The Problem of Rational Knowledge

Mark Jago

Forthcoming in *Erkenntnis*. Draft of April 2013.

Abstract: Real-world agents do not know all consequences of what they know. But we are reluctant to say that a rational agent can fail to know some trivial consequence of what she knows. Since every consequence of what she knows can be reached via chains of trivial consequences of what she knows, we have a paradox. I argue that the problem cannot be dismissed easily, as some have attempted to do. Rather, a solution must give adequate weight to the normative requirements on rational agents' epistemic states, without treating those agents as mathematically ideal reasoners. I'll argue that agents *can* fail to know trivial consequences of what they know, but never *determinately*. Such cases are *epistemic oversights* on behalf of the agent in question, and the facts about epistemic oversights are always indeterminate facts. As a result, we are never in a position to assert that such-and-such constitutes an epistemic oversight for agent *i* (for we may rationally assert only determinate truths). I then develop formal epistemic models according to which epistemic accessibility relations are vague. Given these models, we can show that epistemic oversights always concern indeterminate cases of knowledge.

Keywords: Bounded rationality, logical omniscience, epistemic possibility, knowledge, rationality

1 Introduction

Here are some of the things I know. It's currently sunny outside; Billy Bragg is on the stereo; if it's warm this afternoon the washing will dry; it's warm this afternoon. Do I know that the washing will dry? Of course! Do I know it's sunny and Billy Bragg is on the stereo? Of course! It's tempting to systematise these remarks by claiming that what I know is closed under trivial consequence (including *conjunction introduction* and *modus ponens*). If knowledge is so closed, then that must be in part due to the meanings of the logical constants 'and' and 'if … then'. It must be that the relationship between those meanings and inference rules such as *conjunction introduction* and *modus ponens* dictates that one's knowledge is closed under those rules.

If so, then one's knowledge must be closed under all the standard introduction and elimination rules. For the standard rules for a connective stand to the meaning of that connective just as *conjunction introduction* and *modus ponens* stand to the meaning of 'and' and 'if … then'. And, since those rules taken together form a classically complete system of deduction, one's knowledge must be closed under (first-order) consequence. I must know all logical consequences of what I know and, as a special case, I must automatically know all logical truths, irrespective of whether I've explicitly considered those propositions. Real-world agents are not logically omniscient in this way, of course. Deductive reasoning is typically

cognitively costly, often informative and sometimes surprising. We may know the premises (or have assumed them for the purposes of argument), and yet the conclusions we draw from them are often informative.

It seems that *(i)* rational agents seemingly know the trivial consequences of what they know, but *(ii)* they do not know all logical consequences of what they know. The problem of rational knowledge is that (i) and (ii) are incompatible. Any (first-order) logical consequence of a set of premises is derivable from those premises via a chain of trivial inferences and so, if one does not know some logical consequence of what one knows, then one must fail to know some trivial consequence of what one knows. Let's call such a case an *epistemic oversight*: a particular case in which a given agent fails to know a particular trivial consequence of what she knows. We can then express the problem of rational knowledge as follows: real-world agents must suffer from epistemic oversights, yet we can never attribute a particular epistemic oversight to an agent without thereby treating her as being irrational. (The problem of rational belief is analogous: agents do not believe all consequences of what they believe, yet a rational agent seemingly cannot fail to believe some trivial consequence of what she believes. I'll focus on the knowledge case only here, but almost all of what I say carries over to the doxastic case.)

My aim in this paper is to explore and provide a solution to the problem of rational knowledge. I am not interested merely in the logical omniscience problem, taken as a technical problem for a particular semantics for knowledge. The problem of rational knowledge does not arise merely within a particular formal semantics: it is not an artefact of a particular theoretical framework. (Notice how there was no mention of possible worlds, propositions, deductively closed epistemic possibilities or the like in setting up the problem.) The problem concerns the nature of rational states quite generally. Discussing a similar problem, Stalnaker remarks that:

> The problem does not arise from any easily identifiable philosophical dogma which might be given up to avoid it. … the conclusion really derives not from any substantive assumption about the source of knowledge, but from the abstract concept of content or information. The difficulty is [that] … [w]hether the source of my information is my senses, authority, or a faculty of intellectual intuition with access to a Platonic realm of abstract entities, its deliverances are not news unless they might have been different. (Stalnaker 1984, 25)

Given I know that it's sunny, and I know that Billy Bragg is on the stereo, it isn't news that it's sunny and Billy Bragg is on the stereo. The move from conjuncts (taken individually) to conjunction isn't informative. Since it isn't informative, it seems I can't learn anything new by explicitly reasoning from conjuncts to conjunction; and this could be only because I already know the conjunction. So the problem is quite general, and not the result of a particular semantic account.

The rest of the paper is set out as follows. In §2, I consider and reject several attempts to diffuse the problem of rational knowledge. In §3, I draw a link between the problem of rational knowledge and the sorties which, I claim, sheds light on the

problem (but does not solve it). In §4, I re-cast the problem in terms of epistemic possibility. I argue that the epistemically possible worlds must include logically impossible worlds, but that not all such worlds count as epistemic possibilities. In §5, I consider a serious problem faced by any impossible worlds-based theory of epistemic states (and of content in general). §6 presents formal models based on the philosophical discussion from §4 and §5. §7 establishes some formal results for the models presented in §6, including the main formal result of the paper: an agent cannot determinately fail to know some trivial consequence of what she determinately knows. §8 is a short conclusion.

2 Some Responses Dismissed

In this section, I consider two attempts to diffuse the problem of rational knowledge by arguing that there is no rational requirement for an agent to know trivial consequences of what she knows. The first runs as follows. I know that A only if I believe that A. And I may come to believe that such-and-such, which trivially entails 'A', and yet not realise that 'A' follows. To make this concrete, suppose I'm writing down a proof, from premises I believe. At some stage of the proof, I'm in a position to infer 'A' easily at the next step. But my mind wanders; I don't write 'A'; I don't even consider whether A. According to this line of thought, this kind of mental meander easily explains how I can fail to believe, and hence fail to know, some trivial consequences of what I know.

This response focuses on a fine-grained notion of explicit belief. The view might be put in terms of having sentence tokens (in the language of thought) in the 'belief box'. That picture of the mind is certainly questionable. But I needn't enter into those controversies here, for our ordinary concept of belief extends far beyond this fairly restricted kind of explicit belief. When I make plans for next week, one input to my mental process is my belief that the world won't end before next week; there's also my belief that I won't be abducted by aliens, or drafted for a foreign war, before then. Of course, I don't consider these matters explicitly, for in cognitive terms to do so would be far too costly. Yet I believe all those things (else I wouldn't make those rather dull plans for next week); I believe them implicitly. Indeed, by any reasonable standard, I know all those things.

A good (but fallible) test for implicit belief that A is whether the agent (in sincere, attentive mood) would assent if asked whether A. Belief in trivial consequences of what one believes seems to pass this test. Having inferred 'A' and 'B', one may then fail to make the trivial inference to '$A \wedge B$' explicitly in one's mind but, if asked whether $A \wedge B$, a sincere, attentive agent should assent. Similarly for other trivial inferences. So that fact that one may fail to pay attention to the trivial consequences of what one believes will not help to avoid the problem.

A different way to undermine the idea that a rational agent's beliefs must be closed under trivial consequence is to hold that an agent's doxastic state may be fragmented or divided into clusters. One may consider some matter in one 'frame of mind', forming beliefs in one cluster, and later consider some unrelated matter in some other frame of mind, forming beliefs in some other cluster. Lewis (1982),

Fagin and Halpern (1988) and Stalnaker (1984) present versions of the idea. Lewis motivates his version of it as follows:

> I used to think that Nassau Street ran roughly east-west; that the railroad nearby ran roughly north-south; and that the two were roughly parallel. ... So each sentence in an inconsistent triple was true according to my beliefs, but not everything was true according to my beliefs. ... My system of beliefs was broken into (overlapping) fragments. Different fragments came into action in different situations, and the whole system of beliefs never manifested itself all at once. (Lewis 1982, 436)

We can view each of Lewis's fragments as an individual belief state, so that an agent's total system of beliefs is divided into multiple (and perhaps overlapping) states. On Stalnaker's view, each of an agent's multiple belief states corresponds to a rational dispositional state:

> A person may be disposed, in one kind of context, or with respect to one kind of action, to behave in ways that are correctly explained by one belief state, and at the same time be disposed in another kind of context or with respect to another kind of action to behave in ways that would be explained by a different belief state. This need not be a matter of shifting from one state to another or vacillating between states; the agent might, at the same time, be in two stable belief states, be in two different dispositional states which are displayed in different kinds of situations. (Stalnaker 1984, 83)

An agent is then thought of as believing that A if and only if 'A' is true according to at least one of the agent's belief states. (Fagin and Halpern (1988) give the formal details of this approach.) Then it may be the case that the agent believes that A, and believes that B, but does not believe that $A \wedge B$. This is the case when 'A' is true according to one of the agent's belief states, 'B' is true according to another, but both are true according to none of them. In this way, one may fail to believe trivial consequences of one's beliefs.

I'm quite sure that the kind of situation Lewis describes is commonplace, and that the approaches Lewis and Stalnaker propose are reasonable models of such situations. Yet it is implausible that whenever an agent does not believe some consequence of what she believes, this is so because she hasn't put the relevant premises together. Consider again our agent, trying to complete a complex derivation from some premises $\Gamma = \{`A_1', \ldots, `A_n'\}$ (which, let's suppose, she believes). On the fragments of belief approach, she must either automatically believe all consequences of Γ, or else entertain the 'A_i's in different belief clusters. Both options are untenable. Given the task confronting her, she simultaneously entertains all the 'A_i's in a single frame of mind (look! she's written them all down on her page!). And yet, contrary to the fragments account, she still fails to believe (even implicitly) most of the consequences of Γ.

We can push the objection further. Suppose, having considered all the 'A_i's as individual premises, she then forms (and believes) their conjunction, '$A_1 \wedge \cdots \wedge A_n$'. She must then believe all consequences of Γ, on the fragments of belief approach. The picture is that all the agent's deductive effort goes into combining the premises

into a single belief state, that is, in the relatively trivial deductive move from 'A_1',...,'A_n' to '$A_1 \wedge \cdots \wedge A_n$'. The move from '$A_1 \wedge \cdots \wedge A_n$' to any logical consequence of Γ then comes for free. But this is implausible. Our agent will fail to believe almost all consequences of Γ, irrespective of whether she first forms the conjunction of the premises in Γ.

I've argued that the problem of rational knowledge is a genuine problem which cannot be dismissed easily. In the next section, I offer my own analysis of the problem.

3 Vagueness and Bounded Rationality

The problem of rational knowledge can be formulated in terms of a step-by-step deduction D from premises Γ which the agent in question clearly knows, to a conclusion 'A' that the agent clearly does not know. By assumption, not every step of reasoning in D preserves the agent's knowledge (since we eventually arrive at a conclusion the agent does not know). Yet any attempt to say precisely at which point in the deduction the agent's knowledge gives out is doomed to failure (and not merely because we can't elicit a precise answer from the agent). The problem is one of rationality. Wherever we locate the failure of knowledge in D, we will ascribe to the agent a rational failure, namely the failure to know (even implicitly) some trivial consequence of what she knows. (I'm assuming here, as seems correct, that explicit deduction extends one's knowledge.)

Formulating the problem in this way brings out its structural similarity with the sorites paradox. In this case, the principle that rational agents know the trivial consequences of what they know plays the role that tolerance conditionals play in the sorites. The tolerance conditionals for 'red', for example, say that (in a sorites series of colour samples), if sample n is red then so is sample $n+1$. Clearly, not all such conditionals are true; but we cannot say or discover which is false.

Suppose guests at a party are invited to take some number of sweets from a box. The greedy guests are those who take many sweets, but there is no precise number we can give to answer the question, 'what is the minimum number of sweets a guest could take, and thereby be greedy?' Why can't we give an answer, even in principle? There must be a greatest n for which (in the context of the example) 'taking n sweets is greedy' fails to be fully true. Perhaps for this n (and in that context), 'taking n sweets is greedy' has some truth-value less that full truth; perhaps it is neither true nor false ('there is no fact of the matter'); perhaps it is true on some but not all precisifications of 'greedy'; or perhaps it is straightforwardly false. Regardless, there is a greatest such n, and which n that is (in that particular context) has a lot to do with the meaning of 'greedy'. So, in this sense, the meaning of 'greedy' has some precision to it.

Had things (including our uses of 'greedy') been ever so slightly different, the meaning of 'greedy' (together with the context) may have determined some other least n for which 'taking n sweets is greedy' would fail to be fully true (in that context). But we cannot epistemically discriminate between these possibilities, and so we cannot *know* that $n+1$ is the correct answer to 'what is the minimum

number of sweets a guest could take, and thereby be greedy?'. (We can know that the answer is in the vicinity of n: we know that taking, say, 3 sweets isn't greedy, whereas taking 1,000 is. The point is that we can have *inexact* knowledge only (Williamson 1992).) As a consequence, we may not rationally *assert* the answer to be n. For one should assert only what one knows, or at least, what one has good reason to believe. In asserting 'n is the minimum number of sweets a guest could take and thereby be greedy', one is thereby claiming some evidence in favour of n in answer to the question, but one cannot have such evidence. In all such cases of vagueness, we simply aren't at liberty to make precise claims in borderline cases, *even if* we happen (purely by chance) to hit on the truth. This is the phenomena of *unassertibility at the borderline*.

The situation is very much the same in the case of rational knowledge ascriptions. We cannot rationally assert that it is *this* particular trivial inference in our deduction D which does not preserve the agent's knowledge, even if that is the truth. Given the argument above, there must be some such trivial inference (for a particular context): a cut-off point for 'agent i knows that ...'. My view is that such instances of knowledge failure are always indeterminate instances of knowledge failure. There is never a case in which agent i determinately knows such-and-such, from which it trivially follows that A, such that it is determinate that i does not know that A. So, if 'A' follows trivially from what agent i determinately knows, then either i also knows that A, or else it is indeterminate whether i knows that A.

If this analysis is correct, then our accounts of knowledge should account for the link between trivial inferences and knowledge, so that determinate epistemic oversights (that is, cases in which an agent determinately fails to know some trivial consequence of what she determinately knows) never occur. It is not enough simply to apply whatever one thinks is the correct theory of vagueness to 'trivial inference' and 'knows', for that will not automatically preserve the correct links between the two notions. Our account of knowledge itself needs to guarantee the correct link between trivial inferences and knowledge, so that the theory itself entails that there are no determinate epistemic oversights.

There are various ways of fiddling a formal model of knowledge so that it validates this principle. But by 'writing in by hand' the principle we want to establish, the resulting models take on the appearance of broken furniture held together by gaffa-tape. Not only is this approach unsightly, it is also liable to break sooner or later. So this will not be my approach. Rather, I will look to develop formal models of knowledge with independent motivation (§§4–6). I'll then show (§7) that these models validate the correct link between trivial inferences and indeterminate knowledge. I'll begin, in the next section, by recasting the problem of rational knowledge in terms of epistemic possibility, as this is a key notion in developing adequate epistemic models.

4 Knowledge and Epistemic Possibility

The most successful framework for modelling knowledge is the worlds-based account, which began with Hintikka (1962). Each agent i is assigned an epistemic accessibility relation R_i between worlds. Intuitively, 'R_iwu' means that world u is possible, for all i knows at w. As is well-known (and as Hintikka himself realised), Hintikka's original proposal suffers from the logical omniscience problem. Agents are modelled as automatically knowing all consequences of their knowledge (and hence as automatically knowing all logical truths). As is often remarked, there are many modelling contexts in which this assumption is harmless (Fagin et al. 1995); but clearly, it will not do for a *general* account of knowledge. There are cases in which we want to model the very features of human resource-bounded reasoning which are precluded by the possible-worlds approach.

A popular (and, in my view, the only plausible) approach to the issue is to widen the domain of worlds used in the account, to include some that are not logically possible. Hintikka (1975), Rantala (1975; 1982), Belnap (1977), Levesque (1984), Lakemeyer (1986; 1987) and Fagin et al. (1990) take this approach. One could include worlds analogous to the models of paraconsistent logic, according to which a sentence may be both true and false (Belnap 1977). With such worlds in play, *modus ponens* and *disjunctive syllogism* are invalid (they do not preserve truth-according-to-a-world across all words). This makes it possible to model agents whose knowledge is not closed under classical consequence. (It remains closed under paraconsistent consequence, however. It is moot just how much of a problem this is for the approach.)

On this approach, whenever such a world w is classically impossible, it is trivially impossible: some 'A' is both true and false, according to w. Yet if a world is to enter into epistemic accessibility relations, it must be the kind of world that an agent could consider as a possibility, given what she knows. In considering a world w a possibility, in this epistemic sense, an agent is allowing that the actual world could be like that. But no rational agent considers explicitly contradictory worlds to be such cases.

When one is in the dark about some purported theorem of analysis – one doesn't know whether it's in fact a theorem or not – that's because one doesn't have a proof or other reliable evidence to hand. So one considers both eventualities (the purported theorem is/is not in fact a theorem) as genuine options for how things might in fact be. One thereby considers some mathematically impossible world to be possible (for if the purported theorem is in fact a theorem, then it is so by mathematical necessity; similarly if it is not a theorem). One does not thereby consider some numbers or number-theoretic operators to have contradictory properties! If asked whether some number is the successor of itself, our agent will likely answer, 'of course not!'. If our agent knows that no number is the successor of itself, then she does not consider as epistemically possible any world according to which some number is the successor of itself. Similarly for any other trivial absurdity 'A'. If (as seems reasonable) our agent knows that it is not the case that A, then she does not consider as epistemically possible any world according to which A. But then, the non-classical worlds corresponding to paraconsistent

models cannot play a role in the model of knowledge.

Our problem, therefore, is not merely to avoid logical omniscience in our models of knowledge (which we could do easily by including non-deductively-closed worlds). The problem is to provide a notion of a world which is logically impossible, but not trivially so. Lewis (in complaining about paraconsistent logic) puts the point nicely:

> I'm increasingly convinced that I can and do reason about impossible situations. ... But I don't really understand how that works. Paraconsistent logic ... allows (a limited amount of) reasoning about *blatantly* impossible situations. Whereas what I find myself doing is reasoning about *subtly* impossible situations, and rejecting suppositions that lead fairly to blatant impossibilities. (Lewis 2004, 176)

Lewis allows the theoretical usefulness of 'make-believedly possible impossibilities', whilst noting (correctly) that

> The trouble is that all these uses seem to require a distinction between the subtle ones and the blatant ones (very likely context-dependent, very likely a matter of degree) and that's just what I don't understand. (Lewis 2004, 177)

Hintikka (1975), whilst addressing the logical omniscience problem head-on, makes a similar point. He argues that, for epistemic purposes, impossible worlds must be 'subtly inconsistent' worlds (1975, 478) which 'look possible but which contain hidden contradictions' (1975, 476).

We're in a bind. To avoid logical omniscience, we require impossible worlds; to avoid treating agents as being irrational, we must exclude the trivially impossible worlds. There's obviously no clear way to demarcate the obviously impossible from the subtly impossible worlds.

In previous work (Jago 2009; 2013), my approach to this predicament was to assign relative degrees of impossibility or epistemic implausibility to logically impossible worlds. This approach is one way to flesh out some ideas Chalmers (2010) puts forward concerning (what he calls) *non-ideal epistemic space*: a space of epistemic possibilities which need not all be *a priori* consistent. Chalmers suggests thinking about such possibilities as ones which cannot be ruled out (alternatively, whose negation cannot be known, or proved, or otherwise established) 'through such-and-such amount of *a priori* reasoning' (Chalmers 2010, 44).

If such an approach is to be a rational one, it must interact appropriately with the meanings of the logical constants. My approach was to hold that *(i)* the meanings of the logical constants are associated with inference rules for those constants; but *(ii)* inference rules may establish normative requirements *other* than by closure. They may establish normative requirements in a step-by-step way. By way of example, *rational commitment* is a notion for which the inference rules play a closure role, so that one is rationally committed to all logical consequences of one's rational commitments. But things are otherwise when considering rational epistemic possibility. In this case, the inference rules play a step-by-step normative

role. Explicitly contradictory worlds are a limiting case of epistemic implausibility, for they take no inferential effort to recognise as such. The more inferential steps required to derive an explicit contradiction from what a world represents, the greater that world's degree of epistemic plausibility. Logically consistent worlds are maximally epistemically plausible (even if they are representationally incomplete).

Now for some of the details. I take worlds in general to be ersatz set-theoretic constructions in some *worldmaking* language. That language is constructed in the Lagadonian way (Lewis 1986) with particulars, properties and relations playing the role of names and predicates, interpreted so as to to self-refer. (I take up the difficult question of how to referent non-actual particulars and properties in Jago 2012.) A world w is a pair (w^+, w^-) of sets of worldmaking sentences, where w^+ captures what w represents as being the case, and w^- captures what w represents as failing to be the case. Any such pair counts as a world, so that w^+ and w^- may overlap (producing representational inconsistency) and may fail to be jointly exhaustive (producing representational incompleteness).

Worlds so defined have both a worldly nature (they are ultimately constructed from actual particulars, properties and relations) and a linguistic nature (they have fine-grained linguistic structure). The former feature circumvents the worry that the account is 'purely syntactic'. The latter feature allows us to view a world w as highly structured *sequent*, $w^+ \vdash w^-$. In general, a multi-conclusion sequent $\Gamma \vdash \Delta$ says that the Γs cannot all be true whilst all the Δs are false (Restall 2005; 2008a). Hence a derivable sequent $w^+ \vdash w^-$ tells us that world w is logically impossible. (In this approach, sequents correspond to what Restall (2008a;b) calls an *argument positions*, and a world w where $w^+ \vdash w^-$ corresponds to an *incoherent position*.) Left and right sequent rules, of the form

$$\frac{u^+ \vdash u^-}{w^+ \vdash w^-} \qquad \text{or} \qquad \frac{u^+ \vdash u^- \qquad v^+ \vdash v^-}{w^+ \vdash w^-}$$

can then be seen as relating worlds together in a normative, proof-theoretic way. The former kind of rule can be read as: if u is logically impossible, then so is w. The later kind can be read as: if both u and v are logically impossible, then so is w.

Not any choice of sequent rules is suitable for our current purpose. Relationships between worlds must be dictated purely by the meanings of the logical constants, and so we should not employ any sequent rules that are not directly dictated by the meaning of some logical constant. In general, a sequent calculus employs *logical* rules to fix the meaning of the logical constants, and *structural* rules to give syntactical information about manipulating sequents. This strongly suggests that we should confine the target system to logical rules only. Moreover, since worlds are by definition pairs of sets of sentences, we require logical rules which operate on sets, rather than lists or multisets, of sentences. Given these desiderata, Kleene's system G4 (Kleene 2002), without the cut rule, is an ideal choice. This system has just logical rules plus identity axioms of the form $\Gamma \cup \{A\} \vdash \Delta \cup \{A\}$, and is sound and complete for classical semantics. Since there is a tight relationship between the meanings of the logical constants and each of these sequent rules,

we are entitled to think of the resulting proof structures on worlds as normative, rational and meaning-governed structures.

Such a structure is a *proof* when $u^+ \cap u^- \neq \varnothing$ for each leaf node $u^+ \vdash u^-$. It is a proof of the sequent $w^+ \vdash w^-$ at its root. The size of the smallest proof of $w^+ \vdash w^-$ tells us just how explicit contradictions in world w are. The larger the structure, the more deeply buried are those contradictions and so the more epistemically plausible w is. The *rank* of a world encodes this feature: w's rank $\#w$ is the size (the number of nodes) of the smallest proof of $w^+ \vdash w^-$, if there is one, and ω (the first limit ordinal) otherwise. In general, w is at least as epistemically plausible as u when $\#w \geq \#u$. Given what I said above, this notion of epistemic plausibility is a normative, rational and meaning-governed matter; it is not intended to capture a psychological notion of obvious triviality (or obvious inconsistency).

This scalable notion of epistemic plausibility stands to epistemic possibility as height stands to tallness. There is no clear (or determinate, or definite) value of epistemic plausibility that captures all and only those worlds that are genuinely epistemically possible, just as there is no clear, determinate height which captures all and only the tall people. In each case, just where the borderline falls is indeterminate, unknowable and highly dependent on context. Nevertheless, there are clear cases of epistemic possibility and clear cases of epistemic impossibility, and we can theorise accordingly. When (and only when) a world is epistemically possible, I'll call it an *epistemic scenario*. It is then an indeterminate matter just which worlds count as epistemic scenarios.

We now have a handle (albeit an indeterminate one!) on which worlds count as epistemic scenarios and so we are on our way to building suitable models of knowledge. To do so, we must say which scenarios are epistemically accessible to which agents. This is the topic of the next section.

5 Epistemic Accessibility and Epistemic Oversight

To obtain models of knowledge, given a set of epistemic scenarios, we impose epistemic accessibility relations on those worlds. In the traditional approach to such models, on which the worlds are all logically possible, this is a simple matter. Accessibility relations are primitive characteristics of the model: in specifying the model, we associate one such relation R_i with each agent i whose knowledge we are modelling. In our case, however, there is a further subtlety we must consider.

Suppose you consider a hypothetical situation according to which it's windy and snowing outside. Does that scenario represent that it's snowing outside? The answer seems to be: of course! It is but a short step from here to conclude that, if a scenario represents that $A \wedge B$, then it represents both that A and that B, and vice versa. (After all, there's nothing special, from the point of view of how scenarios represent, about the content of 'it's windy' and 'it's raining'.) And it is but a short step from *this* principle – that what scenarios represent is closed under *conjunction introduction/elimination* – to the total deductive closure of what scenarios represent. After all, *conjunction introduction/elimination* stand to the (conjunctive) meaning of 'and' just as the other introduction and elimination rules

stand to the meanings of the other logical connectives. Yet, as we have already seen, what scenarios represent is *not* deductively closed. So we must resist this plausible chain of reasoning somewhere. The point is not so much that we must employ scenarios which are not deductively closed, but rather that *however* we do so will seem counterintuitive, just as drawing some precise cut-off between red and non-red colour samples does. (Bjerring (2011) uses this point to argue that we *cannot* formulate a non-ideal, non-trivial notion of epistemic space. I disagree.)

Our situation is this. We must use scenarios which fail to represent some trivial consequence of what they represent, such as representing that it's windy and snowing without representing that it's snowing. That's counterintuitive, because if we explicitly describe a scenario as one according to which it's windy and snowing, it seems clear that the scenario in question also represents that it's snowing. We should note that all such cases concern a trivial consequence '*A*' of what a scenario represents as being the case, on which the scenario remains silent. The scenario in question neither represents that *A* is the case nor that *A* is not the case. '*A*' is a representation-gap for that scenario. For if the scenario represented that *A* is not the case (where '*A*' is a trivial consequence of what that scenario represents as being the case), then the scenario would be trivially contradictory, and hence not a scenario after all. So we are considering cases in which a scenario is silent about some situation which, intuitively, it should represent as being the case.

We can explain the tension in one of two ways. On the first way, it's indeterminate what the scenario in question represents. The indeterminacy is in the *represents* relation, holding between scenarios and contents. On the second explanation, it's indeterminate which scenario is the scenario in question. The indeterminacy is in an agent's epistemic access to scenarios.

Each explanation will resolve the tension because, when an agent describes a hypothetical scenario, she describes its determinate features only (determinate with respect to which scenario is under consideration and to what it represents). If it's determinate that the scenario in question represents that it's windy and snowing, then that scenario does not suffer a representational failure in the deductive neighbourhood of 'it's windy and snowing'. That neighbourhood includes both 'it's windy' and 'it's snowing'. So, whenever an agent describes some scenario, it appears to be closed under trivial inference. This is compatible with scenario-representation failing some trivial inference, so long as the failure is not a determinate failure. This is just the picture I have in mind: it is never determinate that the situation in question fails to represent some trivial consequence of what it represents.

Which of the two explanations should we adopt? If we're thinking about representational entities such as novels, or mental imaginings, then it's quite likely that the former explanation is most accurate. But for our purposes in building a model of epistemic notions, the second explanation seems more appropriate. So I will continue to treat scenarios as pairs of sets of worldmaking sentences, and treat the indeterminacy just discussed as indeterminacy in an agent's epistemic accessibility relation.

In what follows, I'll assume that, for each agent, one such relation specifies (sharply, precisely) which worlds are accessible from which. But associated with this accessibility relation is a family of alternative accessibility relations. What is

true relative to all of these relations is determinately true; what is true relative to some but not all of these relations is indeterminate. I'll develop formal epistemic models along these lines in the next section.

6 Epistemic Models

The aim in this section is to build formal models which capture the ideas discussed in §4 and §5. First, as we want to reason about determinate and indeterminate knowledge, we expand the object language $\mathcal{L}$ (over primitive sentences $\mathcal{P}$) to include a determinacy operator '$\triangle$'. '$\triangle A$' means 'it is determinate that A is the case'. We can introduce an indeterminacy operator '$\triangledown$' by definition:

$$\triangledown A \ =_{\text{df}} \ \neg\triangle A \wedge \neg\triangle\neg A$$

with '$\triangledown A$' read as 'it is indeterminate whether A is the case'. In the model theory, states of knowledge will be captured using epistemic projection functions f_i, rather than accessibility relations R_i. Each function $f_i w$ gives the set of worlds that are epistemically accessible from world w for agent i. The projection function f_i captures the accessibility relation R_i such that: $R_i wu$ iff $u \in f_i w$. Working with projection functions is merely a convenient notational change.

In §4, I defined worlds as pairs of sets of sentences, (w^+, w^-). I'll adopt a more general approach in the formal models, which will involve arbitrary sets of entities W^P and W^I, thought of as sets of possible and impossible worlds (but which can be any sets we like). I'll follow standard practise and use labelling functions V^+ and V^-, with each assigning a set of sentences to each world. We think of V^+w as the set of sentences which are true according to world w, and V^-w as the set of sentences which are false according to w. Thus, the sets V^+w and V^-w correspond to the sets w^+ and w^- from §4, except that V^+w and V^-w are sets of object language sentences (which need not be the special worldmaking language from §4).

Relative to $\mathcal{L}$ (which, I am assuming, is fixed throughout), we define models and related notions as follows.

Definition 1 (Epistemic model) *Let W^P and W^I be arbitrary sets (thought of as sets of possible and impossible worlds, respectively), and let $W^\cup = W^P \cup W^I$. An epistemic model for k agents is a tuple $M = \langle W^P, W^I, V^+, V^-, f_1, \ldots, f_k\rangle$, where $V^+ : W^\cup \longrightarrow 2^\mathcal{L}$ and $V^- : W^\cup \longrightarrow 2^\mathcal{L}$ are labelling functions, such that $V^+w \cup V^-w = \mathcal{P}$ when $w \in W^P$, and each $f_i : W^\cup \longrightarrow 2^{W^\cup}$ is an epistemic projection function.*

We carry over our definition of the *rank* of a world $\#w$ from §4:

Definition 2 (Rank) *The rank function $\# : W^\cup \longrightarrow \mathbb{Z}^+ \cup \{\omega\}$ assigns a rank to each world. $\#w$ is the size (number of nodes) in the shortest proof structure rooted at $V^+w \vdash V^-w$, if there is one, and ω otherwise. The rank of model M is $\min\{\#w \mid w \in W^\cup\}$.*

To capture the idea that states of knowledge may be indeterminate, we use the idea from §5 that each accessibility function comes with a family of alternative accessibility functions. (Together, these act as something like precisifications of the agent's epistemic states.) Given any accessibility projection function f_i in M and any sentence $A \in \mathcal{L}$, we define f_i^A, the A-variant of f_i, as follows:

Definition 3 (***A*-variant of f_i**)

$$f_i^A w = \begin{cases} (f_i w \cap \{w \mid A \in V^+ w\}) \cup (f_i w \cap W^P) & \text{if } f_i w \subseteq \{w \mid A \notin V^- w\} \\ f_i w & \text{otherwise} \end{cases}$$

Let $f_i^{\mathcal{L}} = \{f_i\} \cup \{f_i^A \mid A \in \mathcal{L}\}$.

Definition 4 (**Alternative sequences**) *For an epistemic model M as above, let* $\alpha_M = \{\langle g_1 \cdots g_k \rangle \mid g_i \in f_i^{\mathcal{L}}, i \leq k\}$.

I will use the notation '$\vec{g}$' to denote alternative sequences (i.e., sequences of epistemic projection functions) and '$\vec{g}^i$' to denote the ith function of the sequence $\vec{g}$. Thus, if $\vec{g} = \langle g_i \cdots g_k \rangle$ then '$\vec{g}^i$' denotes function g_i. We can think of each alternative sequence $\vec{g} \in \alpha_M$ as a sharpening of 'epistemic accessibility' for our k agents. On this way of thinking, models are defined relative to a particular sharpening $\langle f_1 \cdots f_k \rangle$, from which all the others in α_M are generated. We think of $\langle f_1 \cdots f_k \rangle$ as the sharpening that gets things right: it tells us what's true (*simpliciter*) in that model, whereas what's determinately true in that model is a matter of what is true according to all alternatives in α_M. This will allow us to give a classical definition of truth, on which $M \Vdash A$ or $M \Vdash \neg A$ for all (pointed) models M. (Alternatively, we could give a supervaluationist-style treatment by defining models relative to a set of alternative sequences, and define truth (*simpliciter*) in that model as truth on all alternatives. It would still be the case that $M \Vdash A \vee \neg A$ but not always that $M \Vdash A$ or $M \Vdash \neg A$. Theorem 3, the main result of the paper, would still hold.)

Definition 5 (**$\vec{g}$-truth and $\vec{g}$-falsity**) *Given an epistemic model M as above and an alternative sequence $\vec{g} \in \alpha_M$, we define $\vec{g}$-relative truth and falsity in M, $\Vdash_{\vec{g}}$ and $\dashv\Vdash_{\vec{g}}$, as follows. (M is implicit in each clause.) For possible worlds $w \in W^P$:*

$$\begin{aligned}
w \Vdash_{\vec{g}} p \quad & \textit{iff} \quad p \in V^+ w \\
w \Vdash_{\vec{g}} \neg A \quad & \textit{iff} \quad w \nVdash_{\vec{g}} A \\
w \Vdash_{\vec{g}} A \wedge B \quad & \textit{iff} \quad w \Vdash_{\vec{g}} A \textit{ and } w \Vdash_{\vec{g}} B \\
w \Vdash_{\vec{g}} A \vee B \quad & \textit{iff} \quad w \Vdash_{\vec{g}} A \textit{ or } w \Vdash_{\vec{g}} B \\
w \Vdash_{\vec{g}} A \rightarrow B \quad & \textit{iff} \quad w \nVdash_{\vec{g}} A \textit{ or } w \Vdash_{\vec{g}} B \\
w \Vdash_{\vec{g}} K_i A \quad & \textit{iff} \quad u \Vdash_{\vec{g}} A \textit{ for all } u \in \vec{g}^i w \\
w \Vdash_{\vec{g}} \triangle A \quad & \textit{iff} \quad w \Vdash_{\vec{h}} A \textit{ for all } \vec{h} \in \alpha_M \\
w \dashv\Vdash_{\vec{g}} A \quad & \textit{iff} \quad w \nVdash_{\vec{g}} A
\end{aligned}$$

For impossible worlds $w \in \mathbb{W}^I$:

$$w \Vdash_{\vec{g}} A \quad \textit{iff} \quad A \in V^+w$$
$$w \dashv\Vdash_{\vec{g}} A \quad \textit{iff} \quad A \in V^-w$$

Definition 6 (***n*-entailment**) *A pointed model is a pair $\langle M, w\rangle$ where M is as above and $w \in \mathbb{W}^P$ in M. I'll use the notation 'M^w' to abbreviate '$\langle M, w\rangle$'. We define truth relative to M^w as:*

$$M^w \Vdash A \quad \textit{iff} \quad w, \langle f_1, \ldots, f_n\rangle \Vdash A$$

$M^w \Vdash \Gamma$ iff $M^w \Vdash A$ for each $A \in \Gamma$. For any $n \in \mathbb{Z}^+ \cup \{\omega\}$, logical n-entailment is then defined as:

$\Gamma \vDash_n A$ *iff, for every pointed model M^w of rank $r \geq n$, $M \Vdash \Gamma$ only if $M^w \Vdash A$*

It is then easy to see that $\vDash_n$ extends classical entailment:

Theorem 1 *For any $n \in \mathbb{Z}^+ \cup \{\omega\}$: if Γ classically entails A, then $\Gamma \vDash_n A$.*

Proof: Suppose $\Gamma \nvDash_n A$. Then there is an epistemic model M, world $w \in \mathbb{W}^P$ in M and alternative $\vec{g} \in \alpha_M$ such that $w \Vdash_{\vec{g}} B$ for all $B \in \Gamma$ but $w \nVdash_{\vec{g}} A$. Let $v : \mathcal{P} \longrightarrow \{\textit{true}, \textit{false}\}$ be a classical valuation such that $vp = \textit{true}$ iff $p \in V^+w$ in M. Now extend v for '$\neg$', '$\wedge$', '$\vee$' and '$\rightarrow$' in the usual way. We show that vC satisfies each $C \in \Gamma \cup \{\neg A\}$ by induction on the complexity of C. If $C := p$, then $p \in V^+w$ hence $vp = \textit{true}$. Now assume that, for all $C' \in \Gamma \cup \{\neg A\}$ of lower complexity than C, $vC' = \textit{true}$ iff $w \Vdash_{\vec{g}} C'$. If $C := \neg C_1$ then $w \Vdash_{\vec{g}} \neg C_1$, hence $w \nVdash_{\vec{g}} C_1$ and, by hypothesis, $vC_1 = \textit{false}$. Then $vC = \textit{true}$. If $C := C_1 \wedge C_2$, then $w \Vdash_{\vec{g}} C_1$ and $w \Vdash_{\vec{g}} C_2$ and, by hypothesis, $vC_1 = vC_2 = \textit{true}$, hence $v(C_1 \wedge C_2) = \textit{true}$. The '$\vee$' and '$\rightarrow$' cases are similar. Hence $\Gamma \nvDash_n A$ only if $\Gamma \cup \{\neg A\}$ is classically satisfiable, and theorem 1 follows by contraposition. ■

With the epistemic models we need defined, we now need to check that they guarantee the correct contention between trivial inference and (in)determinate knowledge. This is the topic of the next section.

7 Trivial Inference and Indeterminate Knowledge

In this section, I'll first return to the notion of trivial inference (and trivial consequence). As with the notions of epistemic plausibility and possibility, this is intended as a normative, rational notion, not a psychological one. I'll then show that, with respect to this normative notion and with respect to the epistemic models just defined, any instance in which an agent fails to know some trivial consequence of what she knows is an indeterminate instance of knowledge failure. Agents never determinately fail to know trivial consequences of what they determinately know.

Relative to any precise delineation n of which worlds count as epistemic scenarios (i.e., such that all and only worlds of rank $\#w \geq n$ count as epistemic scenarios), we define A as a trivial consequence of Γ, $\text{triv}_n(\Gamma, A)$ as follows.

Definition 7 (Trivial consequence) *With respect to any integer $n \in \mathbb{Z}^+ \cup \{\omega\}$, $A \in \mathcal{L}$ is a trivial consequence of $\Gamma \subseteq \mathcal{L}$,* $\mathrm{triv}_n(\Gamma, A)$*, if and only if, for all pointed models M^w of rank $r > n$, $M^w \Vdash \Gamma$ only if $M^w \nVdash A$.*

As a definition of (a kind of) consequence, this definition is rather unusual: the clause has '$M^w \nVdash A$' where we would usually have '$M^w \Vdash A$'. This is because trivial consequence is not purely about truth-preservation across all epistemic scenarios. In fact, *no* inference (other than *identity*, $A \vdash A$) is preserved across all epistemic scenarios. Rather, a consequence counts as trivial in the current sense when the truth of the premises guarantee *falsity avoidance* for the conclusion across all epistemic scenarios. So, $\mathrm{triv}_n(\Gamma, A)$ behaves as a consequence relation in some ways, but not in others, as the following results highlight.

Theorem 2 triv_n *has the following properties, for all $n \geq 1 \in \mathbb{Z}^+ \cup \{\omega\}$:*

(a) $\mathrm{triv}_n \subseteq \mathrm{triv}_{n+1}$*: if* $\mathrm{triv}_n(\Gamma, A)$*, then* $\mathrm{triv}_{n+1}(\Gamma, A)$.

(b) triv_n *is monotonic: if* $\mathrm{triv}_n(\Gamma, A)$ *and* $\Gamma \subseteq \Delta$ *then* $\mathrm{triv}_n(\Delta, A)$.

(c) $\mathrm{triv}_n(\Gamma, A)$ *only if* Γ *classically entails* A.

(d) triv_n *is reflexive.*

(e) $\mathrm{triv}_1(\Gamma, A)$ *if and only if* $A \in \Gamma$.

(f) For $n > 1$, triv_n *is non-transitive and does not satisfy cut: it is not the case that if* $\mathrm{triv}_n(\Gamma, A)$ *and* $\mathrm{triv}_n(\Gamma \cup \{A\}, B)$ *then* $\mathrm{triv}_n(\Gamma, B)$.

Proof: For (a), suppose that $\mathrm{triv}_n(\Gamma, A)$ and that pointed model M has a rank $r > n+1$ s.t. $M \Vdash \Gamma$. Then for all M' of rank $r > n$ s.t. $M' \Vdash \Gamma$, $M' \nVdash A$. Hence $M \nVdash A$, and so $\mathrm{triv}_{n+1}(\Gamma, A)$. For (b), suppose $\mathrm{triv}_n(\Gamma, A)$, $\Gamma \subseteq \Delta$ and $M \Vdash \Delta$. Then $M \Vdash \Gamma$ and so, by definition, $M \nVdash A$. Hence $\mathrm{triv}_n(\Delta, A)$. For (c), suppose Γ does not classically entail A. Then there is no closed tree in our sequent system with $\Gamma \vdash A$ at its root. Let $M^w = \langle \{w\}, \varnothing, V^+, V^-, \mathrm{Id}_1, \ldots, \mathrm{Id}_k \rangle$ where $V^+w = \Gamma$, $V^-w = \{A\}$ and the Id_is are identity functions, so that $M^w \Vdash \Gamma$ and $M^w \dashv\Vdash A$. Since $\#w = \omega$, M^w has rank ω and hence $\neg\mathrm{triv}_n(\Gamma, A)$ for any $n \leq \omega$. (c) follows by contraposition.

For (d), first suppose that M has rank $r > 1$. Then there is no $w \in W^\cup$ in M such that $A \in V^+w \cap V^-w$ (since for any such world, $\#w = 1$). Hence if $M \Vdash A$ then $M \nVdash A$, and so $\mathrm{triv}_1(\{A\}, A)$. By (b), if $A \in \Gamma$ then $\mathrm{triv}_1(\Gamma, A)$ and, by (a), $\mathrm{triv}_n(\Gamma, A)$ for any $n \in \mathbb{Z}^+$, hence each triv_n is reflexive. For (e), the 'if' follows from reflexivity. For the 'only if', suppose $A \notin \Gamma$ and let $M^w = \langle \{w\}, \varnothing, V^+, V^-, \mathrm{Id}_1, \ldots, \mathrm{Id}_k \rangle$ where $V^+w = \Gamma$ and $V^-w = \{A\}$. Then $M^w \Vdash \Gamma$ and $M^w \nVdash A$. Since $A \notin \Gamma$, $\#w > 1$, hence M^w has a rank $r > 1$, hence $\neg\mathrm{triv}_1(\Gamma, A)$.

For (f), note that triv_1 trivially satisfies cut. So suppose $n > 1$ and let $\Gamma_n = \{p_1 \wedge (p_2 \wedge (\cdots \wedge p_n) \cdots)\}$. Then $\mathrm{triv}_n(\Gamma, p_{n-1} \wedge p_n)$ and $\mathrm{triv}_n(\Gamma \cup \{p_{n-1} \wedge p_n\}, p_n)$, but $\neg\mathrm{triv}_n(\Gamma, p_n)$. For a counterexample, let $M^w = \langle \{w\}, \varnothing, V^+, V^-, \mathrm{Id}_1, \ldots, \mathrm{Id}_k \rangle$ where $V^+w = \Gamma$ and $V^-w = \{p_n\}$. We have $M^w \Vdash \Gamma$ and $M^w \dashv\Vdash p_n$. Since

$\#w = n+1$, M^w has rank $n+1$ and so $\neg\mathrm{triv}_n(\Gamma, p_n)$. Hence, for $n > 1$, triv_n does not satisfy cut. Since cut is a form of transitivity, the argument that triv_n is non-transitive is similar. ■

So long as n is not too small, the trivial consequences (so defined) include all the inferences we usually call trivial. For example, we have:

$$\mathrm{triv}_2(\{A \wedge B\}, A) \qquad \mathrm{triv}_3(\{A, B\}, A \wedge B)$$
$$\mathrm{triv}_2(\{A\}, A \vee B) \qquad \mathrm{triv}_4(\{A \vee B, \neg A\}, B)$$
$$\mathrm{triv}_3(\{A \to B, A\}, B) \qquad \mathrm{triv}_5(\{A \to B, \neg B\}, \neg A)$$
$$\mathrm{triv}_7(\{\neg(A \wedge B)\}, \neg A \vee \neg B) \qquad \mathrm{triv}_7(\{\neg(A \vee B)\}, \neg A \wedge \neg B)$$

and so on. These are the minimal n-values for which $\mathrm{triv}_n(-,-)$ holds. For example, $\neg\mathrm{triv}_3(\{A \vee B, \neg A\}, B)$. To see why, consider a single-world model M such that $V^+w = \{A \vee B, \neg A\}$ and $V^- = \{B\}$. Then $\#w = 4$ and hence M is of rank 4.

I've given a way to differentiate between trivial and non-trivial valid inferences. I now turn to the main result of the paper, concerning trivial inference, knowledge and (in)determinacy:

Theorem 3 *For any* $n \in \mathbb{Z}^+ \cup \{\omega\}$, *if* $\mathrm{triv}_n(\Gamma, A)$ *then* $\{\triangle K_i B \mid B \in \Gamma\} \models_n \neg\triangle\neg K_i A$.

Proof: Assume that $\mathrm{triv}_n(\Gamma, A)$ and, for all $B \in \Gamma$, $M^w \Vdash \triangle K_i B$, where $M = \langle W^P, W^I, V^+, V^-, \vec{f}\rangle$ and $w \in W^P$. Then for all $B \in \Gamma$ and all $\vec{g} \in \alpha_M$, $w \Vdash_{\vec{g}} K_i B$. Hence $u \Vdash_{\vec{g}} B$, for all for all $u \in \vec{g}_i w$ and all $B \in \Gamma$. Thus, for any $u \in \vec{g}_i w$, $M^u \Vdash \Gamma$ and, given $\mathrm{triv}_n(\Gamma, A)$, $M^u \nVdash A$ and hence $M, u \nVdash_{\vec{g}} A$, for all $u \in \vec{g}_i w$ and all $\vec{g} \in \alpha_M$. This guarantees that $f_i w \subseteq \{v \mid A \notin V^- v\}$ and so, by definition 3:

$$f_i^A w = f_i w \cap \{v \mid A \in V^+ v\} \cup (f_i w \cap W^P)$$

Then, for every world $u \in f_i^A w$ and alternative $\vec{h} \in \alpha_M$ such that $\vec{h}_i = f_i^A$, $M, u \Vdash_{\vec{h}} A$. Since $w \in W^P$, this gives us $M, w \Vdash_{\vec{h}} K_i A$, hence $M, w \nVdash_{\vec{f}} \triangle\neg K_i A$ and so $M, w \Vdash_{\vec{f}} \neg\triangle\neg K_i A$. Then $M^w \Vdash \neg\triangle\neg K_i A$ and so $\{\triangle K_i B \mid B \in \Gamma\} \models_n \neg\triangle\neg K_i A$. ■

Corollary 1 *For any* $n \in \mathbb{Z}^+ \cup \{\omega\}$*:*

(a) If $\mathrm{triv}_n(\Gamma, A)$ *then* $\{\triangle K_i B \mid B \in \Gamma\} \cup \{\neg K_i A\} \models_n \triangledown K_i A$.

(b) If $n \geq 3$, *then* $\models_n \neg\triangle\neg K_i(A \vee \neg A)$ *and* $\neg K_i(A \vee \neg A) \models_n \triangledown K_i(A \vee \neg A)$.

Proof: For (a), suppose $\mathrm{triv}_n(\Gamma, A)$. Then by theorem 3, $\{\triangle K_i B \mid B \in \Gamma\} \models_n \neg\triangle\neg K_i A$ and, since $\triangle B \models_n B$, $\neg K_i A \models_n \neg\triangle K_i A$. Hence $\{\triangle K_i B \mid B \in \Gamma\} \cup \{\neg K_i A\} \models_n \triangle\neg K_i A \wedge \neg\triangle\neg K_i A$ and, by definition of '$\triangledown$', $\{\triangle K_i B \mid B \in \Gamma\} \cup \{\neg K_i A\} \models_n \triangledown K_i A$. For (b), suppose $(A \vee \neg A) \in V^- w$ in some model M^w. Then $\#w \leq 3$ and so M has a rank $r \leq 3$. Hence for any pointed model M^w of rank $r > 3$, $M^w \nVdash A \vee \neg A$ and so, for all $n \geq 3$, $\mathrm{triv}_n(\varnothing, A \vee \neg A)$. By theorem 3, $\models_n \neg\triangle\neg K_i(A \vee \neg A)$. Further, by (a), $\neg K_i(A \vee \neg A) \models_n \triangledown K_i(A \vee \neg A)$. ■

These results are a very pleasing feature of the theory. Theorem 3 then tells us that, however we choose a precise delineation of 'epistemic scenario' and 'trivial consequence', if the inference from Γ to 'A' is trivial then determinate knowledge of Γ entails the agent does not determinately fail to know that A. So if 'A' is a trivial consequence of what agent i knows, then it is never determinate that i fails to know that A. Equivalently (as the corollary says), if agent i does not know some trivial consequence 'A' of what she knows, then it is indeterminate whether she knows that 'A'. So, on the account proposed, there are no determinate epistemic oversights. Equivalently, each case of an epistemic oversight is an indeterminate case.

Since what is indeterminate is not rationally assertible, it is then never rational to assert that agent i suffers from a particular epistemic oversight. If an agent is not logically omniscient, then we can be sure that she suffers from some epistemic oversight. Indeed, it is determinate that real-world agents are not logically omniscient, and hence determinate that real-world agents suffer from epistemic oversights. But we can never say what they are: we cannot locate them in a rational agent's epistemic state. Whenever we focus on a particular trivial consequence 'A' of agent i's knowledge, it is never rational to assert that she does not know that A (even if that's the case). Epistemic oversights are elusive, just as counterexamples to tolerance principles for vague predicates are.

8 Conclusion

For any particular trivial consequence 'A' of what a rational agent knows, it is tempting to say that the agent must also know that A. But we are not likewise tempted to say that agents know all logical consequences of what they know. My aim in this paper has been to explain these seemingly conflicting features of the knowledge of rational (but non-ideal) agents.

Epistemic closure principles are very much like tolerance principles for vague predicates such as 'greedy' in this respect. My approach has been to treat the failure of epistemic closure principles in an analogous way to failures of tolerance principles for vague predicates. In any sorites case, one must of course deny the tolerance principle, but one must also explain *why* the principle seems so reasonable. When a case a is a counterexample to a tolerance principle for a predicate 'F', it is unknowable and hence not assertible that a is F. We can never rationally assert any counterexample to the tolerance principle, and this is how the principle acquires its rational appeal.

In just the same way, to avoid treating real-world agents as being logically omniscient we must deny the epistemic closure principles. Yet we must also explain why they seem so reasonable. My answer is that no counterexample to any *trivial* closure principle is ever rationally assertible. It is false that agents know all trivial consequences of what they know. What is true is that agents never determinately fail to know any trivial consequence of their determinate knowledge. Trivial inferences never take us from clear cases of knowledge to clear cases of knowledge failure. So we can never rationally assert any counterexample to the knowledge

closure principles (§3).

To implement this idea in formal models of knowledge, we require appropriate notions of an epistemic scenario (§4) and of epistemic accessibility (§5). We can then build formal models in a more-or-less standard way (§6). According to these models, there are no determinate instances of epistemic oversights: an agent never determinately fails to know any trivial consequence of what she determinately knows (§7). In this way, the formal approach supports the philosophical contention that epistemic oversights are always elusive.

References

Belnap, N. (1977). A useful four-valued logic, in J. Dunn and G. Epstein (eds), *Modern Use of Multiple-valued Logic*, Reidel, Dordrecht.

Bjerring, J. (2011). Impossible worlds and logical omniscience: an impossibility result, *Synthese*, DOI 10.1007/s11229-011-0038-y.

Chalmers, D. (2010). The nature of epistemic space, in A. Egan and B. Weatherson (eds), *Epistemic Modality*, Oxford University Press.

Fagin, R. and Halpern, J. (1988). Belief, awareness and limited reasoning, *Artificial Intelligence* 34: 39–76.

Fagin, R., Halpern, J., Moses, Y. and Vardi, M. (1995). *Reasoning About Knowledge*, MIT press.

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

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

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

Jago, M. (2009). Logical information and epistemic space. *Synthese*, 167(2):327–341, 2009.

Jago, M. (2012). Constructing worlds. *Synthese*, 189(1):59–74, 2012.

Jago, M. (2013). The content of deduction. *Journal of Philosophical Logic*, 42(2):317–334, 2013.

Kleene, S. C. (2002). *Mathematical logic*, Dover, New York.

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

Lakemeyer, G. (1987). Tractable metareasoning in propositional logic of belief, *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pp. 401–408.

Levesque, H. J. (1984). A logic of implicit and explicit belief, *Proceedings of the Fourth National Conference on Artificial Intelligence*, pp. 198–202.

Lewis, D. (1982). Logic for equivocators, *Noûs* 16(3): 431–441.

Lewis, D. (1986). *On the Plurality of Worlds*, Blackwell, Oxford.

Lewis, D. (2004). Letters to Priest and Beall, in B. Armour-Garb, J. Beall and G. Priest (eds), *The Law of Non-Contradiction—New Philosophical Essays*, Oxford University Press, Oxford, pp. 176–177.

Rantala, V. (1975). Urn models, *Journal of Philosophical Logic* 4: 455–474.

Rantala, V. (1982). Impossible worlds semantics and logical omniscience, *Acta Philosophica Fennica* 35: 18–24.

Restall, G. (2005). Multiple conclusions, *Logic, methodology and philosophy of science: Proceedings of the twelfth international congress*, pp. 189–205.

Restall, G. (2008a). Assertion and denial, commitment and entitlement, and incompatibility (and some consequence), *Studies in Logic* 1: 26–36.

Restall, G. (2008b). Assertion, denial and non-classical theories, *Proceedings of the Fourth World Congress of Paraconsistency, Melbourne*.

Stalnaker, R. (1984). *Inquiry*, MIT Press, Cambridge, MA.

Williamson, T. (1992). Inexact knowledge, *Mind* 101(402): 217–242.