Logique & Analyse **196** (2006), 345–370

IMAGINE THE POSSIBILITIES
INFORMATION WITHOUT OVERLOAD

MARK JAGO

Abstract

Information is often modelled as a set of relevant possibilities, treated as logically possible worlds. However, this has the unintuitive consequence that the logical consequences of an agent's information cannot be informative for that agent. There are many scenarios in which such consequences are clearly informative for the agent in question. Attempts to weaken the logic underlying each possible world are misguided. Instead, I provide a genuinely psychological notion of epistemic possibility and show how it can be used to model static and dynamic information.

1. *Introduction*

I will concentrate on the concept of information that speakers typically associate with episodes of *becoming informed* of some event or state of affairs. There is an intuitive notion to be captured here as, for instance, when I truthfully tell you that you have picked up my laptop instead of your own and you take what I say to be the case. In such cases, I would expect you to check which laptop you have in fact picked up and, on discovering it to be mine, return it to me. Becoming informed, in appropriate circumstances and with appropriate desires (not to upset me, or break the law) may trigger predictable action. Following Quine's and Dennett's line on other cognitive notions, one might even claim that the question of just what information *is*, is not a question of reduction, say to brain processes or symbolic manipulation: "[t]he problem is not one of hidden facts, such as might be uncovered by learning more about the brain physiology of thought processes" [Qui70, p. 180] and as a result intentional idioms, including 'a was informed that p', are "practically indispensable" [Qui60, p. 219].

The kind of information under discussion here might be called *declarative information*, in contrast with procedural or instructive information of the kind we find in an instruction manual or cooking recipe. Declarative information is *alethically qualified* [Flo05a, p. 3], i.e. expressions of declarative information are truth-apt.[1] There are conflicting intuitions concerning declarative information. One is that a set of premises must contain all of the information contained in their consequences. On this view, as Wittgenstein has it, "there can *never* be surprises in logic" [Wit22, §6.1251]. Information is an objective phenomenon, such that there may be information that no one could ever cognize (because, for example, of the total number of fundamental particles that exist in the universe — just pick a propositional tautology containing more propositional letters than this). On the other hand, we have the intuition just mentioned, that becoming informed disposes an agent with appropriate desires to act in a certain way. This is only possible if the agent could, given its cognitive limitations, realize that it was so informed.

The former intuition concerns a static notion of an information state, the latter a dynamic one (or, as Floridi makes the distinction elsewhere in this volume, we have *statal* and *actional* notions of information [Flo06]). Intuitively, the two should be connected by the principle that the dynamic notion of information is no more than the disposition to update one (static) information state to another. There is a notion of information that does not conform to this principle: that of, say, a book containing the information that p, for one cannot inform a book. The difference here is that the sense of information appealed to is not a cognitive sense. Nevertheless, we should only say that a non-cognitive system (such as a book) contains information when it is able, potentially, to effect a change in a cognitive information state (i.e. by being read). It is the cognitive notion of information that I am concerned with here.

2. *Information Update*

One of the key directions in the logical analysis of information is to treat the information that an agent has as the set of all relevant possibilities that she entertains (e.g. [VB03]). If agent a has the information that $p \vee q$, but a's information does not indicate which disjunct is true, then there are three relevant ways the world could be, for all the information that a has. There are worlds in which p is true but q false, worlds in which q is true but p false and worlds in which both are true. Assuming negation behaves classically, we may talk of the three kinds of possibilities as $p\neg q$ worlds, $\neg pq$ worlds and pq worlds. Suppose that the agent is then informed that p is false, in

[1] Perhaps declarative information is not so different from non-declarative information. In order to comprehend the question, 'is it the case that p?', one must first know what information the declarative sentence 'p' conveys. Similarly, one must know what information 'p' contains in order to follow the instruction: 'make it the case that p!'

the sense that a accepts the information as true. Then both the $p\neg q$ and pq worlds are ruled out of a's considerations. The only candidates for how the world could be, given the information that a has, are then the $\neg pq$ worlds. For any agent in a's initial information state, containing the information that $p \vee q$, the additional information that $\neg p$ also contains the information that q, just as we would expect; see figure 1.

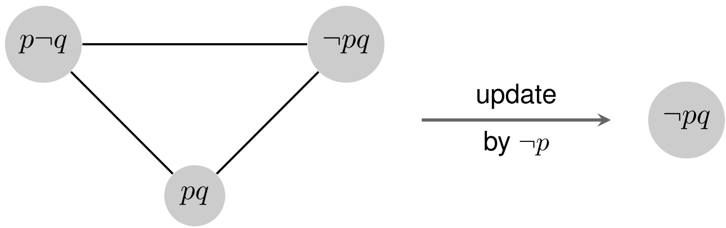


Figure 1. Updating by $\neg p$

As mentioned in the introduction, we have both a static and a dynamic notion of information. We might contrast *being* informed that p, in the sense that the agent in question has the information that p, with *becoming* informed that p, which is an active, dynamic process. Throughout, I will use 'being informed that p' as a synonym for 'having the information that p'. In the former category, we have the *informational state* of agent a at a particular time. This is modelled by saying that the information which a has does not discriminate between certain kinds of worlds: if a has the information that $p \vee q$ but not that p or that q, then a cannot tell whether the actual world is a $\neg pq$ world, a $p\neg q$ world or a pq world. These types of worlds are *epistemically indistinguishable* for agent a, modelled formally as a relation $\sim_a$ between possible worlds, such that $w_1 \sim_a w_2$ means that a's information does not distinguish between w_1 and w_2. If w_1 differs from w_2 in that r is true at the former but not the latter, then $w_1 \sim_a w_2$ implies that a's information state does not include the information that r, or that $\neg r$. An information state, then, is modelled by a class of worlds which the agent cannot distinguish between.

Intuitively, the larger this class of indistinguishable worlds is, the less information the agent possesses. Information is a resource that an agent can use to discriminate the actual world from merely possible ones. Our dynamic sense of information, then, is a narrowing of the class of indistinguishable worlds. Genuine information never excludes the actual world in this process, if it were there to begin with, but misinformation may cause a trusting agent to consider things to be other than they actually are. It is the information contained in a declarative utterance that causes this change in an agent's information state. Therefore, we may model the informational content of the assertion that p as an update on agent a's indistinguishability relation $\sim_a$ such that, after the update, a can distinguish those worlds in which p holds from those in which it does not.

It should be pointed out that the kind of world we have been discussing cannot be the traditional philosophical notion of a metaphysically possible world. For example, it is informative to learn that water is H_2O and yet, since this is a true identity statement, it is a necessary truth. 'Water is H_2O' holds in all possible worlds that contain water (see [Kri80]). But if it is informative to an agent that water is H_2O, they must have previously entertained the possibility of water and H_2O being distinct. It is common to term such possibilities *epistemic possibilities* (e.g. [Hin62]), but it is not often remarked just how different from genuine, metaphysical worlds such possibilities actually are, either in Lewis' sense of genuine concrete entities [Lew86] or Kripke's more parsimonious notion of *ways the world could have been*. I will return to this line of thought in section 5 — suffice to say here that the terminology *worlds* is rather misleading. The logical points that we take to be epistemic possibilities can only be just that — logical points, and hence we can obtain at most a formal model of information. An *account* of what information is remains parasitic on a genuine account of epistemic possibility.

This account of informational content is analogous to the theory of knowledge update. Following Hintikka [Hin62], a static account of knowledge can be given in terms of the worlds that an agent cannot distinguish between. Agent a knows that ϕ in a state s iff ϕ is true at all states s' which a cannot distinguish from s. Gaining new knowledge is thus a matter of restricting indistinguishability between worlds, i.e. of restricting $\sim_a$. Modelling the information contained within the assertion that p is a similar exercise to modelling the change in a's epistemic state when it comes to know that p.[2]

In the remainder of the paper, I will investigate an unintuitive consequence of this framework: an agent cannot be informed about the consequences of its knowledge and logical truths cannot be informative. In section 3, I will argue that this is unacceptable. However, the problem cannot be avoided by weakening the underlying logic (section 4). The problem is not to be located within the analysis of information itself. Rather, the problem arises with a false conception of epistemic possibility, which also plagues epistemic logic. I propose an alternative notion of epistemic possibility in section 5 and show how it results in an improved account of information.

[2] There are, however, a number of differences between information and knowledge, principally that one can be informed that p without knowing that p, even if one holds that information must be true, e.g. [Flo05a]. Stopped clocks may be informative twice a day, but never give rise to knowledge of the time. Those who do not hold that information must be true may make the analogy between informational content and the change in an agent's belief state, rather than its state of knowledge.

3. *Informative Inference*

Let us say that a sentence ϕ is informative for an agent a when an utterance of ϕ could cause a change in a's information state. Now consider the following two cases:

1. Suppose a is informed that $\phi \rightarrow \psi$ and that ϕ. Can ψ then be informative?
2. Suppose $\phi \rightarrow \psi$ is valid. If a is informed that ϕ, can ψ then be informative?

According to the account of *being informed* as an indistinguishability relation $\sim$ on worlds, and of *becoming informed* as an update on $\sim$, the answer to both questions is *no*. In the first case, after becoming informed that $\phi \rightarrow \psi$ and that ϕ, a first excludes all worlds in which $\phi \wedge \neg\psi$ is true, and then excludes worlds in which $\neg\phi$ is true. There only remain worlds at which ψ is true; hence becoming informed that ψ produces no update effect. This is a case of closure under *informed* implication. Suppose an agent has the information that $\phi \rightarrow \psi$. Then being informed that ϕ implies being informed that ψ and becoming informed that ϕ implies becoming informed that ψ. As a consequence, being informed that ϕ is analyzed as exactly the same state as being informed that ψ, and becoming informed that ϕ as the same event as becoming informed that ψ, whenever the agent has the information that $\phi \leftrightarrow \psi$.

In the second case, a is informed that ϕ, so excludes worlds where $\neg\phi$ holds. But since $\phi \rightarrow \psi$ is valid, it holds at all worlds, hence ψ also holds at all worlds which a considers possible. Then updating by ψ produced no change in the worlds that a considers possible, hence ψ has no informative content over and above ϕ. This is a case of closure under *valid* implication. As a consequence, being informed that ϕ is necessarily the same state of information as being informed that ψ, and becoming informed that ϕ is necessarily the same event as becoming informed that ψ, whenever ϕ and ψ are logically equivalent.

This has been termed the problem of *information overload*.[3] If an agent is informed that ϕ, it is also informed of the infinite number of sentences that follow logically from ϕ. Thus the consequences of a set of sentences S contain at most the informational content that the sentences in S themselves contain. The view is very much that the conclusion is contained in the

[3] In the case of knowledge, rather than information, the problem is termed *logical omniscience*. See [Sta91, Whi03] for discussions of this related problem.

premises. So long as we remain within the possible worlds framework, information overload in some form or another cannot be avoided. Both closure under informed implication and under valid implication are present in the weakest normal logic of information, K.[4] In the case of knowledge, rather than information, many find this consequence of the possible worlds framework implausible. Hintikka explicitly says that there are a, ϕ, ψ such that a knows that ϕ, ϕ logically implies ψ and yet a does not know that ψ [Hin75, p. 476]. The question to be discussed, then, is whether the same holds of being and becoming informed.

As a special case of closure under valid implication, this account of information implies that tautologies cannot be informative at all. According to Floridi [Flo05b], "most philosophers agree that tautologies convey no information at all." This is partly because the informativeness of a statement is often linked to how likely that statement is to be true, such that the informativeness of p is inversely related to the subjective probability of p. Thus tautologies, which have a probability of 1, are completely uninformative.[5] Floridi defends this conception elsewhere in this volume, calling tautologies "empty" of informational content: "If the information that p is "empty" … as it is the case of e.g. a tautology … then a can hold the (empty) information that [p], but cannot be informed by receiving it" [Flo06]. Wittgenstein expressed a somewhat similar idea in the *Tractatus* in saying that tautologies literally lack sense (are *sinnlos*). Saying that ϕ logically implies ψ literally says nothing (although it does *show* something, namely that ψ follows from ϕ) [Wit22, §§4 ff]. If one wants to know whether to take an umbrella, it is completely uninformative to be told that either it is raining or it is not.

However, this last example, which seemingly highlights the informational emptiness of tautologies, is a sentence whose tautological nature could be recognized by any competent speaker of the language. But consider the following example. If a sentence is a tautology, then the fact that it is a tautology is also a tautology (of the metalanguage, rather than the object language). The sentence 'ϕ is a tautology' is true precisely when ϕ can be derived in the propositional calculus. It follows that 'ϕ is a tautology'

[4] It is possible to use the weaker Scott-Montague semantics to model information, according to which $\sim$ relates *sets* of worlds, but then one loses the intuition about information update as a restriction of epistemic possibility. Besides, information remains closed under equivalent sentences, itself a form of information overload.

[5] This has the unintuitive consequence that contradictions have maximum informational content. This is known as the Bar-Hillel-Carnap paradox, discussed in [Flo04]. Bar-Hillel and Carnap remark that contradictions are "too informative to be true" [BH64, p. 229]. The problem can be avoided by taking informativeness to imply truth [Flo04], although the elegance of the mathematical model is then lost.

cannot be informative. If true, it is 'empty'; if false, it is misinformation. However, for someone who does not recognize the tautological character of some complicated sentence ϕ, it may well be informative to learn that ϕ is a tautology. A simple example is of a student, sitting a logic exam, asked to say which of the sentences written on the exam paper are tautologies. Given that students frequently get the answer to such questions wrong, our student may certainly find it helpful to have the answers. But how could the answers be *helpful* if they are not *informative*?

In the remainder of this section, I describe several cases that highlight how a consequence ψ of information an agent already possesses can nevertheless be informative. In these scenarios, the only sensible explanation of the agent's behaviour will be: the agent learnt something new and, in so learning, became informed.

Scenario 1 Genuine mathematical theorems are true in all possible worlds, so that discovering a proof for a theorem should not be informative (or rather, it may be informative that one can write a proof in *this* way, but not that the theorem is true). But this is at stark odds with the way mathematicians behave. For example, Andrew Wiles reports a moment in 1986:

> Casually in the middle of a conversation [a] friend told me that Ken Ribet had proved a link between [the] Taniyama-Shimura [hypothesis] and Fermat's Last Theorem. I was electrified. I knew that moment that the course of my life was changing. [Wil06]

What was the source of this electrifying moment? We would say that the cause was the friend's informing Wiles of the link. Wiles gained new information — necessarily true, *a priori* information — which allowed him to continue (and eventually complete) his proof of Fermat's Last Theorem. Lest we be tempted to think that there was really no new information here, here is the analogy that Wiles himself used to describe the process of completing the proof:

> You enter the first room of the mansion and it's completely dark. You stumble around bumping into the furniture but gradually you learn where each piece of furniture is. Finally, after six months or so, you find the light switch, you turn it on, and suddenly it's all illuminated. You can see exactly where you were. [Wil06]

Being able to see objects previously hidden is a paradigmatic case of perceptual information; Wiles' metaphor of illumination explicitly links this type of information acquisition to the psychology of mathematical discovery.

Scenario 2 Early in the summer of 1902, the second volume of Frege's *Grundgesetze der Arithmetik* was in press. In the *Grundgesetze*, Frege sets down his logicist principles and attempts to derive arithmetic from the stable foundations of his logic. Russell's famous letter to Frege of June 16 pointed out that Frege's system was inconsistent. Basic Law V — the abstraction principle, stating that any concept determines a set — had allowed Russell to derive a contradiction similar to the one Burali-Forti had discovered in 1897.[6] Frege immediately began asking questions: "Is it always permissible to speak of the extension of a concept, of a class? And if not, how do we recognize the exceptional cases?" [Fre64, p. 127]. It is evident that Frege's viewpoint had changed completely by 1903. How are we to explain his change of mind? Frege explicitly tells us that his worries were "raised by Mr Russell's communication" [Fre64, p. 127]. We would most naturally say that Russell *informed* Frege of the paradox contained within Basic Law V and that it was *becoming informed* of this that caused Frege to abandon logicism.

Scenario 3 Formal verification *via* model checking is a technique extensively used in industry as a way of checking that certain properties hold of a system at the design stage. A formal model of the system is developed and used to check whether it satisfies a certain property, for example, that two users can never access the same account at the same time, or that the algorithm can never enter a cycle from which it will never exit. Even in seemingly simple systems, the number of possible states of the system can be enormous, which is why a formal tool for checking through all such states is required. It has often been the case that model checking has shown up unexpected flaws in the design, which then has to be rethought. Suppose we have a design that we wish to test and a formal model has been build. We might think that our system can never enter a state at which property ϕ holds. What then is the purpose of model checking whether ϕ is satisfied by the model? Model checking *verifies* that either ϕ holds or does not. It is therefore natural to say that the model checker will output *information* as to whether our design is as safe or reliable as we hope it is. If there is a flaw in our design, the model checker will *inform* us of this.

[6] Russell discovered his paradox in the late spring of 1901 and describes the effect his discovery had on him: "At first I supposed that I should be able to overcome the contradiction quite easily, and that probably there was some trivial error in the reasoning. Gradually, however, it became clear that this was not the case" [Rus69].

All of these scenarios are examples of either case 1 or 2 above. They are cases in which someone is genuinely informed by sentences which, according to the possible worlds account of information discussed above, have no right to be called informative. We can only conclude that there is something wrong with the update model of information.

4. *Avoiding Information Overload*

It is instructive to cast the problem of information overload along the lines of Hintikka's analysis in [Hin75]:[7]

1. 'a is informed that ϕ' is true at w iff ϕ is true at every world indistinguishable from w;
2. There are a, ϕ, ψ such that a is informed that ϕ, ϕ logically implies ψ and yet ψ can be informative for a;
3. A sentence is logically true iff it is true at every logically possible world;
4. All worlds related by $\sim$ are logically possible.

(1–4) are clearly inconsistent; I call this *Hintikka's problem*. In the case of knowledge, Hintikka immediately argues that (2) is not the culprit [Hin75, p. 476], that is, there really are such sentences, so related. Instead, he proposes to reject (4) and claim that not all such 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" [Hin75, p. 476] and hence of information overload. Hintikka's reason for supposing that indistinguishable 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. [Hin75, p. 476]

The worlds that a cannot differentiate between should not be thought of as giving us the possibilities left open by the information that a has. Rather, they should give us the *apparent* possibilities — apparent, that is, given a's ability to follow the logical consequences of the information she has.

[7] Hintikka is concerned with the problem as it arises in the case of knowledge, which I have adapted for the case of information here.

Hintikka devotes the remainder of his article [Hin75, pp. 477–483] to describing *impossible possible worlds*, logical models that are inconsistent from a classical point of view, but "so subtly inconsistent that the inconsistency could not be expected to be known (perceived) by an everyday logician, however competent" [Hin75, p. 478].[8] Suppose an agent considers the sentences satisfied by such a model to state genuine possibilities. That agent will thereby be taking some impossibilities to be possible and, in doing so, will not have all valid sentences in her information state. We therefore have some handle on her logical competence, depending on the degree to which contradictions in the model manifest themselves.

The details of such models are provided by Rantala in [Ran75], where he uses the term *urn models*. The domain is conceived of as a huge urn from which individuals may be drawn (the urn metaphor is taken from elementary probability theory). Sequences of quantifiers embedded one within the scope of another are restrictions on draws from the urn. A classical model is one in which the contents of the urn remains constant between draws; such models are known as *invariant* models. Rantala then considers *changing* models, whose urn has a mechanism attached that can alter the contents from one draw to the next. In this way, sentences that are classically invalid can nevertheless be satisfied by an urn model. The level of inconsistency in an urn model is viewed as the number of draws made before a change in the available individuals. Suppose the largest number of nested quantifiers in a sentence ϕ is d (d is said to be the *depth* of ϕ). If the domain/urn in a model M remains constant for at least the first d draws, M will agree with classical models as to the validity or logical falsehood of ϕ. Such models are called d-invariant.

Hintikka's idea is to use the parameter d as a measure of an agent's logical competency, for sentences with deeply embedded quantifiers are harder to understand than those without. The more competent the agent, therefore, the larger the value of d. An agent whose competency is d will be able to recognize the validity of all valid sentences whose depth does not exceed d, but might get it wrong in the case of more complex sentences. By taking possible worlds to be urn models, the update account can explain how a sentence ϕ with quantifier depth $d' > d$ can be informative to an agent a whose competence is d, even when ϕ follows from information that a already has. There will be worlds that a considers possible at which ϕ is false (these are the d''-invariant models, where $d < d'' \leq d'$) so that, on becoming informed that ϕ, $\sim_a$ is updated to exclude these worlds.

[8] The terminology 'impossible possible worlds' is perhaps not the most advisable. Better suggestions include *nonclassical* in [Cre72, Cre73] and *nonstandard* in [RB79]. Levesque claims a different methodology in [Lev84], using a notion of a *situation* (although Levesque's *situations* are remarkably similar to Cresswell's *nonclassical worlds* [Cre72, Cre73]).

However, for any particular d, an agent's information state either includes *all* or *no* valid sentences of depth d. If its competence is no less than d, then all valid sentences of depth d are 'empty' of information. Assuming our agent has rudimentary logical competency, *all* valid sentences containing no embedded quantifiers, including all propositional tautologies, are empty of information for that agent. Thus neither depth-1 formulae nor propositional tautologies can ever be informative. Moreover, at least *some* complex sentences (say of quantifier depth d) are likely to be informative for an agent, but this should not prohibit the agent from having previously been informed of *any* valid sentence of that depth.[9] Thus, Hintikka's solution does not avoid these unintuitive consequences in the case of information.

There are other approaches which share Hintikka's feeling that (4) is the problematic premise in Hintikka's problem, i.e. epistemic possibilities need not be treated as classical logically possible worlds. They also retain Hintikka's notion that each epistemic possibility must be some kind of logical model, but with a notion of consequence that is weaker than in classical logic. Cresswell [Cre73] describes *nonclassical worlds*, which are essentially based on a paraconsistent logic, where negation behaves in nonstandard ways such that the truth of ϕ does not necessarily exclude the truth of $\neg\phi$. A 4-valued approach to truth underlies Levesque's logic of *explicit belief* [Lev84], which is based on Belnap's 4-valued logic [AB75]. A similar account is given by Fagin, Halpern and Vardi in [FHV90], whose semantics is based on relevant logic (see e.g. [DR02]). What all these approaches have in common is that not all classical tautologies hold at all worlds, allowing the satisfaction clause for 'a is informed that ϕ' to be given in terms of indistinguishable worlds without generating Hintikka's problem. As a consequence, we have an account of worlds that can be used in the update account of information to model an agent genuinely becoming informed about some (classical) consequence of information it already possesses.

However, a version of Hintikka's problem can be generated relative to whatever logic underlies such worlds. Suppose that this logic is Λ. If Λ has infinitely many theorems, the following is likely to be true:

$2'$. There are a, ϕ, ψ such that a has been informed that ϕ, ϕ Λ-entails ψ and yet ψ is informative for a.

Yet, ψ will be true at all such worlds and so ($2'$) comes out false. Substituting ($2'$) for (2) generates a contradiction similar to Hintikka's problem. As a consequence, any agent who has been informed that ϕ cannot possibly be

[9] Similar examples are discussed in [Jag06a].

informed that ψ when ϕ Λ-entails ψ; and any Λ-valid sentence will have no informative content whatsoever.

On reflection, weakening the internal logic of worlds seems a badly motivated move, because it models agents as ideal reasoners in a weaker-than-classical logic. The fact that agents do not suffer from information overload in the real world, on the other hand, is not due to their lacking reasoning principles, as if the agents somehow did not know how to apply *modus ponens* or the law of excluded middle. Rather, agents have bounded resources — time, memory, attention and the like — which limit what an agent can infer from the information it already has.

We should conclude that this notion of epistemic possibility is thoroughly flawed. I develop an alternative conception in the following sections. Before I do, I want to evaluate a rather different approach. Rather than rejecting premise (4) of Hintikka's problem, we might adapt Fagin and Halpern's account in [FH88] and suggest that agents are indeed overloaded with the consequences of the information they possess, but that such consequences are filtered through an 'awareness' filter, thus avoiding the problem in practise. An agent can only use information that it is aware of and hence might consider, wrongly, some consequence of its information to be informative.

Awareness, as discussed in [FH88], is a purely syntactic notion. It is therefore possible to alter the properties of awareness without modifying the underlying possible worlds account of information. We need not specify properties of the awareness set *a priori*, but "[o]nce we have a concrete interpretation in mind, we may want to add some restrictions" [FH88, p. 54]. However, it seems essential to the success of the awareness model that, in general, awareness sets have no closure properties whatsoever. As Fagin and Halpern comment,

> people do *not* necessarily identify formulae such as $\psi \wedge \phi$ and $\phi \wedge \psi$. Order of presentation does seem to matter. And a computer program that can determine whether $\phi \wedge \psi$ follows from some initial premises in time τ might not be able to determine whether $\psi \wedge \phi$ follows from those premises in time τ. [FH88, p. 53, their emphasis]

However, given a concrete formulation of awareness we may ask, why could this notion not be used to define a notion of being informed *directly*, using whatever principles were used to determine the properties of the awareness set? A potential notion of awareness given in [FH88, p. 54] is that the elements of the awareness set are precisely those sentences that the agent *could* determine as consequences of information they already possess in a specified space and/or time bound. This is, roughly, the notion I propose below, although I make no use of the evidently spurious notion of awareness.

5. *Epistemic possibility*

In the introduction, I remarked that epistemic possibilities are unlike metaphysical possibilities in that the latter, but not the former, are captured by appeal to possible worlds as a genuine feature of being. Epistemic possibilities, on the other hand, are psychological notions. In the remainder of this section, I discuss a simple kind of model that captures the notion of *potential ways of reasoning* for an agent with certain abilities and a certain resource bound.

Suppose that our agent reasons using rules of some kind, which can be encoded as a set of step-by-step instructions. These rules may be natural deduction rules, for example, or they may be rewrite rules, or a list of axiom schemes together with rules for their instantiation. I will use the notation

$$\alpha_1, \ldots, \alpha_n \rhd \beta$$

to write the rule: from having inferred each α_1 through to α_n, infer β. The important feature of the rules I consider here, which I call *one-step rules*, is that each requires a set amount of the agent's resources to be applied. A paradigmatic case is that of conjunction introduction: $\alpha, \beta \;\rhd\; \alpha \wedge \beta$. This condition excludes rules that are accompanied by provisos such as: *provided that* α *is consistent*, for consistency checking itself can be computationally expensive. The condition also means that rules that would usually be written with the aid of the '$\cdots$' symbol, meaning *go on in the same way*, cannot be used.[10] This excludes the usual natural deduction rules involving assumptions, such as implication introduction, usually given as:

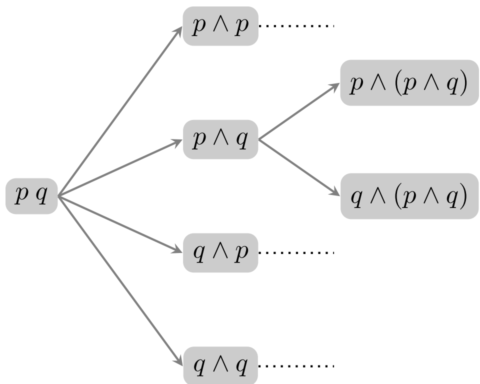
$$\frac{\left|\begin{array}{l}\alpha\\ \vdots\\ \beta\end{array}\right.}{\alpha \to \beta}$$

Here, the vertical line marks the scope of the assumption that α. This rule requires an agent to do several things: firstly, assume α; then reason as normal, within the assumption; then, having derived some formula β, close the assumption and infer the implication. The rule can be encoded as a number of one-step rules as follows. For formulae α inferred within an assumption that γ, write α^γ. At any time, one may infer α^α, so long as α is well-formed: this

[10] The scheme $\alpha, \beta \rhd \alpha \wedge \beta$ from above is not itself a rule, but rather a template for rules. Writing $\alpha_1, \ldots, \alpha_n$ can be understood as shorthand for writing the entire list from α_1 through to α_n in full, so is unproblematic.

rule corresponds to opening the assumption that α (which is clearly a one-step operation). Having inferred β^{α}, one may then infer $\alpha \rightarrow \beta$. Again, this is a one-step operation. Assumptions can be embedded by using sequences of formulae: α^{pq} represents α within the assumption that q, itself within the assumption that p. Additional rules are required to fully encode contextual reasoning.[11] The full details of this approach to encoding assumption-based reasoning can be found in [Jag05, Jag06a]. I offer this brief account here to illustrate how functionally complete, classical reasoning can be encoded using one-step rules.

Given a set of one-step rules, a model of the possible ways of reasoning (using those rules) is built as follows. Models are transition systems, where each transition represents a single one-step rule being applied. The idea is best illustrated diagrammatically. Suppose an agent reasons using conjunction introduction. Then, possible ways of reasoning from p, q include the following (the diagram only shows the newly inferred formulae):



The transition system is just a relational structure: the shaded squares are its states (or elements) and the arrows show the transition relation that holds between them. In this example, each arrow corresponds to applying conjunction introduction but, in general, transitions correspond to applying some (unspecified) one-step rule. Conversely, each possible application of a rule initiates a transition from one state to the next.

[11] For example, if α has been inferred, then α^{β} may also be inferred (one is allowed to use formulae from outside the scope of the assumption within the assumption as well). The other rules must then take account of the assumption-notation. Conjunction introduction, for example, is rewritten as: $\alpha^{\sigma}, \beta^{\sigma} \rhd (\alpha \wedge \beta)^{\sigma}$, where σ is some (possibly empty) sequence of formulae encoding a sequence of assumptions being made. The rule says: having inferred both α and β within the same assumption σ, infer their conjunction $\alpha \wedge \beta$ within the same assumption σ.

These structures can be described using a basic modal language containing $\Diamond$ and its dual $\Box$. I use the brackets '$[-]$' to talk about the formulae that are written within the shaded boxes or, as I shall say from now on, that *label* a state. I shall write $s \Vdash [\alpha]$ to mean that the formula α labels state s (provided that it is clear from the context which model is being talked about). Thus, although the formula $p \wedge q$ does not appear at the root of the model, p does and q does, which we can express as the (metalanguage) formula $\neg[p \wedge q] \wedge [p] \wedge [q]$. Note that $\neg[\alpha]$ is not equivalent to $[\neg\alpha]$. The former means that $\neg\alpha$ labels the current state, whereas the latter means that α does not. It is perfectly possible for neither α nor $\neg\alpha$ to label a state. Note that it is also vital when modelling resource-bounded agents that the formulae that label a state are not deductively closed.

An agent who has been misinformed may infer contradictory sentences, hence both α and $\neg\alpha$ can label a state. Consequently, $[\alpha] \wedge [\neg\alpha]$ is satisfiable. However, $[\alpha] \wedge \neg[\alpha]$ is not satisfiable, since either α labels a state or it does not. This means that negation (and the other Boolean connectives) behave classically apart from when enclosed in brackets. It is best to read the formulae in brackets as a *that*-clause. For example, $[\alpha] \rightarrow [\beta]$ should be read as 'if the agent has inferred α, then it has inferred β as well'. This clearly does not mean the same thing as $[\alpha \rightarrow \beta]$, i.e. 'the agent has inferred that $\alpha \rightarrow \beta$' (and only in the case of ideal reasoners does the latter entail the former).

Let $\mathcal{L}$ be a standard propositional language over a set of primitive propositions $p_1, p_2, \ldots$ and $\mathcal{L}^+$ a propositional modal language over the set $\{[\alpha] \mid \alpha \in \mathcal{L}\}$. (Assume that each language is the smallest such set closed under $\neg$, $\wedge$ and, in the case of $\mathcal{L}^+$ only, $\Diamond$. The other Boolean connectives and $\Box$ can be introduced by definition.) A model M is (at this stage) defined as a tuple $\langle S, T, V\rangle$ where S is a set of states, $T \subseteq S \times S$ is the transition relation on states and V (of type $S \times \mathcal{L} \longrightarrow \{\mathit{true}, \mathit{false}\}$) is the labelling function, assigning a truth-value to each state-formula pair. The support relation $\Vdash$ is then defined recursively, for $\alpha \in \mathcal{L}$ and $\phi, \psi \in \mathcal{L}^+$, as follows:

$M, s \Vdash [\alpha]$ iff $V(s, \alpha) = \mathit{true}$
$M, s \Vdash \neg\phi$ iff $M, s \nVdash \phi$
$M, s \Vdash \phi \wedge \psi$ iff $M, s \Vdash \phi$ and $M, s \Vdash \psi$
$M, s \Vdash \Diamond\phi$ iff there is a state $s' \in S$ such that Tss' and $M, s' \Vdash \phi$

The other Boolean connectives $\vee, \rightarrow, \leftrightarrow$ and the $\Box$ modality can be dealt with by definition.

The particularity of these models comes in the way that T and V interact. As described above, transitions Tsu capture potential applications of some one-step rule, in which the agent infers just one new formula. Thus whenever

Tsu holds, u must be labelled just like s except that, in addition, u is labelled by some additional formula. Here, u is said to *extend* s by that formula. In a model M, whenever a state s may be so extended, there is a state u suitably extending s such that Tsu. Such models are described in detail in [Jag06b, Jag06a].

The agent's resource bounds are then captured by restricting the length of branches in such models. An agent that can only reason for n steps is captured using models in which branches contain at most $n + 1$ states.[12] An agent that can reason for n steps can derive β from $\alpha_1, \ldots, \alpha_n$ iff $[\alpha_1] \wedge \cdots \wedge [\alpha_n] \to \Diamond^n[\beta]$ is valid on the class of models in which the transition relation corresponds to the reasoning abilities of the agent.[13] ($\Diamond^n$ is shorthand for $\Diamond\Diamond \cdots \Diamond$, n times.)

Now, I turn to discussing epistemic possibility. Consider an agent wondering whether this or that is the case. We are not interested in what the agent takes to be a possible state of affairs; rather, we are interested in what, from the agent's point of view and given the information it has, could well be the actual state of affairs. If an agent has the information $\alpha_1, \ldots, \alpha_n$ and cannot infer an explicit contradiction from the set $\{\alpha_1, \ldots, \alpha_n, \beta\}$, given the resources at its disposal, then the agent should consider β as a possibility.[14] We can turn this idea around and say that the epistemic possibilities for an agent are expressed as those sentences whose truth the agent cannot rule out *a priori*, given the other information it has and given its resource bounded nature.

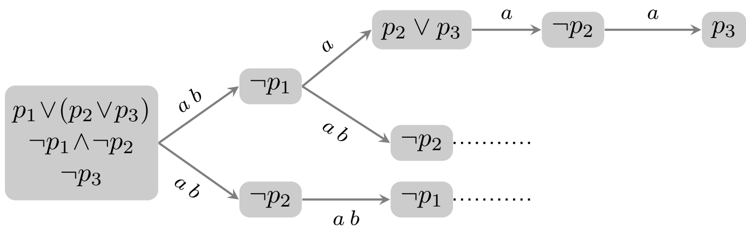
By way of example, suppose that a weak chess player is about to make a bad move m. To that agent, m looks like a good move (or at least, does not look like a particularly bad one). It is epistemically possible, for that agent, that m is a good move, even though it is actually a bad move. In fact, given the rules of chess and the state of the game (which we may assume the agent knows), m could not possibly be a good move. A better chess player might be able to see that m is a bad move, perhaps by using her reasoning ability to discover that making m provides the opponent with a winning strategy. The better player can rule out m as a good move, so its being a good move is not an epistemic possibility for her.

[12] For purely technical reasons it is useful to work with a serial transition relation, so that the final state on each branch is treated as a reflexive point.

[13] Formal definitions of the class of models that correspond to a particular agent's rules are given in [Jag06a, Jag06b, AJL06].

[14] Since the meaning of 'inference' need not be limited to deductive inference, this allows the agent to use common-sense reasoning to try to rule out β.

The agent's ability to reason within its resource bounds affects what is and is not epistemically possible for that agent. Consider a deductive agent a that has the information that $p_1 \vee \cdots \vee p_n \vee p_{n+1}$ and that $\neg p_1 \wedge \cdots \wedge \neg p_n$ (take bracketing to be to the right); and can reason using conjunction elimination and the rule $(\alpha \vee \beta), \neg\alpha \rhd \beta$. Is it epistemically possible, for this agent, that p_{n+1} is false? Clearly not if a has enough time in which to reason, for a can then infer an explicit contradiction $p_{n+1}, \neg p_{n+1}$ in $2n$ applications of its rules. If agent a is allowed $2n$ steps of reasoning, it is possible for it to rule out the truth of $\neg p_{n+1}$, given its prior information. But now consider an agent b with the same initial information as a, the same rules for reasoning with but only n steps in which to reason. It cannot discover an explicit contradiction between the information it has and $\neg p_{n+1}$. Thus, for agent b, $\neg p_{n+1}$ is an epistemic possibility, even though b has the same prior information as agent a, for whom $\neg p_{n+1}$ is not an epistemic possibility. The diagram below shows possible ways of reasoning from the agent's information combined with $\neg p_{n+1}$ for $n = 2$.

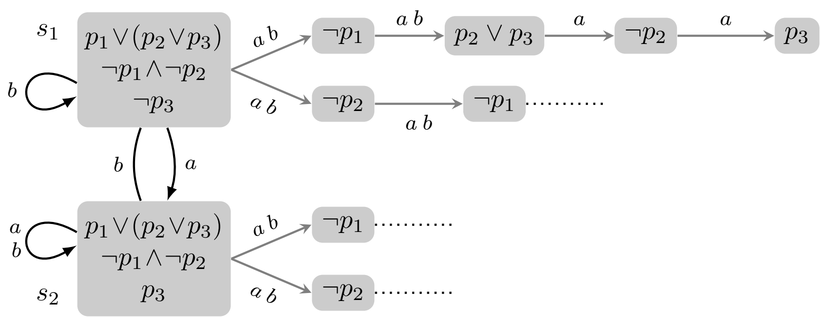


Intuitively, the labels in the leftmost state are each epistemically possible for agent b, but not for a. Using the familiar terminology, I shall say that the root state is *epistemically accessible* to agent b but not to a, for only a can discover the inconsistency implicit at the root of the model. Note that we do not require agent a to actually discover the contradiction: it is enough that the agent could discover it, given its reasoning ability and the resources available to it. It is this reliance on what an agent could do, rather than what it will do, that gives the account a suitable element of normativity.

To make this idea more precise, let i be an arbitrary agent and $\delta_i \in \mathbb{N}$ be agent i's time bound, represented as the number of transitions the agent is allowed to make in the transition system. An explicit contradiction can be inferred by agent i from a state s iff there is a formula $\alpha \in \mathcal{L}$ such that $s \Vdash \Diamond^{\delta_i}([\alpha] \wedge [\neg\alpha])$.[15] In our example, let us set $\delta_a = 4$ and $\delta_b = 2$; let s_1 be the leftmost state in the diagram above; and let s_2 be similar to s_1 but labelled

[15] Note the way both brackets and parentheses are used here: the formula $\Diamond^{\delta_i}([\alpha] \wedge [\neg\alpha])$ says that there is a state s', reachable in δ_i transitions, which supports both $[\alpha]$ and $[\neg\alpha]$, i.e. both α and $\neg\alpha$ label s'.

by p_3 in place of $\neg p_3$. The labels on s_1 are inconsistent and a contradiction can be derived by using the agent's inference rules in four steps. Thus, agent a can discover an explicit contradiction by reasoning from s_1, whereas b cannot. Consequently, s_1 is epistemically accessible to b but cannot be to a. The following diagram shows possible ways of reasoning from s_1 and s_2 using the agent's rules. The grey arcs are transitions between states and the black arcs show show the epistemic accessibility relations (arrow-less black arcs represent bi-directional accessibility).



To accommodate this notion of epistemic accessibility formally, the epistemic accessibility relation for an agent i, written $\sim_i$, must be restricted to states in which agent i cannot find an explicit contradiction, given its resource bound:[16]

$$s \sim_i u \text{ only if } u \nVdash \Diamond^{\delta_i}([\alpha] \wedge [\neg\alpha]) \tag{1}$$

The assumption here is that the agent's reasoning is monotonic, so that $[\alpha] \rightarrow \Box[\alpha]$. If we wish to model a non-monotonic reasoner, then the abbreviation $\Diamond^{\leq \delta_i}$ should be substituted for $\Diamond^{\delta_i}$ in (1), where $\Diamond^{\leq n}\phi$ abbreviates $\phi \vee \Diamond\phi \vee \cdots \vee \Diamond^n \phi$. However, monotonic reasoning is assumed in what follows. Semantically, condition (1) amounts to:

$$s \sim_i u \text{ only if } \forall v \big(T^{\delta_i} uv \text{ implies } v \nVdash [\alpha] \wedge [\neg\alpha]\big) \tag{2}$$

Here, $T^{\delta_i}uv$ means that v is reachable from u in δ_i many transitions. In what follows, the modal language $\mathcal{L}^+$ used above is replaced by a modal language $\mathcal{ML}$ that does not have a $\Diamond$ or $\Box$ modality, so this latter formulation will be

[16] Note that this is not a *definition* of $\sim_i$, but only a restriction on it. Any relation between states restricted in this way can be treated as an epistemic accessibility relation for some agent.

required. As with $\mathcal{L}^+$, the primitives of $\mathcal{ML}$ are of the form $[\alpha]$ for $\alpha \in \mathcal{L}$. If ϕ is a Boolean-free formula of $\mathcal{ML}$ then $\mathsf{E}_i\phi$ and $\mathsf{I}_i\phi$ are wffs of $\mathcal{ML}$; $\mathcal{ML}$ is then the smallest such language closed under the Boolean connectives. As above, I use 'α', 'β' as metavariables ranging over formulae of $\mathcal{L}$ and use 'ϕ', 'ψ' to range over $\mathcal{ML}$.

A formula $\alpha \in \mathcal{L}$ is an epistemic possibility for an agent i, written $\mathsf{E}_i[\alpha]$, if there is a state epistemically accessible to i that supports $[\alpha]$. But given that s_1 in the diagram is accessible to agent b (from either s_1 or s_2) and that b is allowed two steps of reasoning, $\neg p_1$, $\neg p_2$ and $p_2 \vee p_3$ should also be epistemic possibilities for a. The intuition here is that any formula that an agent could infer from any accessible state is itself an epistemic possibility for that agent. In general,

$$M, s \Vdash \mathsf{E}_i\phi \text{ iff } \exists u, v\big(s \sim_i u,\ T^{\delta_i}uv \text{ and } M, v \Vdash \phi\big)$$

Again, the notion of normativity discussed above is in play. Formulae with embedded modalities are well-formed, e.g. $\mathsf{E}_i\mathsf{E}_j[\alpha]$ means that it is epistemically possible for i that it is epistemically possible for j that α. Note that, because epistemic possibility is captured using $\sim_i$ and $\sim_j$, we do not need an E_j modality in the agent's language $\mathcal{L}$.

There is a more restrictive notion of epistemic possibility available to us, based on the same accessibility relation $\sim_i$. This notion, which I call *strong epistemic possibility*, requires an agent to be able to infer the formula in question from an accessible state however it reasons.[17] A formula α is strongly epistemically possible for agent i, written $\mathsf{E}_i^+[\alpha]$, iff there is a state u accessible to i such that i will infer α within its resource bound, however the agent reasons from u. I now count $\mathsf{E}_i^+\phi$ as a wff of $\mathcal{ML}$ so long as ϕ is a Boolean-free wff and define:

$$M, s \Vdash \mathsf{E}_i^+\phi \text{ iff } \exists u \forall v\big(s \sim_i u,\ T^{\delta_i}uv \text{ and } M, v \Vdash \phi\big)$$

Armed with these two notions of epistemic possibility, I now turn to discussing information.

[17] In the language $\mathcal{L}^+$ containing $\Diamond$ and $\Box$, which I am not using here, the difference between regular and strong epistemic possibility is that the former requires a $\Diamond$-formula to hold at an accessible state, whereas the strong version requires a $\Box$-formula to hold.

6. *Information*

In this section, I develop a formal account of information that respects an agent's given resource bounds. Because the aim of the models I have described above is to capture an agent's ways of reasoning, they should be thought of as purely psychological descriptions of the agent's reasoning potential. The sentences that label a state may bear no relation to the facts that obtain in external world. As a consequence, the notion of information that I discuss here is also disconnected from the external world and so does not imply veracity. According to a number of lines of thought, this notion cannot coincide with genuine information which, it is argued, must be veridical. Dretske for one holds that "false information and mis-information are not kinds of information — any more than decoy ducks and rubber ducks are kinds of ducks" [Dre81, 45]. In a similar vein, Grice comments that "false information is not an inferior kind of information; it just is not information" [Gri89, 371]. This view is supported by [BS97, Flo04, Flo05a].

If their arguments prove sound, then the notion I capture here cannot be that of genuine information. For want of a better name, I will call the category that includes both information and misinformation (including false information) *proto-information*. Proto-information is what an agent could take to be genuine information before consulting the world. Although the notion that I discuss here is strictly speaking proto-information, I will use the term 'information' in place of 'proto-information' to avoid unnecessary clumsy phrasing. My principle target is to capture the static notion of an agent *being informed that* α, which I take to mean that the agent has the information that α. For those taking the Dretske-Grice-Floridi line on the veracity of information, read 'being proto-informed' and 'having the proto-information' here.

Because the account of epistemic possibility given above can capture resource bounds, we can follow the standard methodology and encode the information that an agent has using the epistemic accessibility relation. In the diagram above, both agents have the information that $p_1 \vee (p_2 \vee p_3)$ and that $\neg p_1 \wedge \neg p_2$ (i.e. each agent has this information at each state). In addition, agent a also has the information that p_3 at each state because, although $\neg[p_3]$ is supported by s_1, s_1 is not accessible by a from either s_1 or s_2. Agent b does not have the information that p_3 because, as far as b is concerned, s_1 and s_2 are indistinguishable. Since both agents can reason for at least one step, they both realize that the information they have includes the information that $\neg p_1$ and that $\neg p_2$.

Being informed, then, is a universal modality. In general, agent i has the information that α, written $\mathsf{I}_i[\alpha]$, iff i can infer α from an accessible state within its resource bound. Again, embedded modalities are allowed, as in $\mathsf{I}_i\mathsf{I}_j[\alpha]$, which says that i has the information that j has the information that

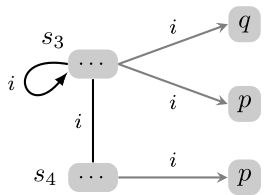
α. In general:

$$M, s \Vdash \mathsf{I}_i\phi \text{ iff } \forall u \exists v \big(s \sim_i u \text{ only if } T^{\delta_i}uv \text{ and } M, v \Vdash \phi\big)$$

A notion of strong information can also be defined, which stands to information as strong epistemic possibility stands to epistemic possibility. Agent i strongly has the information that α, written $\mathsf{I}_i^+[\alpha]$, when i would infer α within its resource bound from any accessible state, however it reasons. I now count $\mathsf{I}_i^+\phi$ as a wff of $\mathcal{ML}$ so long as ϕ is a Boolean-free wff of $\mathcal{ML}$ and set:

$$M, s \Vdash \mathsf{I}_i^+\phi \text{ iff } \forall u, v \big(s \sim_i u \text{ only if } T^{\delta_i}uv \text{ and } M, v \Vdash \phi\big)$$

Note that the dual to any of the modalities I_i, I_i^+, E_i and E_i^+ is not well-formed, for $\mathsf{I}_i\neg\phi$ is not permitted by the rules of $\mathcal{ML}$. It is instructive to lift this restriction and form a language $\mathcal{ML}^*$ that is similar to $\mathcal{ML}$, except that $\mathsf{I}_i\phi$ is well-formed in $\mathcal{ML}^*$ so long as ϕ is (and similarly for the other modalities). In the logic over $\mathcal{ML}^*$, the information modality I_i is not the dual of the epistemic possibility modality E_i. Although $\neg\mathsf{E}_i\neg[\alpha] \rightarrow \mathsf{I}_i[\alpha]$ is valid, the converse is not, as the following diagram shows:[18]

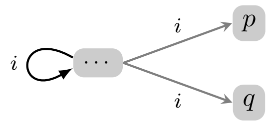


However, although epistemic possibility is not the dual notion to information, *strong* epistemic possibility is. We have: $\mathsf{I}_i[\alpha] \leftrightarrow \neg\mathsf{E}_i^+\neg[\alpha]$. If i has the information that α then there is no accessible state from which i cannot infer α within its resource bound.

Although $\mathcal{ML}^*$ allows for this neat logical characterization of the modalities, it presents a problem when we come to interpret arbitrary formulae in English. As in $\mathcal{ML}$, bracketed formulae function as *that*-clauses, giving the content of the agent's attitudes. The remainder of the constructs in $\mathcal{ML}^*$ either classify the agent's attitude, i.e. I_i and E_i, or else state logical

[18] Suppose $\delta_a = 1$ and so a can infer p from either s_3 or s_4. Consequently, $\mathsf{I}_a[p]$ holds at both s_3 and s_4. The agent could also infer q from s_3; but doing so uses up its resources and so it could not then infer p. There is then a way of reasoning from s_3 in which p is not reached, hence $\mathsf{E}_a\neg[p]$ holds at both s_3 and s_4.

relationships between attitudes, e.g. $\mathsf{I}_i[\alpha] \rightarrow \mathsf{E}_i[\alpha]$ means that if i has the information that α, then it is epistemically possible for i that α. Difficulties in interpreting $\mathcal{ML}^*$ arise when Boolean connectives appear within the scope of a modal operator, but not within brackets. For example, $\mathsf{I}_i\neg[\alpha]$ does not mean that i has the information that $\neg\alpha$, nor does it mean that i does not have the information that α. The diagram below satisfies $\mathsf{I}_i\neg[p] \wedge \mathsf{I}_i[p] \wedge \neg\mathsf{I}_i[\neg p]$:



If one wants to use $\mathcal{ML}^*$ in place of the more restricted $\mathcal{ML}$, then the intuitive meanings of modal formulae must be gleaned from the meanings of the formulae allowed in $\mathcal{ML}$. For example, $\mathsf{I}_i\neg[\alpha]$ is equivalent to $\neg\mathsf{E}_i^+[\alpha]$, and so means that α is not strongly epistemically possible. It might be the case that α is epistemically possible for i, although i could fail to spot this fact. $\mathsf{I}_i\neg[\alpha]$ is at best only indirectly about the information that i has.

As usual in modal logics, other conditions on I_i and E_i and their strong variants can be imposed by altering the properties of the $\sim_i$ relation. For example, $\mathsf{I}_i[\alpha] \rightarrow \mathsf{I}_i\mathsf{I}_i[\alpha]$ is valid on the class of transitive frames. However, $\mathsf{I}_i[\alpha] \rightarrow [\alpha]$ is not valid on the class of reflexive frames, for when $\delta_i > 0$, $\mathsf{I}_i[\alpha]$ may hold at a reflexive point s for which $V(s,\alpha) = \mathit{false}$. This is allowed, for example, if there is a state u accessible from s and $V(u,\alpha) = \mathit{true}$. Just which conditions should hold of the accessibility relation in an account of information is a moot point. The issue is discussed elsewhere in this volume [Flo06] and since a detailed discussion is beyond the scope of the present work, I pass over the issue.

Let us call the account just given the *bounded rationality account* of being informed. It clearly avoids information overload, for the information that an agent has need not be closed under consequence and an agent's information state need not include all tautologies. For example, $\mathsf{I}_i[\alpha] \wedge \mathsf{I}_i[\alpha \rightarrow \beta]$ does not imply $\mathsf{I}_i[\beta]$ and $\neg\mathsf{I}_i[\alpha]$ is satisfiable when α is a propositional tautology. An agent may also be modelled as having inconsistent information, without thereby being informed of every sentence in the language, for $\mathsf{I}_i[\alpha] \wedge \mathsf{I}_i[\neg\alpha]$ is satisfiable.

By way of comparison, Fagin and Halpern have a different approach to avoiding information overload in [FH88].[19] In their account (which uses a

[19] In fact, [FH88] discusses belief rather than information and so concerns the logical omniscience problem, rather than information overload. Nevertheless, their approach to avoiding logical omniscience might be used to avoid overload in a logic of information.

standard modal language), $\mathsf{I}_i\phi \wedge \mathsf{I}_i\psi \to \mathsf{I}_i(\phi \wedge \psi)$ is not valid. Their explanation is based on 'states of mind' of the agent: it may have the information that ϕ in one state of mind and the information that ψ in another, but never put the two pieces of information together as $\phi \wedge \psi$. This also allows an agent to have inconsistent information: it might have the information that ϕ in one state of mind and the information that $\neg\phi$ in another. However, each frame of mind must itself be perfectly consistent and the information that holds within each is closed under consequence. Each state of mind itself then suffers from information overload. This is much less plausible. It seems perfectly possible for an agent to entertain inconsistent information in one and the same state of mind, provided that it does not discover that this information is inconsistent. This shows the superiority of the bounded rationality account of information over accounts such as Fagin and Halpern's.

Another advantage of the bounded rationality account is that it removes the temptation to confuse epistemic with metaphysical possibility. That conceivability (viewed as epistemic possibility) does not entail genuine, metaphysical possibility is evident on this view. We might ask: just what *is* an epistemic possibility? There is a temptation here to make too much of the notion ontologically. What an epistemic possibility *is*, is nothing more than the agent's inability — due to her bounded rationality — to find any explicit contradictions in what she considers possible. This is why epistemic possibility cannot be considered on a par with metaphysical possibility: in the epistemic case, what seems possible to an agent really is epistemically possible for her; though, of course, it might not be metaphysically possible in the slightest.

I will conclude this discussion by showing how to model *becoming informed* as an update on the epistemic accessibility relation $\sim_i$. In becoming informed that α, one no longer considers any states which satisfy $\neg[\alpha]$, or from which $\neg[\alpha]$ can be inferred, to be possible. The information contained in α for agent i can be captured by a restriction on $\sim_i$ to states from which a state that supports $[\alpha]$ is reachable in δ_i transitions. An update by α thus restricts $\sim_i$ to pairs (s, u) such that there is a state v with $T^{\delta_i}uv$ and $v \Vdash [\alpha]$. On becoming informed that α, agent i gains the information that α, $\mathsf{I}_i[\alpha]$. This is a fairly weak notion of becoming informed, for becoming informed that $\alpha \wedge \beta$ need not give rise to the information that α and β.

A stronger notion of becoming informed occurs when an agent is *explicitly* informed that α, e.g. by a public announcement. After the announcement, the agent gains the information that α regardless of its reasoning abilities. Even an agent i for which $\delta_i = 0$ gains information from a public announcement. Becoming explicitly informed that α can be modelled as an update that restricts $\sim_i$ to states that support $[\alpha]$, i.e. to pairs (s, u) such that $u \Vdash [\alpha]$. On becoming explicitly informed that α, agent i gains the *strong* information that α, $\mathsf{I}_i^+[\alpha]$. Moreover, if $\delta_i \geq 0$ then, after becoming explicitly informed that $\alpha \wedge \beta$, i has the information that α and β, $\mathsf{I}_i[\alpha] \wedge \mathsf{I}_i[\beta]$. For such agents, becoming explicitly informed that $\alpha \wedge \beta$ entails becoming informed that α and that β. Together, these two notions of becoming informed capture bounded rationality. They allow us to model agents that are neither perfectly rational ideal reasoners nor completely irrational logical dunces. As [Ho95] terms it (in a different context), we model agents that are neither logically omniscient nor logically ignorant.

To see how this account is advantageous, let us return to the example of agents a and b from above. Recall that they both have information that entails p_3 although agent b does not realize this. Consequently, both p_3 and $\neg p_3$ are epistemic possibilities for b. If an external source remarks to both a and b that p_3, what should we expect their reactions to be? Agent a, for whom all states from which $\neg p_3$ are derivable have been ruled out, will find this utterly uninformative. The announcement initiates an update that restricts $\sim_i$ to states from which $\neg p$ is not derivable in δ_a transitions. But, since all such states are inaccessible to a to begin with, the update has no effect on $\sim_a$. Thus, the announcement that p_3 has no informational content, as far as agent a is concerned.

However, the update does have an effect on $\sim_b$, eliminating the pairs (s_1, s_1) and (s_2, s_1) (recall that $s_1 \Vdash [\neg p_3]$). Agent b becomes explicitly informed that p_3 and so, for agent b, the announcement that p_3 does have informational content. This content is modelled precisely by the effect that the update has on $\sim_b$. This is just as it should be. As discussed in section 3, the consequences of an agent's information, including logical truths, can be informative. I trust this highlights the benefits of the bounded rationality account of epistemic possibility over the traditional notion in terms of metaphysically or logically possible worlds.

REFERENCES

[AB75] A. R. Anderson and N. D. Belnap. *Entailment — the logic of relevance and necessity*. Princeton University Press, Princeton, NJ, 1975.

[AJL06] N. Alechina, M. Jago, and B. Logan. Modal logics for communicating rule-based agents. In *Proceedings of ECAI 06*, 2006. To appear.

[BH64] Y. Bar-Hillel. *Language and Information: Selected Essays on Their Theory and Application*. Addison-Wesley, London, 1964.

[BS97] J. Barwise and J. Seligman. *Information Flow: The Logic of Distributed Systems*. Cambridge University Press, Cambridge, 1997.

[Cre72] M.J. Cresswell. Intensional logics and logical truth. *Journal of Philosophical Logic*, 1:2–15, 1972.

[Cre73] M.J. Cresswell. *Logics and Languages*. Methuen and Co., 1973.

[DR02] J.M. Dunn and G. Restall. Relevance logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 6, chapter Relevance Logic Handbook of Philosophical Logic, pages 1–136. Kluwer Academic, Dordrecht, 2002.

[Dre81] F. I. Dretske. *Knowledge and the Flow of Information*. Blackwell, Oxford, 1981.

[FH88] R. Fagin and J.Y. Halpern. Belief, awareness and limited reasoning. *Artificial Intelligence*, 34:39–76, 1988.

[FHV90] R. Fagin, J.Y. Halpern, and M.Y. Vardi. A nonstandard approach to the logical omniscience problem. In R. Parikh, editor, *Proceedings of the Third Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 41–55. Morgan Kaufmann, 1990.

[Flo04] L. Floridi. Outline of a theory of strongly semantic information. *Minds and Machines*, 14(2):197–222, 2004.

[Flo05a] L. Floridi. Is information meaningful data? *Philosophy and Phenomenological Research*, 70(2):351–370, 2005.

[Flo05b] L. Floridi. Semantic conceptions of information. The Stanford Encyclopedia of Philosophy, Winter 2005.

[Flo06] L. Floridi. The logic of being informed. *Logique et Analyse*, 49, December 2006.

[Fre64] Gottlob Frege. *The Basic Laws of Arithmetic*, chapter The Russell Paradox, pages 127–143. University of California Press, Berkeley, 1964.

[Gri89] H. P. Grice. *Studies in the Way of Words*. Harvard University Press, Cambridge, Mass., 1989.

[Hin62] J. Hintikka. *Knowledge and belief: an introduction to the logic of the two notions*. Cornell University Press, Ithaca, N.Y., 1962.

[Hin75] J. Hintikka. Impossible possible worlds vindicated. *Journal of Philisophical Logic*, 4:475–484, 1975.

[Ho95] D.N. Ho. Logical omniscience vs. logical ignorance. In C.P. Pereira and N. Mamede, editors, *Proceedings of EPIA'95*, volume 990 of *LNAI*, pages 237–248. Springer, 1995.

[Jag05] Mark Jago. Modelling assumption-based reasoning using contexts. In *Proceedings of the Contextual Representation and Reasoning workshop*. CEUR electronic proceedings, 2005.

[Jag06a] Mark Jago. *Logics for Resource-Bounded Agents*. PhD thesis, University of Nottingham, 2006. Forthcoming.

[Jag06b] Mark Jago. Rule-based and resource-bounded: A new look at epistemic logic. In *Proceedings of the Logics for Resource-Bounded Agents workshop*, ESSLLI 06, 2006. To appear.

[Kri80] Saul Kripke. *Naming and Necessity*. Blackwell, Oxford, 1980.

[Lev84] H. J. Levesque. A logic of implicit and explicit belief. In *National Conference on Artificial Intelligence*, pages 1998–202, 1984.

[Lew86] David Lewis. *On the Plurality of Worlds*. Basil Blackwell, Oxford & New York, 1986.

[Qui60] W.V.O. Quine. *Word and Object*. MIT Press, Cambridge, Mass., 1960.

[Qui70] W.V.O. Quine. On the reasons for indeterminacy of translation. *Journal of Philosophy*, 67:178–83, 1970.

[Ran75] V. Rantala. Urn models. *Journal of Philosophical Logic*, 4:455–474, 1975.

[RB79] N. Rescher and R. Brandon. *The Logic of Inconsistency*. Rowman and Littlefield, 1979.

[Rus69] Bertrand Russell. *The Autobiography of Bertrand Russell, 3 vols*. Little, Brown and Company, Boston and Toronto, 1967, 1968, 1969.

[Sta91] R. Stalnaker. The problem of logical omniscience I. *Synthese*, 89:425–440, 1991.

[VB03] J. Van Benthem. Logic and the dynamics of information. *Minds and Machines*, 13(4):503–519, 2003.

[Whi03] M. Whitsey. Logical omniscience: a survey. Technical Report NOTTCS-WP-2003-2, School of Computer Science and IT, University of Nottingham, 2003.

[Wil06] Interview with Andrew Wiles. `http://www.simonsingh.net/Andrew_Wiles_Interview.html`, originally published in Nova magazine. Accessed March, 2006.

[Wit22] Ludwig Wittgenstein. *Tractatus Logico-Philosophicus*. Kegan Paul, 1922.