"Now that you mention it, I wonder..." : Awareness, attention, assumption

dejager, S.T.

Citation for published version (APA):
Chapter 1

Introduction

The solution was obvious, as obvious as it had seemed insoluble for as long as he hadn’t solved it[.]

Georges Perec, Life A User’s Manual

1 · An idea too simple to disagree with

We are surrounded all our lives by myriads of possibilities, strictly speaking perhaps limitless possibilities. We, however, are most definitely limited beings; limited certainly in our capacity to understand what possibilities really obtain, but even limited in our awareness of what possibilities might obtain. There may be countless unrealised possibilities, but at any given moment we are only aware of a very small subset of them.

If you accept this truism, you are halfway to accepting the argument of this dissertation. The other half consists in the recognition that the semantic structures underpinning our use and understanding of language must reflect this limited awareness; this second half will require a little more justification.

In this chapter I will introduce the core ideas of unawareness and assumption, giving some motivation and describing some existing work that I will build on in the rest of the dissertation. I start with the possible worlds semantics for natural language as defended by Robert Stalnaker; taking his ideas to their logical conclusion leads quickly to an intuitive understanding of the relevance of awareness for formal semantics. In particular, I will raise two kinds of problem for Stalnaker’s account: the finegrained individuation of worlds, and cases of worlds that are apparently neither ruled out nor accepted as possibilities. Both have been addressed by Stalnaker, but without any great emphasis, and the solutions he proposes have not been taken up with the same enthusiasm as the ‘big picture’. If I am correct both problems are to be solved in essentially the same way: by investigating the agent’s conscious beliefs and uncertainties, and then by specifying what should be done with everything that she is not consciously aware of.

In the second half of the chapter I will briefly summarise some work that
has already been done in this direction (not primarily intended for linguistic applications), in the growing literature on awareness for epistemic logic and economic applications. There are two distinct themes in this existing literature, corresponding to different ways of dealing with whatever the agent is not aware of, and these themes provide the division into two parts of the remainder of the dissertation.

To begin with, though, I want to introduce the basic framework that is taken for granted throughout: the possible worlds analysis of meaning, and in particular of belief.

2 · Possible worlds for belief

The possible worlds representation of propositional content has a long philosophical history which I do not intend to recap here. I take as my departure point the theory Robert Stalnaker has proposed for belief (in [Sta84]) and conversational context (in papers collected in [Sta99]). To introduce the notions we will focus in this chapter on the simplest case, that of a single believer; the linguistic applications will become more complicated as they include at least two conversational participants.

According to Stalnaker the proper representation of a proposition (the semantic object of a belief attribution or truth judgement) is as a set of possible worlds. A possible world is a “way things might have been”, and a proposition can be identified with a set of these worlds; the proposition is true at each world in the set, and false at each world outside the set. While a sentence expresses a proposition (its meaning) via systematic rules of interpretation, the meaning does not need to contain or reflect the structure present in the sentence. (The conjunction of two sentences expresses the proposition that is the set intersection of their meanings; this proposition does not ‘remember’ that it is a conjunction, it is simply a set of worlds.) If propositions are represented in this way then beliefs can be as well: an agent’s epistemic state can be represented by the set of worlds she holds possible; she then believes every proposition that is true in all those worlds (that is, since propositions are simply sets of worlds, every superset of her belief state is a belief she holds; see Figure 1.1).

This shallow description conceals philosophical depths which I do not intend to plumb. Stalnaker’s book Inquiry [Sta84] is an elaborate argument for and application of this picture; in particular see Chapter 3, “Possible Worlds”, for the light metaphysical commitments of taking “ways things might have been” seriously, and Chapter 4, “Belief and Belief Attribution”, for the main arguments in favour of representing belief in this way.

---

1Somewhere hidden in those depths is Stalnaker’s rejection of David Lewis’s “centered worlds” (essentially world-individual pairs). I mention this because Stalnaker’s position is perhaps not as popular as Lewis’s; so far as I can see either is compatible with everything I will say about unawareness.
Figure 1.1: A belief set and some beliefs. The square indicates the space of possibilities; the shaded region is a belief set. The two ovals are propositions entailed by the belief set, that is, they are propositions believed by the owner of the belief set.

It is important for Stalnaker’s metaphysics that a “way things might have been” is a full specification: any question we can think to ask about the state of affairs in some possible world \( w \) would be completely answered if we could inspect \( w \) in sufficient detail. While this may be metaphysically sound, it is difficult to reconcile with the notion that we use possible worlds to represent beliefs ‘in our heads’. There will be a multitude of invisibly fine distinctions that we are unaware of, so that the possibilities we distinguish between should to some degree be underspecified.

Stalnaker gives the obvious answer to this objection: an infinitely finely graduated space of possibilities may be partitioned to various degrees of finegrainedness, by ignoring some differences and paying attention to others.

[T]here are are surely an infinite number of possible worlds compatible with anyone’s belief state. But a believer’s representation of a space of possible worlds need not distinguish between them all. Just as a finite perceiver may see a space which consists of an infinite number of points, so a finite believer may represent a space of possible worlds which in fact consists of an infinite number of possible worlds.

We have here a clear distinction between a space of metaphysical possibilities, very large and very detailed, and a representation in the mind of some believer.

[Sta84, pg. 69]
Chapter 1 · Introduction

which is much more limited — a notion which will recur throughout this dissertation. Another recurring notion is that our finite believer is unaware of some distinctions which are (potentially; metaphysically) present.

This idea has been given some formal treatment. [Hul02] noted the connection with partition semantics for questions, and Seth Yalcin devotes a chapter of his recent dissertation [Yal08] to the idea. Yalcin calls a proposition “accessibly believed” if it makes distinctions that the believer is aware of, and “implicitly believed” otherwise (see Figure 1.2).

As the title of Inquiry suggests, Stalnaker is predominantly concerned with conscious belief and investigation; similarly, and most naturally, when he applies this theory to a representation of linguistic context it is mainly the representation of explicit and conscious epistemic states that he considers. This is not to say that he ignores implicit or unconscious belief. Indeed, one benefit of a possible worlds representation for belief is that implicit beliefs can be given a very natural characterisation: the bus driver who stamped your ticket this morning believes that you are not a disguised lizardman from Mars, not in the sense that he is aware of this possibility and has rejected it but simply because it does not hold true in any of the worlds in his belief state.

However there is a distinction to be made here which Stalnaker does not appear to find particularly important. Some implicit beliefs (that you are not a lizardman from Mars, that Big Ben is larger than Frege’s left earlobe) are uncontroversial whether attended to or not. Others, though, seem unstable when they are given explicit attention. Here is an example.

Example 1.1: Walt’s interview. Walt hauls himself blearily out of bed at eleven on a Saturday morning. He’s staring into (not yet drinking) a cup of coffee when Perky Pat waltzes into the kitchen. “Back so soon from your interview? Did it go well?” she asks. With horror he realises that he is already half an hour late for a job interview on the other side of town.

No-one should have any difficulty understanding the epistemic condition Walt finds himself in. It may seem strange to refer to his attitude as a ‘belief’ that he does not have an interview to attend. However it is incontrovertible that he behaves as if he believes he has nothing to do that Saturday morning. In that sense we should be willing to say that he has an implicit belief: in all the worlds he is actively considering as possibilities he has nothing better to do on a Saturday morning than recover from Friday night. However equally clearly, when this implicit belief is brought to his awareness (or explicit consideration), it is immediately discarded.

There are also cases intermediate between lizardmen and forgotten appointments. Suppose I ask you if you’re sure you locked your bicycle (or your car, or your front door) this morning, or whether you turned off the gas after
cooking dinner last night. If you are somewhat forgetful (as I am) you may have to admit that you’re not certain that you have done so. If you are habitually forgetful, or merely somewhat paranoid, you may even be prompted by the question to quickly go and check, just to reassure yourself.\(^2\)

All these implicit beliefs have a kind of (intuitively) negative character: they could all be described as beliefs that something abnormal (a job interview, a mistake in a familiar routine, a house-fire) has not happened or is not happening. Here is another example of such a ‘negatively characterised’ implicit belief, of a slightly different kind.

**Example 1.2:** Walt’s keys. Ten minutes later Walt has endured a blistering rebuke and rescheduled his interview. He brushes his chin, shaves his teeth, and hasn’t time to sort it out: he’ll have to run a few red lights as it is to make the new appointment on time. And where are his keys? Not in his pocket, not on the nail behind the door, not beside the phone, he’s searched all the normal places three times. He’s staring again

\(^2\)There is a distinction to be drawn here between **practical belief** and **theoretical certainty**. If I ask if you’re certain that there isn’t a fire starting at this very moment in a far-off corner of the building, you would be displaying evidence of a mild psychological dysfunction if you felt compelled to reassure yourself by checking — this even though your belief that there is no fire is in a sense even less well-grounded than your belief that you have indeed locked your bicycle, even if you cannot remember doing so. Nonetheless, in either case there simply is no uncertainty (be it practical or merely theoretical) until attention is directed to the question.
into his coffee, hoping for inspiration, when Pat helps him out for the second time that morning. “Did you leave them in the car when you came in drunk last night?” Walt slaps his forehead (immediately wishing he hadn’t) and runs to the car. Alas, the keys are not there either; Pat calls him a taxi (and an idiot), he is late to the interview and doesn’t get the job, and later that afternoon she finds the keys in the dishwasher.

Again, the various epistemic states Walt passes through should be familiar, in general form if not (I hope) in specific detail. But if we are to represent them according to Stalnaker’s theory we find some interesting difficulties.

The question is where in Walt’s epistemic state we should place a world in which the keys are in the car (for this is clearly a “way things might have been”). If it is included as an individuated possibility in his belief set we are unable to explain his behaviour: even if he is not certain if that possibility is the actual one, he should certainly act to investigate it.

Perhaps we might combine that world with others in which the keys are not in the car, in the same way that an infinite space of worlds can be subdivided into a finite number of distinct possibilities by aggregating sets of worlds whose differences the agent does not attend to. But intuitively this is not the kind of case that prompted Stalnaker to make this suggestion. We might happily aggregate possibilities in which the keys hang from the ignition at slightly different angles—these are distinctions that we can imagine that Walt does not make—but their presence or absence in the car seems such a salient distinction that we should be reluctant to say that Walt does not distinguish the two possibilities from each other.

It seems we must exclude this world from his belief set entirely, but on what grounds? After checking his pockets Walt has good grounds for excluding the world in which they contain his keys from the realms of possibility, but obviously this is not the same kind of case. We can see an immediate difference too in his behaviour after hearing Pat’s suggestion: if she had said rather “Did you leave them in your pockets?” he would simply answer “No, I’ve checked there.”

What is missing here is a distinction between two different ways of excluding a world from one’s epistemic state: one can have examined it and ruled it out with evidence (as when Walt checks his pockets and establishes that the keys are not there), but one can also have failed to attend to it in the first place. We can fail to attend to entire possibilities as well as to the distinctions between them; failing to attend to whether p conflates p-worlds with not-p-worlds, but failing to imagine the possibility that p means that only not-p worlds are available to be wondered about.

Just as in the first example, it is strange to claim that Walt ‘believes’, before Pat’s helpful interjection, that the keys are not in the car. However by describing the same state in different terms we improve matters considerably: what he
believes, stated positively, is that the keys \textit{could only be} in his pockets, by the phone, or on their nail beside the door. This belief of course entails the other, but being aware of one need not involve being aware of the other; this is what makes Pat’s question (drawing attention to the possibility he is overlooking) helpful.

\section{2.1 \hspace{1em} A failure of attention}

This is an intuitively satisfying solution to describing Walt’s predicament. He has failed to consider the car as a possible hiding place for the keys, and what Pat achieves with her question is exactly making him aware of this possibility. As with finegrainedness, Stalnaker has anticipated this possibility, but he doesn’t seem to make a firm distinction between ‘beliefs’ of this kind, held only due to unawareness, and of the more common and reliable sort.

Discussing riddles, he writes:

\begin{quote}
More interesting than the case of propositions believed but too obvious to be noticed are those propositions taken for granted only because they are not noticed. With riddles and puzzles as well as with many more serious intellectual problems, often all one needs to see that a certain solution is correct is to think of it—to see it as one of the possibilities. […] One has beliefs, or presuppositions, which exclude the correct answer. \textup{[Sta84, pg. 69]}
\end{quote}

This seems right for joke riddles such as “What is brown and sticky?”\footnote{A stick. Most native English speakers will have encountered this ‘joke’ on the school playground; apparently the characterisation of excrement as sticky is less salient for speakers of other languages. I apologise to any readers who through no fault of their own failed to get the joke and now feel left out.} In this case we can even point to the belief at fault: that “sticky” describes the property of sticking to things, rather than that of being similar to a stick.

Perhaps the same kind of description can be applied to Walt’s predicament; after all, this is in a sense what the belief that the keys are only in one of the three places he searches comes down to. What about his implicit belief that he has no job interview that Saturday morning though? Here there doesn’t seem to be a determinate belief (or presupposition) which \textit{gives rise to} the ‘belief’ that there is no interview, unless it be simply that belief itself. Stalnaker uses the term “presupposition” (with a technical definition that need not concern us), covering both ordinary (explicit) beliefs and things that we might be less comfortable calling ‘belief’ without some sort of hedge. Stalnaker, that is, emphasises the similarity of the two cases by choosing a term suitable for both.\footnote{It should be remarked that a Stalnakerian presupposition is primarily a \textit{public} attitude used in explicating conversational behaviour; I am mixing terminology from his investigations of individual belief and of shared conversational context, for the sake of simplicity.}
I want instead to emphasise the differences between them. I will use the term assumption (introduced in this context in [FJ07]) for a Stalnakerian presupposition (or ‘belief’, with scare quotes) of this character: one that is held due to lack of awareness of alternative possibilities rather than due to consideration and evidence. In describing Walt’s epistemic state we need to distinguish between possibilities that are excluded by assumption (the keys being in the car, or the dishwasher) and those excluded by evidence (the keys in his pocket, by the phone, and so on).\(^5\)

3 · Assumptions

Walt is aware of, entertains, or attends to certain possibilities, while others he ignores; some of those entertained are then ruled out on the basis of evidence.\(^6\) These eliminated possibilities have a different status to those that have never been entertained at all; they are available for examination (“No, I already looked there”) in a way that non-entertained possibilities are not.

3.1 · Sentential awareness

One key difference between assumptions and ‘proper’ beliefs is of course whether they are consciously held or not. We can imagine asking Walt to describe all his beliefs and uncertainties regarding his plans for Saturday. If he did so he would never mention any interview: not to express his ‘belief’ (assumption) that he does not have one, and not even to express the tautological belief that he either has an interview or doesn’t have one. By contrast we can imagine that if he had had some training in logic (and was not so hung over) he might be led to say “I don’t have work today, so I suppose logically speaking I either do or don’t have work today.”

Suppose we could successfully elicit a complete recital of Walt’s conscious beliefs.\(^7\) We can imagine this, idealistically, as a consistent set of sentences in some formal language. The suggestion I want to make is that “interview” is not a term in this language (at least before he has realised his mistake); switching examples, while searching for his keys the term “car” does not appear, while “key” most certainly does. We should think of Walt’s lack of awareness of these

---

\(^5\) We must be careful how we phrase descriptions of assumptions, if we are not to give the wrong impression: saying that “Walt assumes the keys are not in the car” wrongly suggests that this assumption has some distinct status in his epistemic state. In fact it follows only as an entailment from a more general assumption that they are not anywhere but the three places he expects to find them, or equivalently but positively, that they can be only in one of the three places he expects them to be.

\(^6\) Again we must distinguish between ‘properly’ ruling out (the foundation of knowledge, philosophically considered) and ‘practically’ ruling out (the foundation of ordinary belief).

\(^7\) That this is probably impossible in finite time is not very interesting. More relevant for the current discussion is the practical observation that articulating some of these beliefs may after all prompt Walt to remember the interview, changing his epistemic state through the very act of trying to describe it.
possibilities as restricting the language he has available for self-ascription of beliefs.

If we can spell this out in detail, we can keep a possible-worlds representation of belief, while making the distinction between conscious beliefs and assumptions: both are propositions entailed by the worlds held possible according to an agent’s epistemic state, but conscious beliefs are describable in his language of belief (self-)ascription while assumptions are not.

Indeed, we can achieve even more than this: the same language will express precisely how finegrained the agent’s representation of possibilities is. Two (entertained) possibilities can be distinguished by Walt, according to this story, if he can say what would make one but not the other hold. His language of conscious belief establishes which possibilities can be distinguished, while his unconscious assumptions exclude some from consideration entirely.

Suppose Walt is not attending to the question whether \( p \). This might be for either of two quite different reasons: he might be indifferent between \( p \) and not-\( p \) alternatives (perhaps in the normal sense of the word, or perhaps because he simply is not conceptualising the difference between \( p \) and not-\( p \), or he might assume that the question is already settled (that is, he either assumes that \( p \), or that \( \neg p \)). The result of calling attention to \( p \) in both cases is the same: he becomes aware of both \( p \) worlds and not-\( p \) worlds as distinct ‘live’ possibilities (the conscious belief attitude he holds to them —whether he gives them any credence—is a separate matter). However the two mechanisms look different: in the first case a more finegrained distinction is made between possibilities that were previously considered equivalent, while in the second case a genuinely new possibility is brought into the light of conscious consideration.

It may be quite difficult to observe (at least with any certainty) the first kind of awareness dynamics. Having the capacity to make a distinction, after all, is no promise that the distinction will in fact be reflected in any observable behaviour. The second kind of dynamics, however, are often extremely visible; Walt’s forehead-slap is the typical sign of an assumption being overturned.

3.2 · Overturning assumptions

Why is mentioning the car enough to overturn Walt’s assumption that the keys are not in it? This effect shows a second important distinction between assumptions and conscious beliefs: consciously held beliefs are typically justified by various kinds of evidence, while assumptions need not be. If you (consciously) believe \( P \) and are confronted with evidence that \( P \) is not after all the case, you face a difficult task. Revising your belief that \( P \) will likely also require changes to a wide range of attendant beliefs, especially those that partly justified or were justified by the belief that \( P \). This difficulty has been recognised in the formal literature on belief revision [AGM85]. Revision is contrasted with monotonic belief update, in which new information is learned that is consistent with the
previous epistemic state; belief update is easily represented in a possible worlds semantics, simply as the elimination of worlds, but belief revision requires a much more complex operation.

The overturning of an assumption has something of an intermediate character: it resembles belief revision in a formal sense (a previously held conviction is overturned, which cannot be represented by elimination of possibilities) but intuitively it is much more similar to belief update. The key property here is the ease with which an assumption is let go: belief revision involves a rearrangement of a whole network of attendant beliefs, while assumptions seem to be overturned more or less in isolation.

This is to be expected, once we recognise what kind of ‘belief’ an assumption is. A belief is formed consciously by eliminating possibilities according to evidence; if that belief is to be overturned then the evidence that lead to those eliminations must also be reexamined. But an assumption is not based on evidence at all: it is based, in fact, precisely on not considering all the available evidence. Part of the difficulty of belief revision is the attitude to propositions that are not direct consequences of the revised belief but are also not independent: those that provide support or evidence for a proposition now deemed false, and that therefore must be considered suspect. These simply are not present in the case of an assumption, which makes the revision process much simpler.\(^8\)

If assumptions are not justified by evidence, though, where do they come from? Intuitively it is clear that not just any proposition is a plausible candidate to be assumed. If the explanation of Walt’s epistemic changes had rested on his assuming that the keys were in the dishwasher, then either the explanation or the use of the term “assumption” would have to be questioned. But what makes “The keys are not in the car” a plausible assumption while “The keys are in the dishwasher” is not?

### 3.3 · Where assumptions come from
Assumptions come fundamentally from a failure of imagination. We are aware of fewer possibilities than in fact exist, and whatever we mistakenly ignore we have assumed away. It might seem from this description that we can easily ‘read off’ a specification of assumptions from checking what the agent does not attend to: Walt does not attend to the possibility of his interview, and so assumes it does not exist; he does not attend to the possibility that his keys might be in the car, and so assumes that they are not.

What makes Walt’s assumptions deceptively simple-looking is that they

---

\(^8\)Formally speaking we have to \textit{add} worlds but never to remove any that we were entertaining. After an update by belief revision proper the new set of worlds may be disjoint from the old one, and it is the complicated relation between these disjoint sets that makes this process difficult to describe formally.
have a clearly negative character: there is some particular thing he is not thinking about (an interview; the car), and his assumption roughly corresponds to the absence of that thing (more about this in a moment). But there are plenty of assumptions that do not have this negative character. Every day when I arrive at work I assume that my key will open the door of my office; I likewise assume that the same key will not open any other office in the same hallway. I assume that my computer is still on my desk and running, and that I can log on to do my work. These are all assumptions in the technical sense I intend: they are implicit beliefs that I act on without any conscious attention to the concepts involved or to whether the beliefs are reasonable or not.

The existence of these ‘positive assumptions’ means that we cannot simply read off assumptions from unawareness: just because I do not attend to the question of whether this particular key will work in this particular lock, you cannot immediately tell whether I assume it will (the lock of my office) or it won’t (someone else’s office).

Still there seems to be a distinction between these cases and Walt’s assumption that his keys are not in the car. We might say that the former are assumptions due to induction, while the latter assumption is due to limited awareness of objects.

3.3.1 · Awareness of objects

Walt’s situation, after looking for the keys in every place he can think of, is quite likely familiar to the reader, who will recognise the feeling that accompanies it: a nagging sense that if one could just remember a few more places the keys might be, one would immediately know where they are. The things being ignored here are not primarily propositions, cognitively speaking, but objects: possible hiding places for the keys.

This distinction need not, per se, be represented in our formal theory. Stalnaker has argued [Sta84, p. 61] that ‘aboutness’ does not need to be encoded in the structure of a proposition: the fact that the sentence “Socrates is mortal” is about Socrates does not need to be explicitly represented in the structure of its meaning; the set of worlds in which Socrates in fact is mortal (the proposition, in other words) will do just fine. In the same way, we can use the possible worlds machinery to represent attending to or ignoring an object: the ‘aboutness’ of that attention need be nothing more than a constraint on which worlds are entertained and which assumed away. Nonetheless, in this case we cannot entirely ignore the cognitive intuition, for we need to define our agent’s language of self-ascription of belief: it is here that the absence of a term “car” appears.

Aboutness is not done away with completely, of course. That “Socrates is mortal” is about Socrates is encoded in the manner by which we attach exactly that proposition to exactly that sentence. What is given up is the idea that this relation is directly visible in the meaning of the sentence itself.
If “car” is not in Walt’s conscious vocabulary of awareness, then he attends to no propositions ‘about’ the car. But we must still specify which beliefs he holds, if only implicitly, about these propositions. That is, we must justify the absence of any worlds in which the keys are in the car from his assumption set. Here the criterion seems to be relevance: whatever objects Walt isn’t aware of, he assumes to be (and thus to have properties making them) irrelevant to his problem of finding the keys. It is this that justifies excluding the world where the keys are in the car from his conscious epistemic state, just as it justifies excluding the world where the keys are in the sugar jar: by implicitly judging the sugar jar and the car irrelevant (by not attending to them) Walt has ensured that he will not consider the possibility that they are hiding places for the key.

One might reasonably ask what is cause and what effect here: does Walt ignore the car because he considers it irrelevant, or does he consider it irrelevant because he has not attended to it properly? In fact neither perspective seems quite right; the connection between irrelevance and unawareness has more the nature of a consistency constraint than a causal relation.

What about other properties of the car, such as its colour (likely irrelevant for the key question) or even its existence, or at least its presence in the garage (which in fact is quite pertinent)? Assumptions about these properties, I suggest, are settled in the same way that the assumption that your bicycle is locked gets formed: by induction.

3.3.2 · Inductive assumptions
A large number of assumptions do not seem to derive from ignoring objects. Instead, they are ‘properly propositional’ in the sense that the natural way to express what is being ignored is “the possibility that (some proposition) p does (or does not) hold”. These seem to come from various kinds of inductive reasoning, which neatly explains our intuitions about which assumptions are plausible and which are not.10

There are assumptions that are clearly induction from previously observed instances; the assumption that the building you are currently inside is not on fire, for instance, that the company you work for has not gone out of business overnight, and so on. (It is important for these examples to recall that ‘assumption’, in the sense I use the term, refers only to beliefs held due to inattention and unawareness. In both these cases we might continue to believe the proposition under consideration even after attending to it, but we should have to admit that we had no grounds but those of precedent to do so. These are ‘assumptions’ in the normal sense but not in the technical sense I intend.)

The assumption that “sticky” refers to the property of sticking to things

10“Inductive reasoning” is perhaps too strong: assumptions take in ‘inductive’ generalisation from a single observed instance, which hardly deserves to be called reasoning at all. The ‘inductive assumption’ is that what the agent has seen is everything there is to see.
might come under this heading also, or it might represent a more general
tendency we have to resolve ambiguities by fiat rather than with full attention
to the range of possibilities. This is especially visible in our attitude to language.
Most sentences are strictly speaking ambiguous in some way or another, but in
most natural contexts one reading is strongly preferred (this is not a coincidence:
if we are to communicate effectively we have to use expressions that can be
easily disambiguated to do it with). The observation linking this to awareness is
that typically this potential ambiguity does not reach the level of consciousness,
not even as a recognition that an alternative potential interpretation is being
rejected. In the case of the riddle this resolution is probably based on two
distinct mechanisms: that the interpretation being assumed is the more common
one, and that it gives rise to a plausible (if scatological) answer to the question
and thus doesn’t trigger any extra effort towards exploring more unlikely
possibilities.

Another kind of inductive assumption might be called “There it’s like
here”: the assumption that conditions are similar in distant places or times
(apart from whatever distinctions are being consciously attended to, of course).
Inexperienced travellers are constantly surprised by the variations in standards
of politeness around the world; actors in period dramas are tall, unmarked by
disease, and clearly bathe regularly, and we don’t notice any incongruity.\footnote{Two
days after I wrote this Brendan Adkins asked on his blog “Not Falling Down”,
Why do the same people who complain about sound in space, the proper rigging for
catapults, or the relative strength of a katana versus a broadsword never mention the
way women in medieval or even Victorian settings are always depicted with shaved
legs?
Inattentiveness and assumption is not the only possible answer to the question, but it is a plausible
one.}

I will not have much to say about the dynamic process of forming assump-
tions; whether approached philosophically or psychologically the formation
of inductive generalisations is a vexed question that I will sidestep as far as
possible. But one potential source of assumptions seems clear: any belief at all
can be converted into an assumption if it goes unchallenged and unexamined
for long enough. Having eliminated a possibility (by examination and evidence)
we do not in fact cling to it for the rest of our lives; in time it fades from our
consciousness and what was once a conscious belief becomes nothing more
than an assumption.

3.4 · What are assumptions for?

We notice assumptions most often when they are overturned, or when they
turn out to be false. Remember, though, that a finite epistemic state supports
an infinite number of assumptions; the vast majority of these will never be
attended to and would be quickly ratified, rather than overturned, even if
they were. (Big Ben is larger than Frege’s left earlobe. My feet have toes, my hands have fingers. I am not a butterfly dreaming it is a PhD student.) In this dissertation, however, we will be particularly interested in the cases where assumptions go wrong: it is these that provide the interesting pragmatic and semantic possibilities that make this notion so important for describing and explaining the various ways we use language. Looking at these cases one might start to wonder whether having assumptions is a smart idea after all. Might we be better off if we could do without them? Life would certainly get a lot simpler, wouldn’t it?

The first partial answer is of course that we cannot do without inattentionness, since the space of possibility is too large and our minds too limited to comprehend it all. However, besides being a negative and rather trivial statement, this does not mean we need assumptions: we might get along fine just using finegrainedness (as in [Yal08]) to carve the world up into manageable chunks.

A more positive answer is that assumptions make reasoning much easier. In deciding how best to fetch that banana we need not consider the possibility that it will fly out of reach when we get close to it; given the unlikelihood of the possibility and the fact that if it did eventuate we would have no sensible strategy for dealing with the problem, not having to consider it seems a mercy. In other words, assumptions do not just reduce the space of possibility so that it can fit inside a believer’s head; they select the possibilities that are most relevant for the problem-solving the believer will have to perform, those that are likely enough to be worth attending to, and also those that the believer can, with sufficient planning, deal with if they eventuate.

If this is what assumptions are for, then, we can distinguish two properties they need to have if they are going to do their job properly. They need to be generally true, and they need to be defeasible. (These two properties, in turn, suggest why assumptions can typically be described as inductive generalisations.)

That assumptions should be generally true should by now be obvious. Excluding the correct answer by assumption, as in the case of riddles, only makes life difficult. It is impossible to reason outside one’s own assumptions (if you are reasoning about the possibility that \( p \), you are aware of it). This means that if we are ever confronted with a genuinely unentertained possibility, we will have to very quickly construct a plan to deal with it. Clearly it will be more comfortable to avoid this kind of last-minute plan formation as much as possible: we should hope that most of our assumptions turn out to be acceptable, most of the time.

And indeed, evolution seems to have equipped us with a magnificently effective machine for forming just the right inductive judgements in just the
right ways, so that they tend to be true. In fact it can be argued that all of our beliefs rest upon a foundation of assumption rather than of evidence strictly conceived. The fact that our beliefs tend in the main to be true ones speaks strongly for the reliability of our assumption-formation apparatus.

The other requirement I stated is that assumptions should be defeasible. This is strictly speaking a requirement not on the notion of assumption but on how they are recorded and represented in the heads of believers; they will by their very nature be defeasible, but they need to be represented in a way that makes it easy to revoke them. This is both an observation (assumptions are easily revoked, as I argued when introducing the notion) and a prediction: assumptions must be revokable with a minimum of difficulty, when contradicted by evidence. This is because they may be confounded by the world; if this occurs the believer should give up assumptions before giving up belief in more tangible evidence. Walt might check twice that the keys are not hanging on the nail behind the door, but if he repeatedly gives up the belief that the evidence of his eyes is correct, rather than the assumption that the keys cannot be somewhere else, we would be justified in concluding that he suffers some psychological disturbance.

Attention or awareness?

I have used “attention” and “awareness” more or less interchangeably, and will continue to do so in the rest of the dissertation. They carry different associations which allow for slightly smoother exposition: “being aware” is a passive state while “attending to” is an active action, and (as we will see) many linguistic

---

12Wittgenstein advances a similar position in On Certainty, with the added observation that certain of these assumptions cannot even meaningfully be questioned, without undermining the sense of the language in which the questioning is attempted. Or rather, perhaps, it is not possible to question all of these assumptions at once: some basis of certainty is required, otherwise a doubt cannot even be meaningfully expressed.

79. That I am a man and not a woman can be verified, but if I were to say I was a woman, and then tried to explain the error by saying I hadn’t checked the statement, the explanation would not be accepted.

80. The truth of my statements is the test of my understanding of these statements.

81. That is to say: if I make certain false statements, it becomes uncertain whether I understand them.

83. The truth of certain empirical propositions belongs to our frame of reference.

96. It might be imagined that some propositions, of the form of empirical propositions, were hardened and functioned as channels for such empirical propositions as were not hardened but fluid; and that this relation altered with time, in that fluid propositions hardened, and hard ones became fluid.

341. That is to say, the questions that we raise and our doubts depend on the fact that some propositions are exempt from doubt, as it were like hinges on which those turn.

[Wit69, §§79–81,83,96,341]
utterances are best thought of as drawing attention in this active sense.

"Unawareness" can also denote a lack of conceptual grasp; the man on the street is unaware of the principles of higher mathematics, and Darwin was unaware of the existence of DNA when he proposed his theory of natural selection. The examples I am mainly interested are more humble: they concern concepts that the agent in principle understands, but has not thought to apply to the case at hand. "Inattention" is perhaps a more appropriate term for these moments of forgetfulness and lack of insight.

However there is at least one potential confusion risked by using "attention" in this way, which needs an explicit comment. In linguistics attention is often a relative quantity, allied with notions such as salience. When the ‘focus of attention’ is directed at something in particular, it is withdrawn from whatever else might have been under discussion. This is emphatically not the sense in which I intend the word. I have tried to use “awareness” particularly in cases where this interpretation seems tempting, since ‘becoming aware’ is much less seductively misleading than ‘shifting attention elsewhere’. I have of course also used “awareness” where it is a technical term in existing literature, as in the three models discussed in the rest of this chapter.

4 · Three models of awareness

We have seen how thinking about formal semantics and philosophy of language can lead us to notions of awareness and assumption. There is a growing field of research investigating the notion of awareness in epistemic logic (and related fields such as certain branches of economics), without any attention to language at all; this field has so far largely ignored the notion of ‘assumption’, which is so crucial for understanding the behaviour of agents suffering from unawareness.

In the rest of this chapter I will survey two very influential systems based on very different notions of unawareness (the logic of general awareness of [FH88] and the subjective state-space approach of [HMS06]), paying particular attention to the space they leave for a representation of assumption. A third proposal that is gathering support is the object-based model of [BC07], which to some extent bridges the gap between the first two models; in particular, while the object-based semantics is a closer formal cousin to the logic of general awareness, the natural notion of assumption it suggests is more closely connected to the state-space approach. These three papers inform the structure of the rest of the dissertation: very broadly speaking, Part I extends models along the lines of [FH88], while Part II combines the object-based approach of [BC07] with the subjective state-spaces of [HMS06].

4.1 · The logic of general awareness

Fagin and Halpern’s paper [FH88] is generally credited as the origin of the current research field of unawareness in epistemic logic. The paper deals with
the problem of logical omniscience: the property of standard possible-worlds analyses of knowledge that an agent knows all logical consequences of her knowledge. This leads to unintuitive results when we match our models against the cognitive limitations real reasoners suffer; for instance the naive prediction from such a model is that a reasoner should know all logical tautologies and mathematical truths.

Fagin and Halpern point out that ‘logical omniscience’ is best thought of as an umbrella term covering a number of distinct problems, and they propose several systems dealing with different aspects of the problem. The one we are concerned with here is the logic of general awareness (for this chapter the LGA, found in their Section 5).

The essence of this system is a distinction between implicit and explicit belief. ‘Implicit’ belief is a new name for the familiar notion of belief interpreted on Kripke structures: a box modality on a serial, transitive and Euclidean accessibility relation (giving rise to the logic KD45). Explicit belief, on the other hand, is modelled via awareness. An awareness function for each agent assigns to each world the set of sentences the agent is aware of (note the syntactic nature of this component). The agent explicitly believes a formula $\phi$ if she implicitly believes the proposition that $\phi$ expresses (standard belief in Kripke structures) and she is aware of the formula. (Among other things this means that every explicit belief is also implicit. While formally convenient this grates somewhat on the natural usage of the terms: typically we would say “implicit belief” to mean precisely a belief that is not explicit. I will follow the ‘logician’s terminology’ of Fagin and Halpern here, and in similar cases that emerge in the later models, and distinguish beliefs that are not explicit by calling them ‘purely’ or ‘strictly’ implicit when the distinction is important.)

The representation of awareness is left completely unconstrained: the awareness function can select any set of sentences whatsoever at each world. (It is not required, for instance, that the agent be aware of $p \land q$ whenever she is aware of $q \land p$.) For Fagin and Halpern this is an advantage, since it gives their system the flexibility to represent many different kinds of ‘unawareness’ (in a rather broad sense) leading to different kinds of failures of logical omniscience. For our purposes though it seems we want to interpret awareness as something like linguistic resource: that the agent has the conceptual vocabulary to describe explicit beliefs is what distinguishes them from those that are strictly implicit. In particular, Fagin and Halpern discuss a set of constraints on the awareness function that reduces awareness of sentences to awareness of atomic formulae (primitive propositions): the agent is aware of a complex formula iff she is aware of all the atomic formulae that occur within it.\(^\text{13}\) All the models I propose

\(^{13}\text{They also treat possible interactions between belief and awareness; for instance, an ‘awareness introspection’ property that the agent knows what she is aware of can be guaranteed by requiring}
in this dissertation have this ‘combinatorial’ property.\textsuperscript{14}

It seems that the \textsc{lga} provides a fairly close fit for the notion of awareness we need. It is less successful, unfortunately, in representing assumptions.

\subsection*{4.1.1 Assumptions in the \textsc{lga}}

It might seem that implicit belief is precisely what is intended by the intuitive notion of assumption (an assumption is distinguished from a ‘real’ belief by being strictly implicit while the latter is explicit: the purely syntactic distinction of [FH\textsc{88}]). However there is a further \textit{semantic} distinction that can be drawn, which becomes important for representing changes in awareness over time: the distinction between worlds that the agent ‘has in mind’ and those that she does not.

In 1957 the \textsc{bbc} current affairs program \textit{Panorama} reported on a bumper spaghetti harvest in Switzerland.\textsuperscript{15} Viewers doubtless formed the conscious, explicit belief that the mild winter had been good for the growth of spaghetti, since that was what the programme announced. That this belief rests on the assumption (entirely implicit) that the \textsc{bbc} reports only the truth (the broadcast was of course an April Fools’ joke) does not make the belief itself any less explicit.

Suppose we were to represent this scenario in a model for the \textsc{lga}. The model representing a viewer named Vera is shown in Figure \ref{fig:intro-lga}: she (strictly) implicitly believes the \textsc{bbc} is entirely trustworthy ($\neg j$, where $j$ stands for “joking”), and explicitly believes that spaghetti is grown in the south of Switzerland and dried in the alpine sun ($s$).

The problem with this picture is that it does not distinguish between Vera’s attitude, at the actual world $w_2$, to $w_0$ and to $w_2$ itself. Both are excluded from her possibility set, but intuitively for very different reasons: $w_2$ is a possibility she is not even imagining (despite its being actual), while $w_0$ is a possibility she imagines but has ruled out (on the basis of the \textsc{bbc} broadcast). This intuitive difference shows in a difference in behaviour, once we take changes in awareness into account. If we should politely draw Vera’s attention to the date, raising the possibility that the \textsc{bbc} is having a bit of fun, the worlds $w_0$ and $w_2$ do not behave the same way in her resulting epistemic upheaval. We expect her to come to hold $w_2$ possible exactly because it was previously excluded by an assumption; she should not come to hold $w_0$ possible, because it was excluded

\begin{flushleft}
\textsuperscript{14}Under this constraint we can think of the agent’s awareness as providing a partition on logical space representing a level of finegrainedness, as in Figure \ref{fig:partition} on pg. 5. This is not in general possible for structures of the \textsc{lga}; for instance the agent might explicitly believe both $\varphi$ and $\psi$ but only implicitly believe $\varphi \land \psi$, which cannot be represented via such a partition.
\end{flushleft}

\begin{flushleft}
\textsuperscript{15}The segment can be viewed online at the \textsc{bbc} website: \url{http://news.bbc.co.uk/onthisdhday/hi/dates/stories/april/1/newsid_2819000/2819261.stm}
\end{flushleft}
Three models of awareness

Figure 1.3: Vera, a victim of a BBC April Fools' hoax in the 1950’s, modelled using the logic of general awareness. Proposition letters $j$ and $s$ stand respectively for “joking” (that the BBC broadcast was not in earnest) and “spaghetti” (that spaghetti is grown on trees). Arrows represent her accessibility relation; the actual world is $w_2$ and the ‘thought balloon’ represents her awareness function (shown only at that world): she is aware of $s$ but not of $j$.

by conscious consideration of the evidence.\textsuperscript{16}

The tactic I will employ in the following chapters is effectively to define two accessibility relations: one representing the possibilities that the agent ‘has in mind’ (regardless of her attitude of belief or scepticism towards them) and one representing her beliefs (regardless of whether these are implicit or explicit). Any world she does not ‘have in mind’ is automatically excluded from her belief set; the worlds she ‘has in mind’ define her assumptions, in the sense that Vera’s belief that there are spaghetti trees in Switzerland is based on her assumption that the BBC broadcasts only factual reporting.

Segue: formal epistemic economics

The approach of [FH88] takes implicit belief and awareness as primitive notions, and then derives explicit belief from them. An opposing theme in the formal epistemic economics community\textsuperscript{17} is to take knowledge as primitive (roughly corresponding to explicit belief, although as we will see the identification of the two is not without problems), and defining (un)awareness as a derived notion.

\textsuperscript{16}This point is liable to misinterpretation. I do not mean that the BBC’s announcement should be treated as true (which would certainly prevent Vera from reinstating $w_0$, but would also keep $w_2$ out of the picture). The broadcast itself remains as an objective fact, and that fact is incompatible with the combination of there being no spaghetti trees in Switzerland and the broadcast being truthful. This combination is present in $w_0$, and thus ruled out by the objective evidence (the existence of the broadcast), even after an awareness change.

\textsuperscript{17}This is the best term I could find for the sub-field of economics exemplified by the biannual conference TARK (Theoretical Aspects of Rationality and Knowledge) and results such as the no-agreeing-to-disagree and no-trade theorems [Aum76; MS82]. The field stands in roughly the same relation to general economics as model-theoretic semantics does to general linguistics. I don’t imagine that all “economists” are concerned with common knowledge any more than all “linguists” make use of Kripke structures, but I will continue to use the terms as though this were the case, since these are the only linguists and economists I am directly concerned with.
Chapter 1 · Introduction

The idea begins with the observation that unawareness leads to a failure of negative introspection: if the agent is unaware of $p$, she can fail to know $p$ but not know that she does not know it. Conversely, she is aware of $p$ just if she either knows that $p$ or she knows that she does not know it. $S_5$ knowledge of course supports negative introspection; the standard approach in economics to modelling agents without negative introspection has been the use of non-partitional information structures: Kripke structures for $S_4$. Unfortunately, [MR94] proved that such structures cannot be suitable for modelling unawareness: adding a symmetry axiom requiring that the agent be aware of $\varphi$ just if she is aware of $\neg \varphi$ (eminently reasonable under our ‘conceptual vocabulary’ interpretation of awareness) again yields $S_5$. Modica and Rustichini also investigated a larger class of models where knowledge is given not by an accessibility relation but by a function mapping each event $E \subseteq W$ to the event of knowing $E$. While such models can incorporate symmetry without collapsing to $S_5$, Modica and Rustichini showed that they nonetheless only give rise to trivial unawareness (either the agent is aware of everything or she is aware of nothing).

This negative result has acquired the label “Standard state-spaces preclude unawareness”, after a later paper extending the treatment [DLR98]. This later paper gave explicit attention to a major problem with the knowledge-based analysis of awareness: the treatment of tautologies. Standard models based on possible worlds translate any tautologous sentence to the (one and only) necessary proposition (the entire state space). But on the idea of awareness as representing conceptual vocabulary, the agent might very well be aware of $p \lor \neg p$ while remaining unaware of $q \lor \neg q$. For models with syntactic awareness as a primitive this is of course no problem, but economists seem to have viewed such models with some scepticism because of the ‘pollution’ of semantics with syntax.

Instead, the economics community has turned its attention towards non-standard state spaces, in particular towards models in which not every ‘world’ represents a genuine fully-specified possibility. If a proposition is a set of partial possibilities, rather than full possible worlds, than a partial possibility unspecified for the value of $p$ might reasonably fall outside the proposition expressed by $p \lor \neg p$, thus reinstating the ability to distinguish between even tautological sentences based on the vocabulary employed. The next model we will consider exemplifies the type.

4.2 · The subjective state-space approach

[HMS06] introduces a model (which I will call the hms model) in which some states are associated with partial, rather than total, valuations of the set of atomic

\[\text{An “event” to an economist is a “proposition” to a linguist: a subset of the set of worlds or possibilities.}\]
formulae. The state space is constructed from a lattice of disjoint subspaces, where each subspace intuitively corresponds to the language generated by a particular vocabulary. The topmost subspace contains ‘real possibilities’, or full valuations; worlds in lower subspaces are partial valuations. An agent unaware of $p$ ‘sees’ only worlds in a subspace whose vocabulary does not include $p$, thus whose valuations do not assign a truth value to $p$ (the agent is unaware of $p$ at $w_4$ in Figure 1.4, for example).

The subspaces are partially ordered according to the richness of their vocabularies, and there are projection functions (shown in Figure 1.5 overleaf) saying how a world in a high space ‘appears’ when viewed according to the limited vocabulary of a lower space. (Typically several worlds in any given higher space will project to a single world in a lower space; for instance two worlds differing only in the valuation of $p$ would project to the same world in a space where $p$ was not part of the vocabulary. Each world projects to only one world

19The construction in fact given in the paper makes no formal reference to vocabularies and so on; I follow the presentation of [HMS08]. ‘The’ HMS model is considerably more protean than I pretend here; see especially note 22 on pg. 26 for a more recent variation with quite different properties.
Figure 1.5: Projection relations between worlds. Each world projects (downward) to exactly one world in each lower subspace; it will project (upward) to several in a higher space. Downward projection represents “how the world appears” in the vocabulary of a lower subspace; the world where both \( p \) and \( q \) are true appears as a world where \( p \) is true in the subspace with vocabulary \( \{ p \} \), and where \( q \) is true in the subspace with vocabulary \( \{ q \} \).

Figure 1.6: Part of an \( \text{HMS} \) model showing an upward cone (several subspaces are omitted). The circles all project down to the world labelled \( w \); the event generated by the set \( \{ w \} \) is the upward cone that is the union of the circles.
in each subspace, however, and a world and its projection always agree on the valuation of whatever vocabulary they have in common.)

The lattice structure of the subspaces constrains the accessibility relation: from \( w \) an agent can only see downward in the structure (to spaces with less expressive power) or ‘across’ to worlds within the same subspace as \( w \). Further constraints generalise reflexivity, transitivity, and so on to the lattice structure: if \( w \) sees \( v \) in a lower subspace, for example, \( v \) cannot see \( w \) (because this would be looking ‘up’ the lattice into a space with higher expressive power) but the analogue of reflexivity requires that \( v \) see the projection of \( w \) into the same space \( v \) inhabits.

Instead of just being a set of worlds, an event (or proposition) in this setting is an upward cone through the lattice structure, generated from a single subspace: for some set of states \( B \) lying in a single subspace it contains all states that project into \( B \) (that this is an upward cone follows from a requirement that projections commute down the lattice of subspaces). See Figure 1.6 on the facing page for an example.

Among other things this means that the negation of an event is not simply its set-theoretic complement: such a set would typically not be an event. Instead we generate an upward cone from the complement of \( B \) in its subspace; this process ensures that an agent aware of the event \( P \) (one who sees the subspace generating \( P \)) will also be aware of its negation (the symmetry constraint required by [MR94]).

Awareness is defined based on knowledge (the agent is aware of \( P \) if she knows \( P \) or knows that she does not know \( P \)), and knowledge is given a standard definition: she knows that \( P \) at \( w \) (where \( P \) is an event, i.e., an upward cone) if all her worlds accessible from \( w \) lie within \( P \). However the structure of the subspaces, and the fact that no agent can see ‘upwards’ in that structure, ensures that she only knows that \( P \) from worlds that have enough vocabulary to describe \( P \), that is, from worlds in the same subspace that generates \( P \) or from higher in the lattice.

At first glance this looks a lot like supervaluations. Worlds in subspaces are partial models, and the projection relation tells us which more complete models they may ‘grow into’. However there is an important difference, to do with the very notion of unawareness. Under a supervaluational definition of knowledge, the agent always knows a tautology such as \( p \lor \neg p \) (since it is true at all supervaluations where it is defined); exactly this has to be avoided in a model of awareness. One way to see the distinction is to take the language-oriented definition of knowledge (see Figure 1.7 overleaf). In a supervaluational story, the agent knows \( \varphi \) iff \( \varphi \) holds everywhere in the projection of her belief set on the highest subspace (that is, in all supervaluations of the partial valuations that make up her belief set). In the HMS model, she knows \( \varphi \) iff \( \varphi \) holds everywhere
in projections of her belief set on any (weakly) higher subspace, including the subspace where her belief set itself lives. “Holds everywhere” means *is defined and holds* everywhere; this makes no difference for the supervaluational version, since in the highest subspace all sentences of the language have definite truth-values, but it matters a lot for the hms model: a tautology such as \( p \lor \neg p \) is only defined in the agent’s belief set if she is aware of the proposition letter \( p \).

This has enormous consequences for the representation of assumptions.

### 4.2.1 Assumptions in the hms model

In fact it means that non-trivial assumptions are entirely ruled out. Suppose the agent is unaware of \( q \), and sees only worlds in a subspace with vocabulary \( \{p\} \). She holds possible only the world where \( p \) in fact holds, and she thus knows that \( p \). If we follow the projection relations backwards from her belief set we can collect the sets of *fully specified* worlds (in the highest subspace, with most extensive vocabulary) that project to each of the worlds she holds possible. Whatever knowledge she can acquire by eliminating worlds in the subspace whose vocabulary she can use, it must deal atomically with these ‘knowledge units’: eliminating a world in the lower space eliminates its entire inverse projection set, so no knowledge generated from the lower subspace can ever ‘cut across’ such a set (see Figure 1.8 on the facing page).

But this is precisely what we want assumptions to do! Every world in the subspace with vocabulary \( \{p\} \) is the projection of two worlds in the higher space: one where \( q \) holds, and one where \( \neg q \) does. Think back to Vera and the spaghetti: she is unaware of \( j \) and believes \( s \). To represent her assumption that \( \neg j \), we need to *separate* the two worlds that project to the (partial) world she imagines: the one where \( \neg j \) holds belongs in her belief set, while the one where \( j \) holds does not.

Here is another way of looking at the problem. The set of worlds generating Vera’s knowledge is generated as an upward cone through the lattice of subspaces, so it includes some ‘worlds’ that are more partial than others. Her knowledge is whatever holds at *all* these worlds, and the construction restricts this in two distinct ways. First there is the vocabulary restriction: since some worlds are partial and make no mention of \( j \), she cannot know that \( j \lor \neg j \) (the observation of [DLR98], that unawareness even affects tautologies). We might be tempted to say that she knows this *implicitly*, however, since it is nowhere contradicted in this set of worlds; it is certainly true everywhere in the *complete*...
Three models of awareness

Figure 1.7: Supervaluations compared to HMS structures (in abstract visualisation). The cone from \( a \) represents \( p \lor \neg p \), generated from the subspace where only \( p \) is defined; the cone from \( b \) is generated from the belief set of an agent unaware of \( p \), thus from a lower subspace. Supervaluations quantify only over the worlds in the horizontally shaded region; HMS knowledge quantifies over the entire diagram, including the vertically shaded region where \( p \) (and thus \( p \lor \neg p \)) is undefined.

Figure 1.8: HMS models cannot represent assumptions. Dotted arrows are upwards projections. If the agent is unaware of \( q \) (that is, her set of accessible worlds lives in the lower subspace) then it treats the circles in the upper subspace atomically: each circle is either entirely inside or entirely outside the upward cone generated by her accessible worlds. Her beliefs can never ‘cut across’ the dashed lines.
worlds in the set, which intuitively correspond to the real possibilities still left open. She cannot know that $j$, though, for another reason as well: there are worlds in that set where it is defined but does not hold. If an assumption corresponds to some kind of implicit belief, it will not be found in this set.

4.3 · Object-based unawareness

The object-based model of [BC07] (obu, for “object-based unawareness”) is a first-order model, in which the agent’s awareness of sentences is generated by her awareness of objects. It bridges the gap in various ways between the logical approach of [FH88] and the economic approach of [HMS06].

An event in obu semantics is a pair: a set of worlds (the “sense” of the event) and a set of objects (the “reference”). The reference contains the objects the event is ‘about’; the tautologies $P(a) \lor \neg P(a)$ and $P(b) \lor \neg P(b)$ have the same sense but different references, as you would expect. The model contains an accessibility relation on worlds but also an awareness function, saying at each world which objects the agent is aware of.

In fact this model is in some sense the ‘obvious’ first-order version of the LGA, at least in the case where full sentential awareness is generated by awareness of atomic formulae. (The main interesting new feature is the possibility of quantification: by using models with non-constant domains, Board and Chung can give a formula corresponding to “The agent is uncertain whether there exists some object she is unaware of”, which is a very desirable feature. Even with constant domains, an agent may be uncertain whether any object has the property $P$ while not being uncertain of any (particular) object whether it has the property $P$.)

At the same time, the model is given in the ‘language’ of economics, in terms of events and operators on events rather than a model and an interpreted formal language. This allows a direct comparison with the HMS model, as in a working paper coauthored by originators of both systems [BCS09]. This paper shows, somewhat surprisingly, that obu structures and hms structures are to some extent equivalent. That is, if we take a particular generalisation of hms structures and concern ourselves only with the event structure of the models (ignoring the extra expressive power that quantification brings to the language of object-based unawareness), a model in either system can be transformed into a model in the other which captures the same facts about knowledge and awareness of events.

This is surprising because the obu model, like its close cousin the LGA, allows non-trivial implicit beliefs, while the hms model does not. The intuitive reason for this surprising result is that the equivalence does not take implicit beliefs into account, as these are nowhere defined in the hms system.\footnote{There is a more technical reason also: the hms models do not entirely follow the description I gave...}

Similarly,
[HRo8] showed that HMS structures can represent exactly the same facts about explicit belief and unawareness that the LGA can, but could say nothing about implicit belief.

4.3.1 · Object-based assumptions
Since the OBUN model is such a close cousin to the LGA, it has roughly the same potential to represent assumptions: we have implicit beliefs, but we cannot distinguish between worlds the agent ‘has in mind’ but has ruled out and those that she has not even considered. Similarly, there is no way within the theory to relate unawareness to implicit belief: we cannot say which implicit beliefs an agent should hold if she is unaware of some particular object.

As I argued in the first half of this chapter, there is no general strategy for deriving implicit beliefs or assumptions from unawareness (remember the assumptions of my key unlocking my own office door but not that of my colleagues). However, the particular case of object-based unawareness seems to be different: if I am unaware of some particular object, then it seems perfectly reasonable to say that I assume it does not exist. Certainly this works for Walt and the car keys (and possibly even for the interview, if the interview itself is thought of as an object, rather than a proposition “Walt has an interview”).

In other words, the object-based model offers the best chance to derive assumptions from unawareness. We will still, however, have to do quite a bit of work before we can achieve this: as it stands, the model cannot yet represent the distinction between implicit belief and assumption, which will certainly be needed.

5 · The rest of the dissertation
The work on formalising unawareness so far, mainly in the economics community, has concentrated on the distinction between implicit and explicit belief.
I do not think that this is all there is to the interaction between awareness and belief. An agent’s unawareness of possibilities influences her beliefs in systematic ways that go beyond the question of whether she is conscious of them or not.

I have argued, in a sense, for asking the same kinds of questions about an agent’s disbelief: if a world is not in the agent’s belief set, is it explicitly ruled out, or implicitly excluded because the agent does not ‘have it in mind’? We need the notion of assumption, and we need to ask how assumption interacts with awareness/attention and with belief.

This dissertation is divided into two parts. Figure 1.9 gives an abstract schematic view of the distinction this division embodies. In both pictures the set labelled $A$ represents the worlds the agent ‘has in mind’; how this set relates to the agent’s (un)awareness of concepts is the fundamental problem of the dissertation as a whole (I make therefore no attempt to answer it in the schematic form of the diagrams).

The distinction between the two parts rests on how we treat the possibilities outside the set the agent ‘has in mind’. In Part I (schematically represented in Figure 1.9a), the agent holds definite beliefs about the possibilities within $A$ (the set labelled $B$) but his attitude to what lies outside remains completely unspecified. If his awareness of possibilities increases, he will have to ‘make up his mind’ what to believe; the dotted circles represent potential beliefs he might come to hold, but whether these will be realised can only be seen dynamically, as his state of awareness changes over time.

Part II follows a rather different schema, given in Figure 1.9b. As before, $A$ represents the worlds the agent ‘has in mind’. He has information which is guaranteed to be objectively true (the set labelled $I$); however his interpretation of that information (his beliefs, the shaded region) is influenced by his assumptions. In contrast to the models of the first part, from our external perspective we can see how the agent’s beliefs will develop as his awareness changes (schematically, his information set $I$ is well-defined even outside the worlds $A$ he ‘has in mind’). The interesting feature of these models is rather the relations that are not shown in the picture: between his information and his awareness, and his awareness and his assumptions.
Figure 1.9: Schematic view of the models of Parts I and II.