicate determines a class) without leading to triviality because of the underlying paraconsistent logic.

implication (which in this context I will take to mean known material implication, rather than known logical consequence): the principle that, if an agent knows that ϕ and that $\phi \to \psi$ then that agent must also know that ψ. This principle is false (at least of actual, as opposed to idealized, cognitive agents) and it can be show to fail when nonclassical worlds are admitted. Since we are dealing with material implication, we define '$\phi \to \psi$' as '$\neg\phi \vee \psi$'. The following diagram shows a counter model to this closure principle:

The arrows show the accessibility relation for agent a. '$p \to q$' is true at both worlds, each of which is accessible from the other and from itself, so that '$\mathsf{K}_a(p \to q)$' is true at both worlds. '$\mathsf{K}_a p$' is also true at both worlds yet 'q' is not at w_2, so $\mathsf{K}_a q$ is not true at either world. We have:

$$w_1 \Vdash \mathsf{K}_a p \wedge \mathsf{K}_a(p \to q) \wedge \neg \mathsf{K}_a q$$

and likewise for w_2. If we suppose that w_1 is a classical world then '$\mathsf{K}_a p \wedge \mathsf{K}_a(p \to q) \to \mathsf{K}_a q$' is invalid and hence the principle of closure under known implication is false.[23] The problem is that the following *is* valid:

$$\mathsf{K}_a\phi \wedge \mathsf{K}_a(\phi \to \psi) \to \mathsf{K}_a(\psi \vee (\phi \wedge \neg\phi))$$

If an agent knows both that ϕ and that ϕ implies ψ, then the agent knows that either ψ or else some contradiction holds. The only worlds accessible to an agent that knows that ϕ and that $\phi \to \psi$ are classical worlds at which 'ψ' is true or else are nonclassical worlds. We can call this the principle of *closure under known implication-or-contradiction*. There is no reason to suppose that knowledge is so closed.

When an agent knows that $\phi \vee \psi$ but neither knows that ϕ or that ψ, it is likely that the agent cannot easily rule out (or does not want to investigate the truth of) either disjunct. But it is highly plausible that

[23]Note that '$\mathsf{K}_a\psi$' and '$\neg\mathsf{K}_a\psi$' cannot both be true at a world, even if that world is nonclassical. The closure principle does hold at worlds from which only classical worlds are accessible; but all that is required for its invalidity is a classical world at which it does not hold, i.e. a classical world from which a nonclassical world is accessible. It remains invalid even if we define the valid sentences to be the sentences true at all worlds, classical or nonclassical.

most agents would infer 'ψ' from '$\psi \vee (\phi \wedge \neg\phi)$', particularly if '$\phi \wedge \neg\phi$' is easily identified as an explicit contradiction. For such sentences 'ϕ', it is implausible that each such agent must know that ψ whenever it knows that ϕ and that $\phi \rightarrow \psi$. The only defence that can be offered in favour of *closure under known entailment-or-contradiction* is that agents who know that ϕ and that $\phi \rightarrow \psi$ thereby *implicitly* know that $\psi \vee (\phi \wedge \neg\phi)$. But whatever this may mean, exactly the same could be said about closure under known implication. If this response were adequate, therefore, the machinery of impossible worlds would be unnecessary.

6. Open Worlds

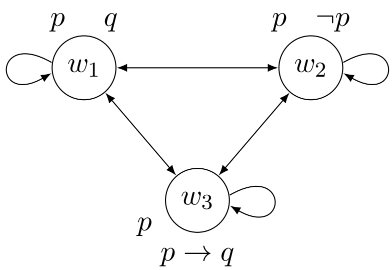
The defender of the possible and impossible worlds model of knowledge must, I contend, accept that there are worlds that are neither classical nor nonclassical. How could this be? Nonclassical worlds have a certain logical structure. They retain just enough of the classical meaning of the Boolean connectives whilst avoiding triviality in the presence of inconsistency. This means that there are (weaker than classical) closure conditions on the set of truths at each nonclassical world, for example, closure under conjunction introduction and elimination, and addition and deletion of double negations. These closure conditions are fairly harmless from a certain normative perspective, for anyone who knows the meaning of '$\neg$' can see that 'ϕ' is equivalent to 'ϕ' preceded by an even number of '$\neg$'s. But if this number is so great that no token of the sentence '$\neg\neg \cdots \neg\phi$' could be uttered or written down in full (say, because it contains more negation signs than the number of electrons in the universe), then how is our agent to decide whether to believe it? If she decides to withhold judgement and does not hold the sentence as true, she cannot know it to be true. On this view (which takes knowing that ϕ to imply that the agent holds the sentence 'ϕ' as true), knowledge is not closed under addition or deletion of double negations.

This notion of knowledge can be modelled by introducing worlds that are not closed under any rule of inference.[24] In so doing, we attain a notion of knowledge that comes far closer to our everyday uses of 'knows' than the possible/impossible worlds account can. I argue for such worlds in [18], where I use them in an analysis of information and epistemic possibility. Priest also makes use of such worlds in [27], where he calls them *open worlds* (in contrast to worlds closed under inference rules). I took an agnostic view

[24] Except, of course, *infer 'ϕ' from 'ϕ'*.

of the ontological commitments of my account in [18], thinking of a open worlds merely as logical tools for analysing intentional notions. Priest is a *noneist*, holding that some objects (including open worlds) do not exist.[25] If the view is plausible in general, then it is an attractive account of open worlds, for we can refer to and quantify over them without offending our sense of what exists.

In the propositional case, open worlds are worlds at which the valuation function assigns a truth value to each sentence in the language, rather than just to the primitive propositions.[26] In the first-order case, each open sentence is assigned an extension and an anti-extension, such that '$\phi(t_1, \ldots, t_n)$' is true iff the tuple containing the denotations of 't_1' through to 't_n' is in the extension of 'ϕ' and false iff this tuple is in the anti-extension of 'ϕ'.[27] Closed worlds avoid the *closure under known implication-or-contradiction* problem, as can be seen in the following diagram, in which w_3 is an open world:



Nothing is true at w_3 except the sentences marked on the diagram. In particular, neither 'q' nor '$p \wedge \neg p$' are true at w_3.

It is not clear that this approach has any benefits, other than rescuing the possible worlds account of knowledge from the logical omniscience problems. The advantage of the possible worlds framework, it is often claimed, is that it uncovers logical properties of knowledge, albeit the knowledge of idealized agents only (that is, agents who know all consequences of their knowledge without computational effort). When we factor open worlds into the account, the modelled concept of knowledge loses any interesting logical properties.[28]

[25] See footnote 20.

[26] In [18], I restricted valuations to sentences that do not contain epistemic operations such as 'K' but this is not essential.

[27] Open sentences are thus treated in the same way that predicates are dealt with in classical first-order semantics.

[28] There remain valid sentences containing the 'K_a' operator, for example ones obtained by uniformly substituting '$\mathsf{K}_a p$' for any primitive in a propositional tautology, but these

This is an advantage to some extent, for it is these very principles (the closure principles) that give rise to the logical omniscience problems.

The worry is not that the resulting account is false but that exactly the same results could be achieved with far less logical machinery. Given agent a's epistemic accessibility relation R_a, set $R_a|w = \{w' \mid R_a ww'\}$. Now set

$$\mathbf{K}_a^w = \bigcap_{w' \in R_a|w} \{\phi \mid w' \Vdash \phi\}$$

Each $\mathbf{K}_a^w$ is an arbitrary unstructured set of sentences. It is easy to see that $w \Vdash \mathsf{K}_a\phi$ iff $\phi \in \mathbf{K}_a^w$. For any agent a, we can define an accessibility *function* f_a from worlds to worlds by setting $f_a(w) = w'$ iff $\{\phi \mid w' \Vdash \phi\} = \mathbf{K}_a^w$. Assuming that no two worlds generate the same set of truths as one another (i.e that indiscernible worlds are identical), f is genuinely functional.[29] Because any set of sentences is the set of truths at some open world, f is total. Then $w \Vdash \mathsf{K}_a\phi$ iff $f(w) \Vdash \phi$.

It follows that any model $\mathcal{M}$ containing an accessibility relation R_a for each agent a can be replaced by a model $\mathcal{M}'$ which contains an accessibility function f_a in place of each R_a. We can define a relation $\Vdash'$ in just the same way as $\Vdash$, except the clause for '$\mathsf{K}_a\phi$' becomes: $w \Vdash' \mathsf{K}_a\phi$ iff $f_a(w) \Vdash' \phi$ in $\mathcal{M}'$. Then:

$$w \Vdash \phi \text{ (in } \mathcal{M}\text{) iff } w \Vdash' \phi \text{ (in } \mathcal{M}'\text{)}$$

(this follows from the agreement of $\Vdash$ and $\Vdash'$ on primitive and Boolean sentences and the definition of the accessibility functions). This shows that each world need be related to just one open world for each agent a. If we are concerned only with the actual knowledge of agents $a_1, \ldots, a_n$, then the model need contain only the actual world @ and n open worlds. In such models, Hintikka's intuition that gaining knowledge amounts to excluding certain possibilities (modelled as worlds) is lost completely and yet the models that Hintikka recommends, modified to include open worlds, appear to provide no logical benefits over the models we have just defined.

The criticism can be taken further by replacing each open world $f_a(w)$ in the model by the set $\mathbf{K}_a^w$, agent a's *knowledge set* at w. The resulting models contain an arbitrary set of sentences $\mathbf{K}_a^w$ for each agent a and $w \in S$ but do not contain any accessibility functions or relations. In these models,

are not logical principles concerning the '$\mathsf{K}_a p$' operator itself, as opposed to any other sentence-forming symbol and hence not *epistemic* principles.

[29] If this is not the case, f can be made functional by picking the first world that fits the definition in some canonical ordering of all worlds.

we set $w \Vdash \mathsf{K}_a\phi$ iff 'ϕ' $\in \mathbf{K}_a^w$.[30] In this approach, knowledge is modelled *sententially*, with the semantics for '$\mathsf{K}_a\phi$' given in terms of membership of certain sets of sentences.[31]

It is often objected that sentential approaches give us "ways of *representing* knowledge rather than *modelling* knowledge" and so, the thought runs, "[o]ne gains very little intuition about knowledge from studying syntactic [i.e. sentential] structures" [9, p. 320]. This is because the sentential approach, it is claimed, "lacks the elegance and intuitive appeal of the semantic approach" [8, p. 40]. But as we have seen, the open worlds semantics collapses into a sentential semantics. We can agree that the possible worlds semantics for intentional attitudes is both elegant and intuitive; the problem is that, without open worlds factored in, it cannot overcome the logical omniscience problems. Open worlds, considered as worlds of any kind, are counterintuitive. To accept them as genuine entities, one must be a plenitudinous (or 'full-blooded') Platonist who accepts a paraconsistent notion of being (that is, any object that could exist, whether it is consistent or not, does exist; see [1]) or else a noneist, holding that some entities, including open worlds, do not exist.[32] The model I have described is certainly elegant in its simplicity, for we need but one knowledge set for each agent per world in the model and have no need for epistemic accessibility relations.

There are a number of responses to this line of criticism. The first is that the possible/open worlds analysis of knowledge (and other intentional attitudes) can accommodate properties of knowledge by imposing or relaxing properties on the epistemic accessibility relations between worlds. Reflexivity guarantees that knowledge entails truth (i.e. '$\mathsf{K}_a\phi \rightarrow \phi$' is valid on reflexive frames), transitivity results in models of agents who know what they know ('$\mathsf{K}_a\phi \rightarrow \mathsf{K}_a\mathsf{K}_a\phi$' is valid on transitive frames) and, by taking the accessibility relation to be Euclidian, we model agents who know that they do not know whatever they do not know: '$\neg\mathsf{K}_a\phi \rightarrow \mathsf{K}_a\neg\mathsf{K}_a\phi$' is valid on

[30] If we are concerned with the agent's actual knowledge only, then we require just one knowledge set $\mathbf{K}_a$ for each agent a and set $@ \Vdash \mathsf{K}_a\phi$ iff 'ϕ' $\in \mathbf{K}_a$.

[31] This approach is often called the *syntactic* approach to epistemic logic (e.g. in [9]) in contrast to the *semantic* approach, in which possible, impossible and open worlds play an integral role in the semantics of intentional attitudes.

[32] Beall describes this version of Platonism as *really full-blooded Platonism*. It is the thesis that every mathematical theory, whether consistent or not, truly describes mathematical reality. Presumably Beall would want *some* restrictions in place on what counts as a mathematical theory, e.g. that it be non-trivial.

Euclidian frames.[33] The second and third of these principles are known as *positive* and *negative* introspection respectively. The duality between properties of a's knowledge and properties of R_a is lost in the sentential models.

This is true, but we should be careful to distinguish between what we *can* model and what, from a logician's perspective, looks tidy. The sentential account can easy accommodate knowledge as a notion that implies truth by imposing the condition:

$$\mathbf{K}_a^w \subseteq \{`\phi\text{'} \mid w \Vdash \phi\}$$

for every possible world w, so that '$\mathsf{K}_a\phi \to \phi$' is valid.[34] Let us turn to positive and negative introspection. If our reasons for rejecting logical omniscience were that there are agents that cannot possibly know all consequences of what they know because of their cognitive limitations, then we are likely to reject both positive and negative introspection as universal principles for precisely the same reasons. If we do want to model introspective agents in the sentential approach, however, we simply require that '$\mathsf{K}_a\phi$' $\in \mathbf{K}_a^w$ if 'ϕ' $\in \mathbf{K}_a^w$ and/or '$\neg\mathsf{K}_a\phi$' $\in \mathbf{K}_a^w$ if 'ϕ' $\notin \mathbf{K}_a^w$, for any sentence 'ϕ'.

A better response for the defender of open worlds is to show that there is a logical advantage to having a multiplicity of open worlds available for modelling even a single agent's intentional attitudes. In [18], I describe one such scenario: modelling what is epistemically possible for an agent. 'It is epistemically possible, for agent a, that ϕ' is taken to be true iff there is a world w', accessible to a from w, such that $w' \Vdash \phi$.[35] To model rational (but not logically omniscient) agents, we need epistemic structures that are richer than the standard ones obtained from possible and open worlds. This is because even the most cognitively burdened agent can see that it is not possible that $p \wedge \neg p$.[36] Roughly, it is epistemically possible for a that ϕ when a cannot spot a conflict (an explicit contradiction) between 'ϕ' and what it already knows. Following this idea, an open world w should be epistemically

[33] R_a is Euclidean iff $\forall xyz(R_axy \wedge R_axz \to R_ayz)$.

[34] We need this to hold only at possible worlds, since validity is defined as truth at all such worlds. Impossible things happen at open worlds and so there is no reason why agents cannot know falsehoods at an open world. This in no way implies that it is possible for any agent to know a falsehood, for open worlds have no bearing on genuine possibility.

[35] This is a direct parallel with alethic possibility, the difference being that not all epistemically accessible worlds need be metaphysically or even logically possible.

[36] I am assuming that 'p' is not the liar sentence or some similarly paradox-inducing construction.

accessible to an agent a only if a could not spot an explicit contradiction within the sentences that hold true at w.

To capture this idea in an epistemic structure, I introduce a further relation T on open worlds. Intuitively, this relation takes a world w and treats the truths at w as premises. For every conclusion 'ϕ' that could be drawn *in just one step of reasoning* from these (e.g. by applying a single natural deduction rule to the truths at w), w is related by T to some world w' at which 'ϕ' is true but is otherwise just like w. Potential ways of reasoning, for the agents in question, from the truths at w_1 are modelled as ω-sequences $w_1 w_2 w_3 \cdots$ of open worlds related by T. The more cognitive resources an agent has, the further it could reason along the sequence. If an explicit contradiction is true at w_{256} and agent a has enough resources to reason this far, then w_1 is not epistemically accessible to a from any possible world, but it may be accessible to another agent who has less resources available. There is no room to describe this framework in detail here and so I refer the interested reader to [18]. If this approach is successful, then it highlights a logical benefit of introducing open worlds into a relational semantics over a purely sentential account of intentional attitudes.

To conclude, Hintikka's notion of defensibility does nothing to remove the sting of the logical omniscience problems. The introduction of impossible worlds is more successful. This is the approach that both Hintikka and Cresswell eventually favour. But so long as all worlds that do not behave as classical possible worlds have a paraconsistent underlying logic, the logic leads to unwarranted conclusions, for example that an agent who knows that ϕ and that $\phi \rightarrow \psi$ must know that $\psi \vee (\phi \wedge \neg\phi)$. We have as little reason to believe this to be true as we do to believe that knowledge is closed under known implication. To avoid the problem, we need to introduce open worlds. But the resulting model collapses into a sentential approach, in which knowledge is modelled by arbitrary sets of sentences. I have suggested a few reasons for thinking that open worlds can be of use in formal epistemology but to do so, we have to take the idea of agents reasoning within their resource bounds seriously and build this feature into our models from the beginning.

References

[1] J.C. Beall, "From full blooded platonism to really full blooded platonism", *Philosophia Mathematica* 7 (1999).

[2] R. Brady, "The consistency of the axioms of abstraction and extensionality in a three-valued logic", *Notre Dame Journal of Formal Logic* 12 (1971), 447–453.

[3] R. Carnap, *Meaning and Necessity*, University of Chicago Press, 1947.

[4] Roderick M. Chisholm, "The logic of knowing", *Journal of Philosophy* 60(25) (1963), 773–795.

[5] M. J. Cresswell, "Classical intensional logics", *Theoria* 36 (1970), 347–72.

[6] M. J. Cresswell, *Logics and Languages*, Methuen and Co., 1973.

[7] N. da Costa, "On the theory of inconsistent formal systems", *Notre Dame Journal of Formal Logic* 15(4) (1974), 497–510.

[8] R. Fagin and J. Y. Halpern, "Belief, awareness and limited reasoning", *Artificial Intelligence* 34 (1988), 39–76.

[9] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi, *Reasoning About Knowledge*, MIT Press, 1995.

[10] R. Fagin, J. Y. Halpern, and M. Y. Vardi, "A nonstandard approach to the logical omniscience problem", in R. Parikh (ed.), *Proceedings of the Third Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 41–55, Morgan Kaufmann, 1990.

[11] Leon Henkin, "Some remarks on infinitely long formulas", in *Infinitistic Methods*, pages 167–183, Pergamon Press, Oxford, 1961.

[12] J. Hintikka, *Knowledge and belief: an introduction to the logic of the two notions*, Cornell University Press, Ithaca, N.Y., 1962.

[13] J. Hintikka, *Logic, Language-Games and Information: Kantian Themes in the Philosophy of Logic*, Clarendon Press, Oxford, 1973.

[14] J. Hintikka, "Surface semantics and its motivation", in H. Leblanc (ed.), *Truth, Syntax and Modality.* North-Holland, Amsterdam, 1973.

[15] J. Hintikka, "Impossible possible worlds vindicated", *Journal of Philosophical Logic* 4 (1975), 475–484.

[16] J. Hintikka, "A second generation epistemic logic and its general significance", in V. F. Hendricks, S. A. Pedersen, and K. F. Jørgensen (eds.), *Knowledge Contributors*, number 322 in Synthese Library. Kluwer Academic, Dordrecht, 2003.

[17] Max O. Hocutt, "Is epistemic logic possible?" *Notre Dame Journal of Formal Logic* 13(4) (1972), 433–453.

[18] Mark Jago, "Imagine the possibilities: Information without overload", *Logique et Analyse* 49(196) (2006), 345–370.

[19] D. Kaplan, "Demonstratives", in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*, chapter 17, pages 481–563, Oxford University Press, New York, 1989.

[20] G. Lakemeyer, "Steps towards a first-order logic of explicit and implict belief", in J. Y. Halpern (ed.), *Proceedings of the First Conference on Theoretical Aspects of Reasoning About Knowledge*, pages 325–340, Morgan Kaufmann, San Francisco, California, 1986.

[21] G Lakemeyer, "A computationally attractive first-order logic of belief", in *Proceedings of JELIA 90*, pages 333–347, Springer, Heidelberg, 1990.

[22] H. J. Levesque, "A logic of implicit and explicit belief", in *National Conference on Artificial Intelligence*, pages 198–202, 1984.

[23] A. Meinong, "Über gegenstandstheorie", in A. Meinong (ed.), *Untersuchungen zur Gegenstadstheorie und Psychologie*, Barth, Leipzig, 1904.

[24] R. Montague, "Universal grammar", *Theoria* 36 (1970), 373–98.

[25] Charles Sanders Peirce, *Reasoning and the Logic of Things: The Cambridge Conferences Lectures of 1898*, Harvard University Press, Cambridge Mass., 1992.

[26] G. Priest, *In Contradiction: A Study of the Transconsistent*, Martinus Nijhoff, Dordrecht, 1987.

[27] G. Priest, *Towards Non-Being*, Clarendon Press, Oxford, 2005.

[28] G. Priest, R. Routley, and J. Norman (eds.), *Paraconsistent Logic: Essays on the Inconsistent*, Philosophia Verlag, München, 1989.

[29] V. Rantala, "Urn models", *Journal of Philosophical Logic* 4 (1975), 455–474.

[30] R. Routley, *Exploring Meinong's Jungle and Beyond*, RSSS, Australian National University, Canberra, 1980.

[31] R. Stalnaker, "The problem of logical omniscience I", *Synthese* 89 (1991), 425–440.

[32] R. Stalnaker, "On logics of knowledge and belief", *Philosophical Studies* 128(1) (2006), 169–199.

[33] W. van der Hoek, B. van Linder, and J.-J. Ch. Meyer, "An integrated modal approach to rational agents", in M. Wooldridge and A. Rao (eds.), *Foundations of Rational Agency*, pages 133–168, Kluwer Academic, Dordrecht, 1999.

[34] M.Y. Vardi, "On epistemic logic and logical omniscience", in J.Y. Halpern (ed.), *Theoretical Aspects of Reasoning about Knowledge: proceedings of the First Conference*. Morgan Kaufman, 1986.

[35] H. Wansing, "A general possible worlds framework for reasoning about knowledge and belief", *Studia Logica* 49(4) (1990).

[36] Timothy Williamson, "Conceptual truth", *The Aristotelian Society supplementary volume* 80 (2006).

[37] Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, Kegan Paul, 1922.

[38] Edward N. Zalta, *Abstract Objects: An Introduction to Axiomatic Metaphysics*, D. Reidel, Dordrecht, 1983.

[39] Edward N. Zalta, "A classically-based theory of impossible worlds", *Notre Dame Journal of Formal Logic* 38(4) (1997), 640–660.

Mark Jago
Department of Philosophy,
University of Nottingham,
University Park,
Nottingham, NG7 2RD,
UK
Mark.Jago@nottingham.ac.uk