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 captured in a formal model, which I call a *fan*. I then show how to use fans to build formal models of *being informed*, as well as knowledge, belief and information update.

1   Introduction

I will concentrate on the concept of information which 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. "The 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

1

the kind we find in an instruction manual or cooking recipe. Declarative information is alethically qualified [Flo05a, p. 3] (i.e. expressions of such declarative information are truth-apt). Perhaps declarative information is not so different from other kinds of information. After all, an imperative can be formed from a declarative: *make it the case that p*! just as an interrogative can be formed: *is it the case that p*? One would expect that, in order to comprehend the latter question, one must first comprehend the information that would be asserted as '*p*', were '*p*' true. That is, one must know what information p would convey, were it true. In a similar way, one must know what information p would contain, were it true, in order to follow the instruction: *make it the case that p*![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 which no one could ever cognise (because, for example, of the finite number of fundamental particles 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, realise 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 which 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 to effect a change in a *cognitive* information state (i.e. by being read). Thus, information must at bottom be potential information *for* someone. I hold that there could not be information that, as a matter of necessity, could not be known.

---

[1] The proviso that '*p*' be true reflects that fact that information must be truthful, that is, *mis*information is not information at all.

2 Information Update

One of the key directions in the logical analysis of information is to treat an agent's static information state as the set of all relevant possibilities that she entertains (e.g. [VB03]). If agent a knows that $p \vee q$, but does not know which disjunct is true, there are three relevant ways in which the world could be, given what a knows. 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. If the agent is then informed that p is false—in the sense that a accepts the information to be true—then two of these possibilities are ruled out. The $p \neg q$ and pq worlds are ruled out of a's considerations, leaving only the $\neg pq$ worlds as candidates for how the world could be (that is, given what information a has). 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.
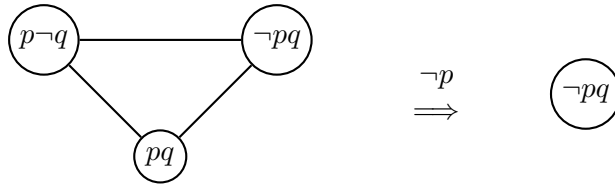

Figure 1: Updating by $\neg p$

As mentioned in the introduction, we have both a static and a dynamic notion of information. We might contrast *being* informed that ϕ, in the sense ϕ is part of the agent's current state of information, with *becoming* informed that ϕ as an active, dynamic process. 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; a cannot tell whether the actual world is a $\neg pq$ world, a $p \neg q$ world or a pq world. Formally, we treat *epistemic indistinguishability* for agent a as a relation $\sim_a$ on 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 tool which 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 of course 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 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.

There is a clear relation between this notion of informational content and knowledge update. Following Hintikka [Hin62], a static account of knowledge can be given in terms of the worlds that an agent cannot distinguish betwen. 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 $\sim_a$. The information contained within ϕ, then, is on a par with the change in a's epistemic state when it comes to know that ϕ.

In linking information to knowledge, there are two points that should be raised. Firstly, some view information as true by definition; *mis*information is not a subspecies of information at all, but only pseudo-information [Flo05a].

4

Those who hold that information may be false should talk of a change in an agent's belief state, rather than its state of knowledge. Secondly, the logic of information may well be stronger than the logic of knowledge (even the knowledge of ideal agents, as described by Hintikka). I have described the indistinguishability relation as one which engenders partitions on the total set of worlds such that, from within a certain partition, an update simply makes certain worlds vanish (as in figure 1).[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 just sketched 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 knowledge, belief and information.

3   Informative Inference

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

1. Suppose a is in informed that $\phi \rightarrow \psi$ and 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

---

[2]Each $\sim$ is thus an equivalence relation, which is too strong for an analysis of knowledge. The scheme $\neg\mathsf{K}\phi \rightarrow \mathsf{K}\neg\mathsf{K}\phi$ (i.e. whenever an agent does not know something, it knows that it does not know it) is S5-valid but, even in the case of agents with ideal reasoning ability, this is implausible. It is more common to take a logic between S4 and S5 (including S4 itself) to be the correct logic of knowledge, such that $\mathsf{K}\phi \rightarrow \mathsf{KK}\phi$ is valid. I comment briefly on this so-called KK-principle (valid on all transitive frames) in section 5 below.

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 $\phi \to \psi$. 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 analysed 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 $\phi \to \psi$ 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 which 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 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 which follow logically from ϕ. Thus the consequences of a set of sentences contain at most the informational content that the sentences themselves contain. The view is very much that *the conclusion is contained in the 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 and under valid implication are present in the weakest normal logic of knowledge, K.[4] I 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

---

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

[4]It is possible to use Scott-Montague semantics to model knowledge, 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.

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*). $\vdash \phi \rightarrow \psi$ 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 recognised by any competent speaker of the language. Now, 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 $\vdash \phi$ in the propositional calculus (hence 'is a tautology' obeys a disquotation scheme for tautologies just as 'is true' does for truths.) It follows that 'ϕ is a tautology' cannot be informative. If true, it is 'empty'; if false, it is misinformation. However, for someone who does not recognise 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 that 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.

---

[5]This has the unintuitive result that contradictions have maximum informational content. This is known as the Bar-Hillel-Carnap paradox; their conclusion is that contradictions are "too informative to be true" [BH64, p. 229]. The problem is avoided by taking informativeness to imply truth, as I have done here, although the elegance of the mathematical model is then lost.

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 one exists at all). 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 how Wiles himself described 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

---

[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].

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.

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 update account of information, 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 along the lines of Hintikka's analysis of the closure of knowledge in [Hin75] as follows:

1. 'a is informed that 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 possible world;

4. All indistinguishable 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 conse-
> quences 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 which are indistinguishable by a 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.

Hintikka devotes the remainder of his article [Hin75, pp. 477–483] to describing *impossible possible worlds*, logical models which 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].[7] 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.

---

[7]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]).

The details of such models are provided by Rantala in [Ran75], where he uses the term *urn models*. The domain is conceived 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. Now, 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 which may alter the contents from one draw to the next. In this way, sentences which are classically invalid may nevertheless be satisfied by an urn model. The level of inconsistency in a urn model is viewed as the number of draws which occur before any change in the available individuals takes place. Suppose the largest number of nested quantifiers in a sentence ϕ is d (d is said to be the *depth* of ϕ). Then, 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 recognise 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 which a already has. There will be worlds which 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.

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* 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.[8] Thus, Hintikka's solution

---

[8]Similar examples are discussed in [Jag06a, ch. 2].

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]. The logic within Levesque's worlds is based on Belnap's 4-valued logic [AB75] and Dunn [Dun76], which makes use of a truth *relation* on $\{true, false\}$ rather than a function). A similar account is given by Fagin, Halpern and Vardi in [FHV90]. The semantics here is based on the Routley star operator approach to relevant logic (see [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 which 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 recursive truth conditions, so that there are infinitely many theorems of Λ, then the following should 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. Adding 2$'$ to 1, 3, 4 generates a contradiction similar to Hintikka's problem. As a consequence, any agent who has been informed that ϕ cannot possibly be 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 denies agents information of the classical principles which are not principles of the chosen logic. The fact that real agents do not suffer from information overload is not due to their lacking reasoning *principles*, as if they 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 derive 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. Agents can only use information that they are aware of and hence may *think* that some consequence of their information is informative (whereas in fact it cannot be).

Awareness 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 formulas 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 and becoming informed *directly*, using whatever principles were used to determine the properties of the awareness set? A potential notion of awareness given in [FH88, 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 will propose below, although I will do so directly in terms of a possible worlds analysis, making 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 (in part) psychological notions. Of course, psychological attitudes are part of being as well. Beliefs and desires really do exist but, according to Dennett, they "can be discerned only from the point of view of one who adopts a certain *predictive strategy*, and [their] existence can be confirmed only by an assessment of the success of that strategy" [Den87, p. 15]. In [Jag06a], I discuss a formal model of belief which makes use of Dennett's predictive strategy but, unlike his account, does not assume that agents are ideally rational reasoners.

In the remainder of this section, I introduce these structures and show how they can be used to develop an account of epistemic possibility. Such structures appeal to what an agent could determine given limited resources. This is similar in some respects to Fagin and Halpern's notion of awareness as the sentences that an agent could determine in a specified space and/or time bound, but treated as a genuine semantic notion. Consider an agent wondering whether this or that set of sentences is compatible with prior information. If the agent can find no explicit contradiction between this information and the set of sentences, then it has no reason to suppose that those sentences do not describe the way things actually are. We can turn this idea around and say that *any* arbitrary set of sentences that the agent could not recognise an explicit contradiction in, given limited resources, may take the place of an epistemic possibility. We can then develop an account of information (as well as knowledge and belief) in more or less the standard way.

Fix a denumerable set of propositional letters $\mathcal{P}$ and let $\mathcal{L}$ be the smallest language closed under $\mathcal{P}$ and the usual Boolean connectives and the sentential operator 'E'. Models are relational structures whose domain is a set of points S (which, following standard practise will be called *states*). Assuming we model a group of n agents, models contain two kinds of relations, namely a serial transition relation T and the indistinguishability relations $\sim_i$ for each agent i. For simplicity of explication, assume that T forms a number of unconnected tree structures on S.[9] We also have a labelling function V that labels each state $s \in S$ with a set of non-modal sentences of $\mathcal{L}$

---

[9]The restriction to models in tree form is inessential, as it is a theorem of normal modal logics that every model is bisimilar to a tree model. See, for example, [BdRV02].

and a function ρ which assigns a set of inference rules to each agent. It is important that the sentences which hold at a state are not deductively closed. In what follows, I assume that ρ assigns at least conjunction introduction to each agent.[10]

The particularity of the models we are interested in comes in the way that T and the $\sim_i$s are fixed. Whenever Tsu, we say that there is a *transition* from s to u. These transitions model potential *atomic* inferences—the act of inferring just one new formula from those that hold at parent states. Thus whenever Tsu holds, u must be labelled just like s except that, in addition, u is labelled by some additional formula. For some formula ϕ, we have $V(u) = V(s) \cup \{\phi\}$ whenever Tsu. Here, I say that u *extends* s by ϕ. A state s may be extended by a formula ϕ when ϕ is the conclusion of a rule of inference whose premises match the sentences which label s (or rather, since such rules tend to be meta-rules containing sentence-variables, we should talk about ϕ being the conclusion under some substitution instance of a rule whose premises, under that same substitution, are all labels of s). In a model M, whenever a state s may be so extended, there is a state u suitably extending s such that Tsu.[11]

Models also contain a function δ assigning a natural number to each agent, which represents how many inferences (applications of the rules assigned to the agent by ρ) an agent may perform before its resources run out. Now, if the entire tree represents the reasoning possibilities of an ideal agent, with one possible line of reasoning per branch, we can limit our attribution of rationality by chopping off each of the branches in the tree at depth δ. We might imagine a wedge-shaped fan, whose sides are of length δ, held over the tree so that its sides run parallel to the outermost branches of the tree. The area within the fan represents reasoning which the agent can perform before its resources run out. An example will illuminate this idea.

Consider the tree τ_1 whose root s_1 is labelled by $\{p \vee q \vee r, \neg p, \neg q, \neg r\}$ and suppose all agents can use the rule

$$\frac{\phi \vee \psi \vee \chi \qquad \neg\phi}{\psi \vee \chi}$$

and, for the sake of the example, assume agents can rearrange disjuncts instantaneously (so rearrangement does not cost a transition). Clearly $V(s_1)$

---

[10]This allows us to say in the syntax of our language that there is a state in which *both* ϕ and ψ hold. Since these sentences are not deductively closed, this is not equivalent to $\phi \wedge \psi$ holding. When $s \Vdash \Diamond^n\phi$ and $s \Vdash \Diamond^n\psi$, we can write $s \Vdash \Diamond^n\Diamond(\phi \wedge \psi)$ to say that ϕ and ψ hold in the *same* state, n transitions away from s.

[11]Such models are considered in more detail in [Jag06c, Jag06b].

is inconsistent and a contradictory pair can be derived in two applications of the above rule. But suppose we have an agent a which, because of lack of resources perhaps, can only reason to depth 1, i.e. $\delta_a = 1$. Then s_1 will appear consistent to a (assume a has no further rules for dealing with disjunctions). In considering an agent b such that $\delta_b > 1$, we should exclude the fan beginning at s_1 from our considerations: it is not epistemically possible for b. But we do not have to exclude this fan in the case of agent a, as a could not recognise the contradiction, given its resource bounds.

Let us now see how we can define *being informed*, as well as knowledge, belief and epistemic possibility, within this framework. Suppose a and b both have the information that $p \vee q \vee r$, that $\neg p$ and that $\neg q$ and that $\delta_a = 1$ whereas $\delta_b > 1$. I think it is natural to say that b also has the available information that r, whereas a does not, for there is no possible way for a to access or make use of such information. This intuition is even more forceful in the case of knowledge: agent b knows that r but a does not, for a has no way to recognise that r follows from what it does know. Agent b can distinguish between s_1 and s_2, because however it reasons, it will sooner or later realise that the labels of s_1 (but not of s_2) are inconsistent. Agent a, on the other hand, cannot; so we have $s_1 \sim_a s_2$, but $s_1 \not\sim_b s_2$. The situation is represented in figure 2—assume that both points are reflexive. I shall talk of *indistinguishable fans* when roots of the fans are indistinguishable.
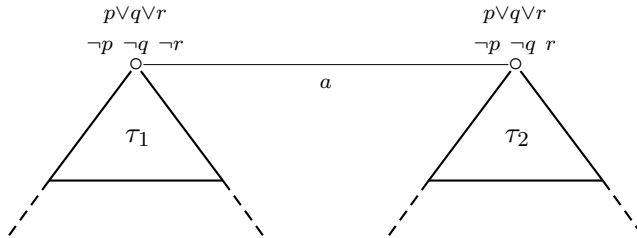

Figure 2: Epistemic possibilities as fans

For any agent i, we only allow $s \sim_i u$ when the u fan contains no explicit contradictions at a single state. I count both a contradictory pair $\phi, \neg\phi$ and a contradictory conjunction $\phi \wedge \neg\phi$ as explicit contradictions.[12] Let us write $T^n su$ when u can be reached from s via n T-transitions. We then have $s \sim_a u$ only if both $\{\phi, \neg\phi\} \not\subseteq V(u')$ and $\phi \wedge \neg\phi \notin V(u')$, where $T^{\delta_a} uu'$, for

---

[12]Recall that labels of states are not deductively closed, so these conditions are distinct. Yet both states at which $\phi, \neg\phi$ hold and those at which $\phi \wedge \neg\phi$ hold are intuitively objectionable.

any ϕ. As a first approximation, suppose we model an agent's information state in the usual way: agent a has the information that ϕ at s iff ϕ appears in all fans which a cannot distinguish from s. Formally, $s \Vdash \mathsf{I}_i\phi$ iff, for all $s' \in S$, $s \sim_i s'$ implies $s' \Vdash \Diamond^{\delta_i}\phi$.

However, as mentioned above, one can only have the information that ϕ if ϕ is true. Misinformation is not a species of information at all. We cannot guarantee that ϕ holds whenever '$\mathsf{I}_i\phi$' holds simply by taking $\sim_i$ to be reflexive, as we would in a regular modal logic of information. Fans are in effect *ways of reasoning*, rather than possible states of affairs. They are genuine epistemic notions and this is why they need not be consistent. But information must be true and thus consistent. We therefore add to the model a set of additional states S^* and a function $*$ on states, associating each state $s \in S \cup S^*$ with a state $s^* \in S^*$ such that $s^{**} = s^*$ and $V(s^*) = V(s)$. We then close the sentences satisfied by states $s^* \in S^*$ under the usual satisfaction rules for Booleans, i.e. $s^* \Vdash \neg\phi$ iff $s^* \nVdash \phi$, and $s^* \Vdash \phi \wedge \psi$ iff $s^* \Vdash \phi$ and $s^* \Vdash \psi$. Thus, if $V(s)$ is classically inconsistent, $V(s^*)$ will contain all non-modal formulae in the language. On the other hand, if $V(s)$ is classically consistent, s^* will behave as a classical possible world.

The correct analysis of an information state is therefore: agent a has the information that ϕ at s iff s^* is consistent and ϕ appears in all fans which a cannot distinguish from s. Formally,

$$s \Vdash \mathsf{I}_i\phi \text{ iff } s^* \nVdash \bot \text{ and, for all } s', s \sim_a s' \text{ implies that } s' \Vdash \Diamond^{\delta_a}\phi$$

Thus, at states s for which s^* is not a possible world (i.e. $V(s)$ is classically inconsistent), an agent has no information at all. Intuitively, s^* gives us the truths which hold in the situation in which the reasoning episode at s takes place. We then add a new modality 'T', with '$\mathsf{T}\phi$' read as 'ϕ is true' such that, for any $s \in S \cup S^*$, $s \Vdash \mathsf{T}\phi$ iff $s^* \Vdash \phi$ and $s^* \Vdash \phi$. This latter proviso rules out 'true contradictions'. Here, the truth modality only applies to consistent states. It follows from these definitions that $\mathsf{I}_i\phi \to \mathsf{T}\phi$ is valid for any agent i. In our above example (figure 2), we see that agent a has the information that $p \vee q \vee r$ at s_2 but merely has misinformation at s_1.

Next, I turn to *epistemic possibility*, a phrase that can be somewhat misleading. The intended sense of 'epistmeic possibility' is that ϕ is epistemically possible for an agent a when, for a, is is epistemically open whether ϕ is true or not. Let us write '$s \Vdash \mathsf{E}_a\phi$' when this is so at a state s. Again, we capture this notion in terms of fans indistinguishable from s by a. Formally, $s \Vdash \mathsf{E}_i\phi$ iff there is a state s' such that $s \sim_i s'$ and $s' \Vdash \Diamond_i^{\delta}\phi$. Given these definitions, it might be thought that *epistemic possibility* is just the

dual of *being informed*, in the sense that being informed closes some of the agent's previously epistemically open questions. However, this is not the case. Whilst $\mathsf{E}_i\phi$ implies $\neg\mathsf{I}_i\neg\phi$, $\neg\mathsf{E}_i\neg\phi$ does not imply that $\mathsf{I}_i\phi$. There could well be a state s, indistinguishable from the current state, at which neither $\Diamond^{\delta_i}\phi$ nor $\Diamond^{\delta_i}\neg\phi$ hold. It seems to be characteristic of epistemic possibilities that they may be partial, as well as inconsistent (but not obviously contradictory) descriptions of states of affairs. 'E' is not the dual of a knowledge or belief operator for similar reasons.

Let us call this the *bounded rationality account* of being informed. It essential to this account that the set of non-modal sentences which are satisfied at a state s is not deductively closed. For example, $\mathsf{I}_i\phi$ and $\mathsf{I}_i\phi \to \psi$ does not imply $\mathsf{I}_i\psi$. However, there is nothing preventing our ascription sentences—those sentences of the form $\mathsf{I}_i\phi$ and $\mathsf{E}_i\phi$—from being closed under classical consequence. If agent a has the information that ϕ, then this is a fact about the world and such facts should behave just like any other. So, for example, $s \Vdash \mathsf{I}_i\phi$ and $s \Vdash \mathsf{I}_i\psi$ implies $s \Vdash \mathsf{I}_i\phi \wedge \mathsf{I}_i\psi$, but $\mathsf{I}_i\phi \wedge \mathsf{I}_i\psi$ is not equivalent to $\mathsf{I}_i(\phi \wedge \psi)$. '$\mathsf{I}_i$' does not distribute over implication or conjunction. The same holds for 'E'.

Conditions on the $\sim_i$ relation do not have their usual effect on these logics. For example, seriality does not guarantee consistency, as states themselves may be inconsistent; and reflexivity does not guarantee truth, for it is not correct to say that the sentences satisfied in some epistemic possibility are true there. A sufficiently complex falsehood might be considered possible, but cannot be true. What is true at s is not what is satisfied at s, but rather what is satisfied at s^*, provided that $\perp$ is not also satisfied there. On the other hand, $\mathsf{I}_i\phi \to \mathsf{I}_i\mathsf{I}_i\phi$ is valid when $\sim_i$ is transitive. On the presentation here, each $\sim_i$ may or may not be transitive. Whether each *should* be transitive is another question. According to the explanation I have given of fans, they tell us what an agent with fixed resources could become aware of, not what they are aware of. Thus, $\sim_i$ should be transitive iff agent i cannot be aware that ϕ follows from its prior information without being aware that it is so aware.

An account of knowledge can be given along the lines of the logic of being informed, with the exception that the relation underlying the definition of knowledge should *not* be transitive, for agents do not always know what they know. Knowledge depends partly on how one's beliefs co-vary across worlds [Noz81].[13] Thus, if one holds that becoming informed implies gaining knowledge, then one should also hold that the relation '$\sim_i$' used to model

---

[13]See [AC05] for a recent defence of a tracking-theory approach to knowledge.

being informed should not be transitive. Whether being informed implies gaining knowledge turns on whether one considers, for instance, a stopped clock to be informative on the two occasions a day when it tells the right time.

An account of belief can also be given in the same way as the logic of being informed by dropping the restriction to consistent states. That is, an agent's beliefs may be inconsistent. If the same underlying indistinguishability relation is used to model both knowledge and belief, then knowledge implies belief (i.e. $\mathsf{K}_i\phi \to \mathsf{B}_i\phi$ is valid). It is of course possible to introduce two families of relations, $\sim_i^{\mathsf{K}}$ and $\sim_i^{\mathsf{B}}$ for each agent i so that the properties of knowledge and belief become decoupled allowing, for example, the latter but not the former to be transitive.[14]

From what has been said, it should be clear that operators for knowledge and belief, 'K' and 'B' so defined, do not distribute over implication or conjunction. One therefore avoids treating agents as logically omniscient. In [FH88], Fagin and Halpern provide a different modal logic of belief in which $\mathsf{B}_i\phi \wedge \mathsf{B}_i\psi \to \mathsf{B}_i(\phi \wedge \psi)$ is not valid. Their explanation is based on 'states of mind' of the agent—it may believe ϕ in one state of mind, and ψ in another, but never put the two together and so never believe $\phi \wedge \psi$. Such episodes do take place and so, as far as it goes, this is a satisfactory explanation. However, each frame of mind, considered on its own, must be perfectly consistent. This sounds much less plausible. It seems perfectly possible for an agent to have inconsistent beliefs even when these beliefs are on the very same topic and entertained in the very same frame of mind, provided that it cannot discover that the beliefs are inconsistent. This shows the superiority of the bounded rationality account of belief 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

---

[14]However, it seems sensible to maintain the scheme $\mathsf{K}\phi \to \mathsf{B}\phi$. Floridi argues otherwise elsewhere in this volume, although the reasons he gives independently of his own logic of information (which is one of those accounts that suffer from the problem of information overload criticised here) are less than conclusive.

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 a relation $\sim_i$. In becoming informed that p, one no longer considers any states which satisfy $\neg p$ to be possible. The information contained in p for agent i can be captured by a restriction on $\sim_i$ to worlds which are not labelled by $\neg p$. An update by ϕ thus restricts $\sim_i$ to pairs (s, u) such that $u \Vdash \Diamond_i^\delta \phi$. Thus, after becoming informed that ϕ, i will not consider $\neg \phi$ to be possible. Similarly, becoming informed that $\neg \phi$ implies that i will not consider ϕ to be possible, as well as not considering $\neg \neg \phi$ to be possible (because $\neg \phi$ can produce a contradictory pair with either ϕ or $\neg \neg \phi$). However, we should note that such updates can only be said to lead to an agent being informed at consistent states, i.e. states s for which s^* is a possible world. Impossible worlds contain no information whatsoever. From our definitions, we can see that informing an agent i that $\phi \wedge \psi$ has the same effect as informing an agent j that ϕ and that ψ, provided that δ_j is in the range $\delta_i \pm 1$ and that both agents use conjunction introduction and elimination rules.

To see how this account is advantageous, let us return to the example of agent a, who is informed that $p \vee q \vee r$, that $\neg p$ and that $\neg q$ but for whom it is open whether r is true (recall that we set $\delta_a = 1$). It is a logical truth that r follows from a's information, but this does not prevent us from informing a that r. 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.

[AC05]   Fred Adams and Murray Clark. Resurrecting the tracking theories. *Australasian journal of philosophy*, 83(2):pp.207–211, June 2005.

[BdRV02]   Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, New York, 2002.

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

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

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

[Den87]   Daniel C. Dennett. *The Intentional Stance*. MIT Press, 1987.

[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.

[Dun76]   J.M. Dunn. Intuitive semantics for first degree entailment and coupled trees. *Philosophicl Studies*, 29:149–68, 1976.

[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, *Theoretical Aspects of Reasoning about Knowledge: Proc. Third Conference*, San Fransisco, California, 1990. Morgan Kaufmann.

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

[Flo05b]   L. Floridi. Semantic conceptions of information. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2005.

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

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

[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.

[Jag06a]   Mark Jago. Belief and bounded rationality. `http://eprints.nottingham.ac.uk/archive/00000383.01.bounded-rationality.pdf`, April 2006.

[Jag06b]   Mark Jago. *Logics for Resource-Bounded Agents.* PhD thesis, University of Nottingham, 2006. Forthcoming.

[Jag06c]   Mark Jago. Rule-based and resource-bounded: A new look at epistemic logic. `http://www.nottingham.ac.uk/philosophy/staff/mark-jago/`, 2006.

[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.

[Noz81]   Robert Nozick. *Philosophical Explanations.* Clarendon press, Oxford, 1981.

[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:pp.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]   Andrew Wiles. Interview with andrew wiles. `http://www.simonsingh.net/Andrew_Wiles_Interview.html`, March 2006.

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