Information Flow

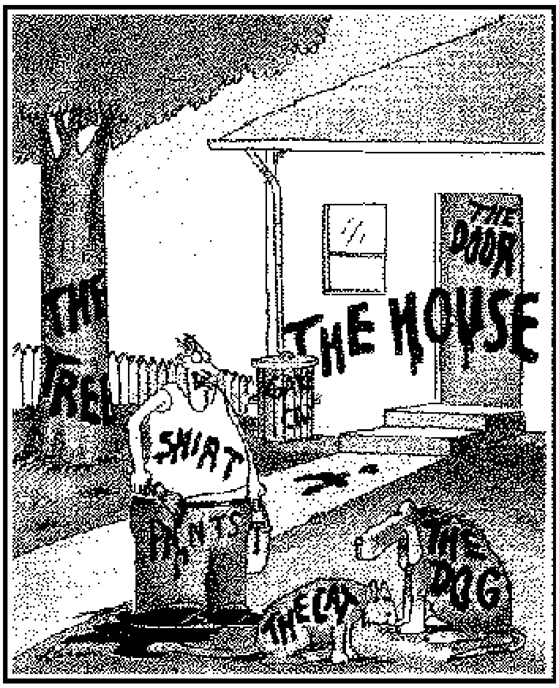
Logics for the (r)age of information

Diploma thesis in Mathematics

Philipp Keller, 2002

Institute for Theoretical Computer Science

University of Berne



"Now! ... That should clear up a few things around here!"

Preface

To mathematically interested philosophers and philosophically interested mathematicians, "information" may appear to be the magic linkage between the different fields, connecting logic and language, models and reality. So it did to me and this is why this study ends on a pessimistic note. Even if I am less than convinced that the high expectations placed on a future substantive and rigorous science of information will pay off in the short term, however, I enjoyed exploring their credentials; and I am the last person to claim that everything worth doing pays off in the short term.

Even if situation theory, due to notational overkill and lack of spectacular solutions to not home-made problems – and due, certainly in part at least, to the changing fashions in philosophy and theoretical computer science –, is widely held to be in decline, at least since Jon Barwise's untimely death, I am convinced that there is something to be learnt of all the work undertaken within this paradigm in the last twenty years. It is highly probable on inductive grounds that some of the motivations behind this approach, some of the problems that emerged for it, and some of the solutions situation theorists have devised, will reappear on stage in due term. Even if situation theory as a whole was a failure (which is certainly too early and too harsh a verdict), there is much to be gained from studying it – or so I believe and hope to show in the following.

Nobody works, nor is able to work, in isolation – and certainly not me. First and foremost I want to thank my friends, who made the last year an agreable one. I also want to thank my supervisor, Prof. Gerhard Jäger, for setting me to work on these problems. The original task he set me, (i) to give a conceptual analysis of the notions of information, data and knowledge and their interrelations and (ii) to apply this analysis to the theory of information flow developed by Jon Barwise and Jerry Seligman in their book *Information Flow: The Logic of Distributed Systems*, sets the agenda for this work. I also want to thank Luca Alberucci for reading parts of sct. 4.5 and my students' advisor Prof. Christine Riedtmann for not having lost her patience with a student who is raising the average time of study.

Contents

Introduction

The diploma thesis at hand is an attempt to pursue some of the ramifications of the formal modelling of some old and notoriously vague notions in relatively new branches of mathematical logic and theoretical computer science – what might perhaps become, one day far in the future, the "information science" so much talked about these days. The notions under study – "information", "knowledge", "meaning" – have been much discussed by philosophers and one is likely to find philosophical remarks about the 'nature' of these concepts in many of the articles written and cited even by more technically oriented protagonists of the field. These remarks are meant to delineate the intended field of application and to justify the general principles of the formal and semi-formal theories. I will try to adjucate their credentials to shed new light on these old and contentious concepts. These remarks typically serve a double purpose: on the one hand, they are supposed to give the intuitive background for the formal theory, while on the other hand they should carve out, in an informal and preliminary way, what the explananda, and therefore the adequacy conditions, of the theories under construction are. While there are obvious coincidences between the informal introductions and the technical development, in terminology at least, it turned out very difficult to test the different theories against some well-demarcated and common background. There just is no bedrock of indisputable and universally accepted intuitions in this field. Agreement risks to stay verbal; very much is up for grabs.

The apparent incommensurability of different theories in the young discipline may be due to different factors. One possible reason lies in the concepts themselves whose 'logic' is said to be laid out. They are vague, and their meaning depends on the meaning of at least some, if not most, of the other concepts in their semantic cluster. No theory of knowledge, e.g., can do without a discussion of such concepts as "justification", "meaning" or "evidence". Another possible reason is just the youth of the science itself: it may well be possible that a common framework, or some especially important applications, will – at some time in the future – provide us with rough and ready criteria of evaluation. A third reason, interconnected with the other two, is the nature of the more general perspective within which much of the studies are pursued.

This today perhaps predominant, "Amsterdam", perspective on logic and modal logic in particular regards logical systems and theories mostly as modelling tools to describe formal, mathematical structures which themselves are taken to stand for real-world structures used and built by computer scientists. Computer networks, say, are first represented as transition systems: modal logics, in turn, are considered tools to describe these. While transition systems (under the name of Kripke frames) traditionally have been used to give the meaning of the syntactic elements a given logic consists of and to help us developing intuitions about the acceptability or non-acceptability of certain axioms, this perspective has lately been reversed: now its models first and logics later. Logic thus becomes empirical in a sense: transitivity of the abut becomes a modelling and engineering *decision*: all depends on the ulterior descriptive aims of the model-builder and ultimately on the purposes of the networks in question.

This empirical turn certainly had a liberating effect on a discipline considered metaphysically dubious and conceptually shaky by many from its beginnings. It is certainly a good thing if researchers working on distributed systems, program verification or the semantics of programming languages nowadays freely draw on the resources provided by modal logicians and it seems plausible to ascribe the possibility of such an unprejudiced new look at old methods to the demise of the traditional, "prescriptive" paradigm, viewing modal logic as a sub-branch of a yet to be developed general theory of rationality. Modal logic has profited immensely of the interest computer science took in it and has blossomed as a result of the number and diversity of the new approaches.

The disadvantage, however, of the Amsterdam perspective – model first, ask questions later – is that it has become very difficult to compare and contrast – conceptually, not with respect to their usefulness or metalogical properties – the different theories on the market. These theories are tied together not so much by a common subject matter, but just by common techniques and (to some extent) notation. This, however, may be just the way sciences go, after having left their mother discipline philosophy and become disciplines of their own. All the better if they still remember and respect their parents – not so for moral, but for practical reasons. Especially in young disciplines the danger of conceptual error looms large. Reflection on the conceptual bases may help to avoid belated and more radical, hence costly, paradigm shifts. In this work, I hope to provide something to some small steps in that direction.

Chapter 1

Meaning, Knowledge, Information

1.1 Three interrelated notions

Three notions will be of particular concern to us in the following: "meaning", "knowledge" and "information". They are obviously inter-connected. Transmission of knowledge and of information presupposes some access to meaning; meaning, in turn, is what can be known and conveyed as a piece of information and become knowledge in the end. Knowledge, at least propositional knowledge, is an epistemic state people may be in with respect to the meaning of what they have been told or the information they received. Information, finally, is what meanings may become and the crucial means by which we gain knowledge.

Among our three key notions, "information" is probably the fuzziest and least entrenched. It seems very much up for grasps how to construe such a notion, and there seems little to the conceptual core of it, which could not be given up in favour of otherwise attractive features of some technical precisification. "Information" shares this fate with kindred notions like "data": like these, "information" is at the same time fashionable and elusive. It is very difficult to avoid the impression that many people working in what nowadays are called the "information sciences" are talking past each other. Many controversies, e.g. the one on the veracity of information (see below, sct. 1.4.5), seem more of a terminological than a substantial nature.

A good analogue from a slightly different field is provided by "ontology", another equally general, elusive and fashionable term. By this term, usually taken to be on a par with what Aristotelian scholars called "metaphysics" ("the science of being *qua* being", as the latter is usually 'explained'), philosophers understand the study of what there is, i.e. the search for general categories classifying all entities, and thus for a high-level picture of the world which would be plausible, if not downright true, with respect to the general structure of the universe. In computer science, however, "ontology" has become to be used for a different, though related, purpose. An ontology is now taken to be some sort of hierarchical inventory of types of things talked about by some information repository, and, ultimately, of suitable data-structures for databases:

> "An ontology is a set of axioms that account for the intended meaning (the intended models) of a vocabulary." (Galgemi et al. 2002: 79)
> "Ontologies are generally held to provide standardised definitions of the terms used to represent knowledge." (Degen et al. 2002: 182)
> "An ontology is in this context a dictionary of terms formulated in a canonical syntax and with commonly accepted definitions designed to yield a lexical and taxonomical framework for knowledge-representation which can be shared by different information systems communities." (Smith 2002: 5)

It is clear from these quotes that, even among computer scientists, ontologies are taken to be very different things. In itself, this is not a bad thing: differences in terminology, like the differences in approaches, interests and emphasis they mirror, may prove fruitful and inspiring. It just becomes difficult, however, to agree on evaluation standards. The objection, e.g., that the *WordView* classification developed at Princeton University, classifying fever as a symptom, hence as evidence, as justification and ultimately as psychological feature, cannot possible true (because fever is not a psychological, but a physiological property of persons), will immediately meet the counter-objection that the *WordView* classification is based on sentences of the form "... is a $\cdots$" judged true by a (more or less representative) sample of English speakers and that the ... is a $\cdots$ relation is commonly taken to be transitive by those speakers. The objection thus misses the point because it is informed by another conception of what an ontology or classification is and hence what conditions a good ontology should meet.

As with "ontology", there are also different uses of "logic". By "logic", one can mean a set of sentences, some sort of grammar of a language, structural traits of the world justifying inferences or a mathematical tool to model relational dependencies (between structures, models, sentences or propositions).[1] The same goes for "information" and, to

[1]Richard M. Martin, e.g., characterises a logic just as a language: "Any formalized language system about a given subject matter may be regarded as providing a "logic" of that subject matter, more particularly a logic of the fundamental notions of that system." (Martin 1962a: 588) (Perhaps this view may be traced back to Ryle's conception of logic as some kind of regimentation of some sectors of ordinary discours Ryle (1954). Martin's "extensional logic of belief", developed in Martin (1962b) and Martin (1962a), then, consists basically in the proposal to regimate "Caesar knew that the capital of the Republic is situated on the Tiber" into "Caesar knew about (stands in a certain relation to) the function capital of, the Republic, the relation of being situated on, and the tiber, that the capital of the Republic is situation on the Tiber"

an even higher degree, "logic of information". Information can alternatively be taken to be an objective feature of the world, what is in the mind when someone thinks a thought, what is conveyed by utterances, what is stored in computers, what is written on a sheet of paper etc. A logic of information can hence taken to be some sort of description of informational dependencies or of the internal structure of some information repository in the world, a normative theory of how to transmit, store or structure information, an account of how natural languages, say, manage to convey information and many things besides. It is therefore difficult to evaluate claims like the following *credo* of Devlin's *Logic and Information* (note the scare quotes):

> "For purposes such as these [to design and to operate information processing devices] a 'logic' based on truth (such as classical logic) is not appropriate; what is required is a 'logic' based on *information.* It is the aim of this book to provide such a logic." (Devlin 1991: 10)

We will later see (in sct. 4.1) that many authors have claimed that epistemic modal logic has been a logic of information all along. Even in Devlin's own case, it is very difficult to assess what he takes the quoted passage to imply.

There is, however, an intuitive core to the notion of information, even if it may not be specific enough to distinguish it from many other notions in the semantic neighbourhood. This intuitive core is best brought out by observing that our notion of information is closely tied up with various activities:

- Information is conveyed, transmitted and broadcast.
- Information is stored, protected and sold.
- Information is received, processed and ready to be taken up.

In this, "information" differs both from "meaning" and "knowledge". While it may be said that meaning(s) or content(s) are conveyed and transmitted, it seems odd to speak of meaning(s) sold or stored, learnt or received. Knowledge, on the other hand, may be learnt (acquired) and considered valuable, but not broadcast or stored. So the notion of information seems to be a generalisation both of that of knowledge and of that of meaning, while retaining a strong kinship to both. It is knowledge "in spe" something that may become knowledge under certain conditions and for certain agents. Meaning, in turn, is a form of information, i.e. some of that information which is carried by representing items (see below, sct. 1.2.3).

---

and to take the last relatum to be referring to a sentence. Against such conceptions, Hintikka urged that a branch of logic be viewed as an "explanatory model in terms of which certain aspects of the workings of our ordinary language can be understood" (Hintikka 1968: 5).

There is another crucial asymmetry between information on the one, and meaning and knowledge on the other hand. On the view of nearly all protagonists of 'information science' the views of which will be discussed later, information is an *objective, mind-independent feature of the world.* Information is *there*, waiting to be picked up, whereas both meaning and knowledge are of our own making, require the existence of intentionally produced symbolic systems and epistemic agents respectively. Dretske forcefully made the relevant contrast in the preface to *Knowledge and the Flow of Information*:

> "Information is an artifact, a way of describing the significance *for some agent* of intrinsically meaningless events. [...] This is one way of thinking about information. It rests on a confusion, the confusion of *information* with *meaning.* Once this distinction is clearly understood, one is free to think about information (though not meaning) as an objective commodity, something whose generation, transmission, and reception do not require or in any way presuppose interpretive processes." (Dretske 1981: vii)

The theory mostly under consideration later, i.e. Jon Barwise's and John Seligman's theory of information flow, is influenced both by a theory of meaning and by a theory of knowledge. After some stage setting, I will discuss them in turn, beginning Fred Dretske's theory of knowledge (ch. 2). We will then address Jon Barwise's and John Perry's situation theory (ch. 3), which is basically a theory of meaning. We will then address some more general issues for a theory or logic of information (ch. 4) and discuss the general framework of epistemic modal logic (sct. 4.5). I will then present the Barwise/Seligman theory of information flow (ch. 5 to ch. 8), incorporate Kripke models into this framework and finally draw some (tentative) conclusions (ch. 9).

1.2 Meaning

1.2.1 Content, tone, and force

The first difficulty a theory of meaning has to face is the identification of its subject-matter. This turns out trickier than it may seem at first. Starting with a preliminary way to pick out what we are after and identifying this with *what is conveyed* by a speaker a at a particular time of utterance t, we already narrow down our range of options.[2] But

[2]It is, for one thing, uncertain that everything of interest to a theory of meaning, i.e. everything potentially meaningful, involves in one way or other something like a speaking person. What about flags and road-signs, menus and the Bible? But even where an (ordinary) person is involved, it does not have to be a speaker. It may be someone making a sign, pointing, or smiling, or just standing at a particular place.

even what is conveyed by a at t may mean different thing to different people. If the chairman rises and leaves the room, he thereby conveys to the audience that he thinks the presentation is awful – but even leaving aside what his action meant to him (he perhaps just went to the toilet), it may mean very little or very much too me (perhaps depending on the fact whether I am the speaker). So we have to relativise our pertinent notion of what is conveyed to a specific audience, both the one that it intended by the speaker and the one that actually takes up what he says. But this is still not enough: a may convey at t to b different things by doing different acts. If he fakes contempt by what he says and at the same time indicates complicity by raising his eye-brow, the eye-brow but not his speech will tell me what he really thinks.

We still have to go further: What a conveys to b at t by doing F may vary with circumstances or context. The action of doing F is individuated at least partly by reference to the actor's intentions, which may differ between different contexts of use. There are other, more easily generalised, context-dependencies as well. This is most obvious with indexicals, but there are other examples as well: "I'm American" may mean different things in a small-talk context or a politician discussion. Its meaning may even depend on which country (as opposed to continent) a comes from. Even with context-dependency taken into account, however, we are not yet done. Within one and the same context, an utterance may have different types of *force*. Compare the following utterances:

- You close the window.
- (Please,) close the window.
- Close the window!
- Do you close the window?
- (Would you mind) closing the window?

It has been said that utterances of these sentence types carry the same content while having different forces. Searle, in his seminal book *Speech Acts*, distinguishes different kinds of forces sentences may have: the assertoric force of an assertion is only one of them. We may perhaps say that *what is said* is the same thing in these different utterances and decide to call that thing (if it is one) the propositional content of these utterances.

We do not want our account of meaning to be too fine-grained neither. Differences

Another point of concern may be the reference to a particular time: if a colourful sunset means that it will be good weather tomorrow, does it mean this at all times in the interval in which it occurs (even after it is completed, or colourful, or a sunset)? What about some particular ornaments, which today mean (to some of us) that this house has been built in the 17^{th} century – did they mean this all along, even to the people who made them (and who did not know perhaps that they lived in the 17^{th} century)?

in *tone*, as between "He asked whether he should leave and then he went" and "He asked whether he should leave but then he went" (differing in their implicatures as to what advice may have been given), do not make for differences in meaning, whether or not on qualifies meaning (as it is often done in this context) as "cognitive" or "cognitively relevant". Differences in tone may be thought of as what is filtered out by *any* (admissible) formalisation, what a translation has to omit to count as a formalisation. Insofar as we are after a logic at all, we have to forsake tone.

So let us now settle on objectified, standardised, atemporalised propositional content as the object under investigation and let our propositional variables p stand for what a speaker conveys to an audience at a certain time by doing some action with some force in some context or other.

1.2.2 Propositional content

Even under all the restrictions mentioned in the previous section (1.2.1), the notion of propositional content remains slippery. One crucial step, often made unnoticed or at least rarely made problematic, lies in the identification of what is said (in the objectified, standardised, atemporalised sense identified above) with what is true or false. "Proposition" may (and is often taken to) stand for both. This is doubtful if we allow for components of force in what is said. Even with assertoric contents, however, the issue is far from clear.

If what is true or false must be completely determinate, the identification of what is said with what is true or false forces upon us a particular theory of ambiguity, according to which either nothing or two things are said by an ambiguous utterance (as opposed to an intrinsically vague proposition). It also besaddles us with the obligation to come up with a rather awkward construal of linguistic indeterminacy and vagueness. If I tell you that Tom, a borderline case of baldness, is bald, is this, even if unbeknownst to me (or even necessarily unbeknownst to me), true or false *simpliciter*? If I say that the cat is on the mat, without having any referential intention deciding between the infinitely many non-overlapping and hence non-identical lumps of feline tissue on the mat (or even only partially on the mat!), do I thereby say something true of false of one determinate cat? In both cases, it rather seems that what is said leaves open several options, on some such options giving us something true, in some others something false.

Again, this is a familiar picture with indexicals. With indexicals, what is said does not yet get at either truth or falsity. It takes the world (the context of utterance) together with the content to fix something which is either true or false. Have we not, however, already excluded indexicality in sct. 1.2.1? Indeed we have, but perhaps this was premature. There is a perspective on indexicality that considers it not so much an isolated feature of our less-than-perfect natural languages, but an essential feature of any language that merits its name. On this view, use of an indexical is not just a handy way to express many different propositions as a function of the time, place and speaker of the utterance, but a *cognitively essential* ingredient of what is said (by a particular speaker, at a particular time and place).

This is the doctrine that came to be known as that of the *essential indexical*, after the title of John Perry's article (1979), the underlying idea being that theory of content should account for what we ascribe to ourselves and others in order to rationalise, explain and predict our and their actions. So let us imagine a distracted philosopher, say John Perry, doing his shopping in the supermarket. He sees himself in a mirror and recognises that the man he sees, to which he refers as "this man" or "the man with the beard over there" or even as "John Perry", carries a bag of sugar with a hole in it, leaving a white trail behind him. The simple thought "this man carries a leaking bag of sugar" will not make him turn round. This effect will only be achieved by his thinking "*I* carry a leaking bag of sugar." This shows that the two sentences

(1) John Perry is making a mess.

(2) I am making a mess.

or, to change the example, the sentences

(3) The department meeting starts at noon.

(4) The department meeting starts now.

say different things, for (2) will make John Perry turn around (and (4) will make him leave his office) while he may stay perfectly comfortable where he is with either (1) or (3). So, the argument goes, there must be something to the *semantics* of (2) and (4) which distinguishes them from (1) and (3) respectively. Their essential indexicality, in Perry's view, forces us to

> "... make a sharp distinction between objects of belief and belief states, and to realize that the connection between them is not so intimate as might have been supposed." (Perry 1979: 28)

This has, among others, the consequence that we should classify belief states not only by propositions believed but also by sentences containing indexicals and by what Perry calls "relativised propositions", i.e. propositions which are not only true at worlds, but true at times, persons, places as well.

1.2.3 Meaning as a kind of information

It thus seems that the information carried by the utterance of a sentence by a person at a time is not adequately captured by the traditional doctrine of propositions.[3] In recent work, John Perry generalised the distinction between proposition expressed and information conveyed in important ways. The former, which he calls the "subject matter content" are the truth-conditions of an utterance in a certain context. The latter, which he calls "reflexive content" includes the truth-conditions the utterance may have if we do not take (all of) the contextual facts as given. The reflexive content can be brought out using what Perry in (Perry 2001a: 125) calls "the content-analyzer":

(CA) Given *such and such*, ϕ is true iff *so and so*.

Given all the contextual facts about the utterance of ϕ in question, *so and so* abbreviates the official content of the utterance. If we allow some such contextual conditions to vary, we get a more inclusive content, which may, on some occasions, correspond to the information conveyed by the actual utterance of ϕ:

> "... only a small part of the truth-conditions of an utterance are usually incorporated into what we think of as its content. The other parts are taken as given and exploited to get us to the subject matter we are interested in." (Perry 2001a: 127)

Generalising the insight from the *Essential Indexical*, Perry argues that the reflexive content has to enter an adequate account of the informativity of utterances and hence is part of the information conveyed by them.[4] As a special case, (CA) allows us to distinguish denotation-loaded from denotation-unloaded content (Perry 2001a: 126-127)[5] of, e.g., Donnellan's

[3]John Perry (1979: 29-30) characterises this doctrine by three tenets: that belief is a relation between a subject and an object, the latter being denoted by a that-clause; that these objects of belief (propositions) have a truth-value in an absolute sense; that their identity-conditions are fine-grained, i.e. that two propositions are identical only if they not only have the same truth-conditions, but involve the same concepts.

[4]Cf. e.g. "Just as the reflexive contents of our statements made clear how two statements with the same subject matter content can have quite different cognitive significance, the reflexive contents of our beliefs make clear how they can have different causal roles, each appropriate to its own reflexive content." (Perry 2001a: 130-131)

[5]We will later discuss an precursor of this general distinction in the special case of loaded and unloaded definite descriptions in Barwise's and Perry's situation semantics (cf. p. 79 in sct. 3.2.3).

famous sentence "The murderer of Smith is insane." In worlds which are members of the unloaded content of that sentence, different people murder Smith and each of them is insane in the respective world. In worlds which are members of the loaded content, Jones, the man who really murdered Smith, is insane (cf. Perry (2001b: 27) and (Perry 2001a: 127)). The loaded content does, while the unloaded content does not, contain Jones as a constituent; instead, the latter contains the identifying condition *being the murderer of Smith* (Perry 2001b: 26).

So there is more information conveyed by a given utterance, according to Perry, than what we learn by it about what can sensibly be called its subject matter.[6] Meaning, i.e. what an utterance says about its subject matter, is only some of the information conveyed by it. We here have an example of what Dretske called the phenomenon of nested information (71 Dretske 1981: cf. def. 2.2.4) described by Devlin as follows:

> "Given the possession of prior information and/or attunement to other constraints, the acquisition by an agent of an item of information Φ can also provide the agent with an additional item of information Ψ." (Devlin 1991: 16)

So we have seen that meaning is one kind and part of the information conveyed by an utterance. What kind of information is meaning? I think a distinction of Paul Grice's in an article called "*Meaning*" (Grice 1948) will turn out useful for answering this question.

Grice distinguishes natural from non-natural meaning. The first is a natural phenomenon: Due to empirical regularities, smoke (naturally) means fire, i.e. when- and wherever we see smoke, we may expect a fire. Natural meaning is factive: "x means that p" entails p. Non-natural meaning, as exemplified by natural language use, is non-factual and, in Grice's view, essentially intentional: In order for me to mean something p by an utterance u, I need to have at least the following intentions:

- To make my audience believe that I believe that p.
- To make my audience believe that I intended to make them believe that I believe that p.
- To make my audience believe that I believe that p *because* I intended to make them believe that I believe that p.

These three intentions characterise what Grice subsequently called *speaker's meaning*, i.e. my (non-naturally) meaning that so-and-so on a particular occasion. As Grice shows

[6] The inclination to think otherwise is taken by Perry to be symptomatic of the "subject matter fallacy": "The subject matter fallacy is supposing that the content of a statement or a belief is wholly constituted by the conditions its truth puts on the subject matter of the statement or belief; that is, the conditions it puts on the objects the words designate or the ideas are of." (Perry 2001b: 50)

with elaborate examples, all these three intentions are individually non-sufficient.[7] It has been a matter of much dispute whether the three intentions are jointly sufficient: confronted with even more elaborate counter-examples, Grice subsequently complicated his theory (cf. his *Utterer's Meaning and Intentions* (Grice 1969: 114), where the new definition runs to almost a page). Given that they are plausibly taken to be necessary, however, this will not matter much for our purposes.

Dretske remarked that meaning is only sharply to be distinguished from information if one means the non-natural variety. Natural meaning, he suggests (Dretske 1981: 242), may be "close (if not equivalent) to the ordinary sense of "information"" and says (Dretske 1990: 115) that the distinction between information and meaning is "roughly equivalent" to the distinction between natural and non-natural meaning.

This is too quick, however, for the third intention required by Grice displays exactly the type of connection crucial to information flow, as we will see later. Grice is very careful in requiring not a causal, but an explanatory relationship between the achievement of the first two intentions:

> "A must intend to induce by x a belief in an audience, and he must also intend his utterance to be recognised as so intended. But these intentions are not independent; the recognition is intended by A to play its part in inducing the belief, and if it does not do so something will have gone wrong with the fulfilment of A's intentions. Moreover, A's intending that the recognition should play this part implies, I think, that he assumes that there is some chance that it will in fact play this part, that he does not regard it as a foregone conclusion that the belief will be induced in the audience whether or not the intention behind the utterance is recognised." (Grice 1948: 219)

We will later see in the discussion of Dretske's own theory (in sct. 2) that he needs basically the same proviso for his own theory of natural meaning.

[7] The first intention is not sufficient because I might leave B's handkerchief at the scene of a murder (intending the police to suspect B) without the handkerchief (non-naturally) meaning anything. The first and second intentions are not jointly sufficient because Herodes presented Salome with the head of St. John the Baptist with both the intention to get her to believe that he is dead and to get her to believe that he did what he did to show her he was dead; but all the same his head did not (non-naturally) mean that St. John the Baptist was dead. He did only mean that if the recognition of Herodes' intention by Salome would have been required for inducing in her the belief that St. John was dead.

1.3 Knowledge

1.3.1 The classical definition and its demise

"Knowledge" has always been one of philosophers' most cherished concepts. Traditionally, the debate has centred about the third clause in the classic definition of "a knows that p", commonly attributed to Plato's *Theatetus*, though Plato does not there endorse it:

Definition 1.3.1 (Classical definition of knowledge). *An agent a knows that p iff*

(i) It is true that p.

ii) a believes that p.

(iii) a is justified in believing that p.

In 1963, Edmund Gettier published a short note showing that this analysis is inadequate: though the three conditions might be necessary, they are not jointly sufficient (Gettier 1963). A counterexample is the following scenario: I believe that one pupil in my class is French, on the basis of my justified, but untrue, belief that Joan is French. If I have a justification for this latter belief, I have one for the existential generalisation. It may happen, however, that the generalisation is true, while the instance is not. There is another pupil in my class who, unbeknownst to me, is French. It seems false to say in this scenario that I know that one pupil in my class is French, though the three conditions are satisfied.

The common reaction to such counterexamples was to look for a fourth condition, or, equivalently, a third condition which subsumes the old one, while being stronger than it. Various candidates have been proposed. It has, e.g., been required that my believe that p should be counterfactually dependent on the truth of p (so that, if p were not true in relevantly similar circumstances, I would not (or cease to) believe that p). Another way of fixing (1.3.1) has been proposed by Alvin I. Goldman (1967). His idea is to require some sort of causal connection between the epistemic state of the believer and the object the alleged knowledge is about.[8] The most popular candidate was to require some kind of *reliable* connection:

> "A truly believes that p, and his belief [is] appropriately produced in a way such that beliefs produced in that way are generally true, where "appropriately" means that the truth of the belief is not accidental relative to what it is about the way of its production which makes that a generally reliable way." (Williams 1978: 45)

[8]To allow for knowledge about the future, Goldman draws a difference between there being a causal connection between a and b and a's being the cause of b (or vice versa): the former, but not the latter holds if a and b are effects of a common cause (Goldman 1967: 364).

Dretske's theory (presented in sct. 2) may be seen a member of the latter, reliabilist, family.

In 1981, Robert Nozick proposed subjunctive third and fourth conditions, imposing a sort of counterfactual dependency between the truth of the belief and the fact that it is had:

Definition 1.3.2 (Nozick's definition of knowledge). *An agent a knows that p iff*

(i) *It is true that p.*
(ii) *a believes that p.*
(iii)* *If p were not true, a would not believe that p. (Nozick 1981: 172)*
(iv)* *If p were true, a would believe that p and it would not be the case that a believes that* $\neg p$*. (Nozick 1981: 178)*

Nozick calls such counterfactual dependence "tracking" and explicitly links it to there being a *constraint* on our beliefs and the way the world is:

> "A person knows that p when he not only does truly believe it, but also would truly believe it and wouldn't falsely believe it. He not only actually has a true belief, he subjunctively has one. It is true that p and he believes it; if it weren't true he wouldn't believe it, and if it were true he would believe it. To know that p is to be someone who would belief it if it were true, and who wouldn't believe it if it were false. [...] To know is to have a belief that tracks the truth. Knowledge is a particular way of being connected to the world, having a specific real factual connection to the world: tracking it." (Nozick 1981: 178)

Nozick's notion of tracking and his requirement that there be a law-like dependency of beliefs on the state of affairs they are about will turn out important in the following.

While Gettier's 1963 article re-awakened the philosophical zeal to find a reductive definition of knowledge, some authors recently argued that no such reductive definition, in particular no definition of the form of (1.3.1) and (1.3.2) may possible be true. Timothy Williamson, in his influential book *Knowledge and Its Limits* (2000), has argued that knowledge is not analysable in terms of belief at all. He argues that no such factorisation of knowledge into non-circular necessary and sufficient conditions is possible, that knowledge has to be acknowledged as a primitive concept. Rather than to define it, philosophers would better concentrate on spelling out its logical behaviour.

Revolutionary as this thesis was in epistemology, it has long ago been endorsed by situation theorists:

> "Common sense says that knowledge is one thing, belief is another. There is no particular reason to think that one is in any sense more fundamental than the other,

or that one can be defined in terms of the other. On the other hand, there are obvious relations between knowledge and beliefs. These relations amount to necessary structural constraints..." (Barwise and Perry 1983: 265)

We will come back to these relations between knowledge and belief in sct.1.3.7. Before doing this, however, we will discuss some of the more important putative properties of knowledge, conceived of as an operator on propositions and represented by "K".

1.3.2 Omniscience

The most important assumption underlying epistemic interpretations of modal logic is that knowledge satisfies the following axiom (K):

(K) $\vdash \mathsf{K}p \wedge \mathsf{K}(p \to q) \ \to\ \mathsf{K}q$

Together with the following closure principle, this ensures that the logic of knowledge is *normal*:

(Nec) $\dfrac{\vdash p}{\vdash \mathsf{K}p}$ Nec

Fred Dretske, in a series of famous articles, has argued that (K) is false of all epistemic operators, in particular of "a knows that" and "a has reason to believe that".[9] After having searched the whole building and not found anyone in it, I may have reason to believe that the church is empty without having reason to believe that it is a church that is empty, though the truth of the latter is entailed and presupposed by the truth of the former and I may know (and a fortiori have reason to believe) that, even if I know that the latter is a consequence of the former (Dretske 1970: 36). Even if I know that zebras are not mules, it does not follow from the fact that I know that the animals I see in the zoo are zebras that I know that they are not mules cleverly disguised to look like zebras (Dretske 1970: 40). He puts forward what has become to be known as a relevant alternatives approach: in order to know that p, I do not have to rule out all possibilities in which $\neg p$, but only relevant alternatives in which $\neg p$. In a given instance of (K), I may be able to rule out the alternatives relevant to the premiss-es without being able to rule out the alternatives relevant to my knowledge claim stated in the conclusion.

[9] Other epistemic operators include "a sees that", "there is evidence to suggest that", "a can prove that", "a learns (discovers, finds out) that", "in relation to our evidence it is probable that" (Dretske 1970: 32).

Convincing as these examples are, they have come under heavy attack. G. C. Stine has famously argued that the presupposition carried by a knowledge claim to p that $\neg p$ is not a relevant alternative is only pragmatic, not semantic, i.e. that it is a presupposition tacitly made by the speaker and not pertaining to the knowledge claim itself (Stine 1976: 255). Pragmatic presuppositions are cancellable[10] and utterances carrying them may be true even if the presuppositions in question are false. So we may either assert the conclusion, despite our committing ourselves thereby to a false pragmatic presupposition, or withdraw our assertion of the first premiss.

The central idea of this line of response to putative counterexamples to (K) is to restrict (K) to reasoning *within* a fixed context and to take the set of relevant alternatives to be an essential feature of a context (Stine 1976: 256). A basically equivalent move would be to take different sets of relevant alternatives to define different meanings of "know" and its derivatives (Williamson 2002: 9). Relevant alternative counterexamples to (K) would then be guilty of a fallacy of equivocation or, at least, an illegitimate change of context.

The problem is that (K) is derivable from the application of the following principles to knowledge K:[11]

(5) $\vdash p \rightarrow q \;\;\Rightarrow\;\; \vdash \Box p \rightarrow \Box q$

(6) $\vdash \Box p \wedge \Box q \;\;\rightarrow\;\; \Box(p \wedge q)$[12]

The problem then re-emerges even with operators which are clearly 'non-penetrating' (this is Dretske's term for non-distributivity over 'boxed' implication). Take, e.g., "it is strange that". Even if Dretske is right that "a concatenation of factors, no one of which is strange or accidental, may itself be strange or accidental" (Dretske 1970: 32) and the converse of (6) is not true, (6) seems ok, as does (5) (at least for a restricted sense of "it is strange that" not applicable to astonishing theorems). So we would have the strange result that "it is strange that" distributes. This is strange only if we do not consider what "$\Box(p \rightarrow q)$" would mean in this case. If it is strange that Susan and Bill married each other but it is

[10] An example of a pragmatic presupposition is the exclusive reading of some uses of "or". If I tell a child standing in front of an ice-cream shop that it might have chocolate or strawberry flavoured ice-cream, I normally do not intend to offer him the possibility of having both. Adding "or both", however, cancels this presupposition and thus shows that it was only pragmatic, not semantic.

[11] We have $\vdash (p \wedge (p \rightarrow q)) \rightarrow q$, hence, by (5), $\vdash \Box(p \wedge (p \rightarrow q)) \rightarrow \Box q$. By (6), we have $\vdash \Box p \wedge \Box(p \rightarrow q) \rightarrow \Box(p \wedge (p \rightarrow q))$ and chaining the two together we get $\vdash \Box p \wedge \Box(p \rightarrow q) \rightarrow \Box q$.

[12] For the knowledge case, (6), i.e. epistemic conjunction, has been denied by Devlin (1991: 186). Unfortunately, he does not give any reasons. Provided one is prepared to accept *any* such logical principles governing knowledge, (6) seems to me one of the most obvious candidates.

not strange that Susan married. The reason for this, however, is not so much the failure of (K) but the falsity of premiss $\Box(p \to q)$: it is not strange at all that if Susan and Bill married then Susan got married.

I do not know whether this strategy works in all cases: there is, after all, nothing bizarre to the idea that some strange fact implies some familiar fact and that this implication is itself strange. The point I want to make is just that putative counterexamples to (K) have to be chosen carefully so as to render the major premiss as plausible as possible.

There is another reason to be sceptical about counterexamples to (K): what the relevant alternatives to consider in the ascription of knowledge claims are depends, as we saw, on context. This context is not just the context of the ascriber, but includes at least some features of the context of the person to which knowledge is ascribed. It is therefore possible to argue that the sceptical scenario (q) becomes a relevant alternative to consider just by our ascribing to the subject in question knowledge that p and that $p \to \neg q$. One way to bring this out is to see the non-obtaining of the sceptical scenario ($\neg q$) as a pragmatic rather than a semantic presupposition which may be cancelled and *is* cancelled in instances of (K), where the possibility of p is explicitly considered in the second conjunct of the antecedent. Another way is to follow David Lewis in taking the relevant/non-relevant distinction between possible worlds to be subject to pragmatic rules, one of which is the rule of actuality:[13]

> "When we say that a possibility is properly ignored, we mean exactly that; we do not mean that it *could have been* properly ignored. Accordingly, a possibility not ignored at all is *ipso facto* not properly ignored. [...] No matter how far-fetched a certain possibility may be, no matter how properly we might have ignored it in some other contexts, if in *this* context we are not in fact ignoring it but attending to it, then for us now it is a relevant alternative." (Lewis 1996: 434)

The sceptical scenario, then, is not properly ignored already in the premisses of the argument because it is mentioned in the second one.

Note that it would not be an appropriate line of reply to insist, as it is often done, that the subjects of interest to us (the epistemic behaviour of which we are modelling) are logically perfect reasoners, drawing all logical conclusions from what they know. This reply only makes our situation worse, for it justifies (Nec) and thus forces us to accept not

[13] The others are the rule of actuality (the actual world is never properly ignored), the rule of belief (a possibility that the putative knower believes to obtain is not properly ignored), the rule of resemblance (we never properly ignore just one of two similar possibilities), the rule of reliability (we properly ignore the possibility that reliable mechanism malfunction) and the rule of conservatism (generally ignored possibilities are properly ignored).

only distribution over known implication, but even distribution over any logical entailment whatsoever.

In order to get clear about this, we have to distinguish a whole family of closure principles (Gochet and Gribomont 2002: 46):

(7)	If $\vdash \phi$ then $\vdash \mathsf{K}\phi$	closure under theoremhood
(8)	If $\vdash \phi \to \psi$ then $\vdash \mathsf{K}\phi \to \mathsf{K}\psi$	closure under logical implication
(9)	If $\vdash \phi \leftrightarrow \psi$ then $\vdash \mathsf{K}\phi \leftrightarrow \mathsf{K}\psi$	closure under logical equivalence
(10)	$\vdash \mathsf{K}(\phi \to \psi) \ \to \mathsf{K}\phi \to \mathsf{K}\psi$	closure under material implication
(11)	$\vdash \mathsf{K}(\phi \leftrightarrow \psi) \ \to \mathsf{K}\phi \leftrightarrow \mathsf{K}\psi$	closure under material equivalence
(12)	$\vdash (\mathsf{K}\phi \wedge \mathsf{K}\psi) \to \mathsf{K}(\phi \wedge \psi)$	closure under conjunction
(13)	$\vdash \mathsf{K}(\phi \wedge \psi) \to (\mathsf{K}\phi \wedge \mathsf{K}\psi)$	closure under simplicification

While (7) corresponds to (Nec) above, (10) is (K) and (12) corresponds to (6). Of these seven principles, the first three clearly are the most problematic ones, "imput[ing] an infinite capacity to the agent" (Gochet and Gribomont 2002: 47). The problem mentioned above is that (K), in the presence of (Nec), gives us (8) and (9). Obviously, (8) implies (9). Given (13), the converse is true as well, i.e. (9) implies (8).[14] Given (13), (9 even implies (7/Nec) provided that I know something.[15]

Our list thus falls naturally into two groups, the first three principles being far more, the last four much less problematic. It seems that they could only be avoided by a very syntax-oriented approach – which in turn will not do justice to the trivial transformations rational epistemic agents *can* (and should be able to) perform. The problem, generally speaking, is how to avoid both logical omniscience and logical blindness, i.e. to draw a line between (9) and (10) on the above list. We will later, in sct. 4.8.2, discuss the strategy of doing without (12) and, in sct. 8.1.3, of dispensing even with (13).

Another general strategy may be suggested by the following consideration. (8) fails if we impose, as Dretske did, a requirement on knowledge that is close to Nozick's notion of tracking (cf. def. 1.3.2): in order for me to know that p, there has to be a reason R such that my belief is based on the fact that R holds and such that, if p were false, R would not hold. The problem then is that it seems not clear how to interpret, on these terms,

[14] Suppose $\vdash p \to q$. Then $\vdash p \leftrightarrow (p \wedge q)$, hence, by (9), $\vdash \mathsf{K}p \leftrightarrow \mathsf{K}(p \wedge q)$, and $\vdash \mathsf{K}p \to \mathsf{K}q$ by (13).

[15] Suppose $\vdash \mathsf{K}p$ for an arbitrary proposition p. If $\vdash q$, $\vdash p \leftrightarrow p \wedge q$. So $\vdash \mathsf{K}(p \wedge q)$ by (9) and $\vdash \mathsf{K}q$ by (13).

the second premiss so as to make the following inference valid:

$$\frac{\begin{array}{l}\text{If } p \text{ were false then } R \text{ would not hold.}\\ p \text{ entails } q.\end{array}}{\text{If } q \text{ were false, then } R \text{ would not hold.}}$$

The situation, however, changes in important respects if we replace the second condition of the tracking account with its contraposition and require that p is true only if R holds. R may then be said to *indicate* that p is true. We then have the following inference, which is valid:

$$\frac{\begin{array}{l}R \text{ would hold only if } p \text{ is true.}\\ p \text{ entails } q.\end{array}}{R \text{ would hold only if } q \text{ were true.}}$$

While this interpretation of knowledge makes (K) acceptable, it suggests at the same time that the notion captured is more precisely described as "having the information that" than as "knowing that".[16] It seems plausible to me that the transitivity of "indicates" is just the manifestation of a more widespread phenomenon, called "nesting" by Dretske, which we will discuss in more detail below (cf. def. 2.2.4).[17]

(Nec), indispensable as it is to the project of modelling knowledge with the tools provided by modal logic, imputes *logical omniscience* on the epistemic agents, a problem which we will discuss in much detail later (cf. sct. 4.8.2). It has been heavily criticised e.g. by Hector-Neri Castañeda in his review of Hintikka's *Knowledge and Belief*:

> "...[Hintikka's] senses of 'knowledge' and 'belief' are much too strong [...] since most people do not know every proposition entailed by what they know; indeed many people do not even understand all deductions from premises they know to be true." (Castañeda 1964: 133–134) [18]

Together with (K), (Nec) gives us (5) and hence closure of knowledge under material implication. This is a weaker principle than (Nec), but it may still be deemed unacceptable.

[16] A somehow intermediate solution, exposed and defended by Duc (2001), as reported by (Gochet and Gribomont 2002: 50–51), would be to change (K) for the temporalised $\mathsf{K}(p \to q) \to \mathsf{K}p \to \Diamond\mathsf{K}q$. The resulting logic eschews all seven forms of closure mentioned in the main text and thus seems to rule out too much. It is a further disadvantage of the system, however, that we are not told *what* the agent has to do to eventually arrive at $\mathsf{K}p$. It is not very helpful to be told that "$\Diamond\mathsf{K}p$" reads "sometimes after using rule R, the agent knows p".

[17] It is somewhat ironical that Dretske (as we will see in ch. 2, in particular in sct. 2.2.2 and sct. 2.3) later advocated an even stronger definition of knowledge according to which it distributes – not only over *known* but – over any necessary connection whatsoever.

[18] While Hintikka defended (Nec) for his notion of knowledge, he grants the point for the – in his view, more general – notion of information: "It simply is not true that there are no objective senses of information in which logical inference can add to our information." (Hintikka 1970: 181)

The presence of (Nec) in a logical system makes it clear that what it models at best a highly idealised notion of knowledge. It has proven very difficult to spell out the degree and the nature of this idealisation in knowledge-based terms.[19]

(K), on the other hand, does not demand such a reinterpretation: requiring that the implication has to be known, it does not for itself force upon us an idealisation. It does not, as (Nec) does, impute infinitary powers on our agents, nor does it make an unconditional assertion about what every agent knows. Instead, it constrains their behaviour as modelled in the system: (K), therefore, is more naturally taken to be an external constraint on our knowledge *ascriptions*: we should not, in other words, deny our agents the capacity to perform *modus ponens*. (Nec), on the other hand, forces us to ascribe only deductively closed belief or knowledge sets, i.e. it forces us to ascribe knowledge that normal agents do not have. I conclude, then, that (K) may be plausibly held to be acceptable, at least in the absence of (Nec) which is the real culprit and a far more problematic principle.

1.3.3 Reflexivity and consistency

Another crucial feature of any acceptable logic of knowledge is the following "truth axiom", incorporating condition (i) in the classical definition of knowledge (1.3.1):

(T) $\vdash \mathsf{K}p \rightarrow p$

While the inclusion of (T) in any logic of knowledge is not controversial, its equivalent for a logic of belief has been subject to disputes. Hintikka, in adapting the logic of knowledge – which he took to be **S4** – to the case of belief replaced (T) with (D):

(D) $\vdash \mathsf{B}p \rightarrow \neg\mathsf{B}\neg p$

thereby ruling out inconsistent belief sets.[20] Lemmon violently criticised this on the ground that (D), even if perhaps a useful principle in the representation the agents gives of his

[19] Cf. the following remarks of Hintikka: "What causes the breakdown of these rules [(8) and (9)] is after all the fact that one cannot usually see all the logical consequences of what one knows and believes. Now it may seem completely impossible to draw a line the implications one sees and those one does not see by means of general logical considerations alone. A genius might readily see quite distant consequences while another man may almost literally 'fail to put two and two together'. Even for one and the same person, the extent to which one follows the logical consequences of what one believes varies with one's mood, training, and degree of concentration." (Hintikka 1970: 181)

[20] One possibility to sacrifice (D) without allowing inconsistent beliefs is to renounce to closure of belief sets under conjunction ($\mathsf{B}p \wedge \mathsf{B}q \rightarrow \mathsf{B}(p \wedge q)$), which is the option chosen by Ruth Barcan Marcus. The resulting doxastic logic, however, then ceases to be regular. We will come back to this in sct. 4.8.2

own belief,[21] is prohibitively crippling from the modelling perspective – since it is plainly possible (and actual) that people hold inconsistent beliefs.

1.3.4 Transitivity

We will later see how, on the semantic side, an agent's knowledge that p is usefully analysed as p's being the case in all of a's epistemic alternatives. The following axiom ensures that the relation of epistemic alternativeness is transitive:

(S4) $\vdash \mathsf{K}p \rightarrow \mathsf{KK}p$

Jaakko Hintikka argued for (S4) on the basis of his peculiar semantics given in terms of defensibility (consistency) of belief sets:

> "That q is the case can be compatible with everything a certain person – let us assume that he is referred to by a – knows only if it cannot be used as an argument to overthrow any true statement of the form "a knows that p." Now this statement can be criticised in two ways. One may either try to show that p is not in fact true or else try to show that the person referred to by a is not in a position or condition to know that it is true. In order to be compatible with everything he knows, q therefore has to be compatible not only with every p which is known to him but also with the truth of all the true statements of the form "a knows that p."" (Hintikka 1962: 18)[22]

The idea, in a nutshell, is that I only know something if it is not excluded by what I know nor by my knowing what I know. This is based on the observation that explicit self-ascription of knowledge is a particularly strong form of endorsing a claim:

> "I am not in a position to say "I know" unless my grounds for saying so are such that they give me the right to disregard any further evidence or information." (Hintikka 1962: 20)

This argument from entitlement or rhetoric force of knowledge claims[23] is inconclusive, however. It may be defended (at least by Hintikka-style arguments) only in contexts

[21] Though he doubted even that: "Finally, in moments of self-honesty, a man may *admit his own* inconsistency (consistently) in the form of Moore's paradoxical assertion - "my problem is that I'm a homosexual but I don't really believe that I am."" (Lemmon 1965: 382–383)

[22] He had an ancillary argument for transitivity, based on the observation that "knowing" and "knowing that one knows" are often used interchangeably in ordinary language – or, at least, that their "basic meanings" as captured by an explanatory model are the same (cf. Hintikka 1968: 8).

[23] Hintikka sometimes also puts the point in terms of justification: "In the primary sense of *know*, if one knows one *ipso facto* knows that one knows. For exactly the same circumstances would justify one's saying "I know that I know" as would justify one's saying "I know" *simpliciter*." (Hintikka 1962: 28,111)

where the "absurdity" correlated to an indefensible belief set is understood in performatory terms.[24] This performatory aspect of knowledge claims, however, is cancellable and applies only to first-person utterances. Whenever we are interested in modelling the epistemic behaviour of (real or idealised) agents, such considerations are inapplicable.

In reply, Hintikka relies on the idealisations he made: he only considers statements made on one and the same occasion and presupposes that the person referred to by "a" in the index of the knowledge operator knows that he is being referred to (Hintikka 1962: 106).

Hintikka considers it one of the major advantages of his theory that it is able to explain the pragmatic weirdness of 'Moore's paradox' by its unbelievability:

(14) p, but I do not believe that p

Assuming doxastic alternativeness (DA) to be transitive, it is straightforward to show that Moore's paradox is unbelievable.[25] We should, however, resist the temptation to take doxastic alternativeness to be transitive. For it is clearly possible to believe that one lacks beliefs one in fact has. What makes the indefensibility of Moore's sentence paradoxical is not only that we know that, for most if not all p, it might be true and for most of the true p's of our language it is actually true. It also is paradoxical because there does not seem anything wrong with ascribing belief in it *to someone else*. Using transitivity to show that Moore's paradox is unbelievable *tout court* misses this feature: for someone who is claiming that p is true although he does not believe it is claiming that transitivity fails for him. Because he, after all, might very well be right, we get a Moorean paradox at the meta-level.

The sense in which (14) is indefensible is better brought out by considering that belief in (14) makes the world in which it is held either doxastically inaccessible to itself

[24] Hintikka is clear on this point: "The absurdity of doxastically indefensible sentences is of *performatory character*; it is due to doing something rather than to the means (to the sentence) which is employed for the purpose." (Hintikka 1962: 77) He calls transitivity also "the quasi-performatory aspect of the verb *know*" (Hintikka 1962: 55). He repeatedly stressed this point in subsequent discussions: "[...] for someone to know that p his evidence [...] has [...] not only to be good but as good as it [...] can be. It has to be such that further inquiry loses its point (in fact, although it is logically possible that such an inquiry might make a difference). The concept of knowledge is in this sense a 'discussion-stopper'. It stops the further questions that otherwise could have been raised without contradicting the speaker." (Hintikka 1968: 13)

[25] Assume there is a possible world w where a believes that p and that a does not believe that p. Assume there is a doxastic alternative v for a in w. Then $v \models p \wedge \neg \mathsf{B}p$. Because of the second conjunct, there is an doxastic alternative u for a in v such that $u \not\models p$. By transitivity, u is also a doxastic alternative for a in w. Hence $u \models p$, which is impossible. So a has no doxastic alternatives in w.

or contradictory. (14) cannot be rationally believed in 'reflexive' worlds, i.e. worlds w such that $w\mathsf{DA}w$.[26] Taking a doxastic alternative to be possibly actual as opposed to just merely possible, however, means taking it to be doxastically accessible to itself. So no one can take a doxastic alternative in which (14) is true to be a possible way *his* actual world, the world of the believer, might be.

What lies at the bottom of indefensibility of (14) is not the transitivity of doxastic alternativeness but the commonly made presupposition that what one says might be true even in a state of complete information. If I utter p, I therefore commit myself to the claim that p might be true even if I knew everything about the actual world there is to know, that is, even if the actual world were my only doxastic alternative.[27] If you utter p, you must consider it possible that you would believe p even if you had a maximally specific belief set, i.e. if you would believe all the truths (or, equivalently, disbelieve all the falsehoods). If the world in which we believe them were our only doxastic alternative, belief in (14) would make that world inaccessible to us.

The strongest argument against (S4), I think, is the simple fact that knowledge may supervene on factors unknown to the agent. This is nicely brought out in Dretske's notion of a channel condition (cf. sct. 2.2.3),[28] but the point can be made independently: Suppose that whether or not I have knowledge of the truth that p depends on the obtaining of some external factors r (i.e. a reliable connection between my belief and the fact that p, the trustworthiness of my conversational partner, the actual absence of other possibly misleading evidence or what have you) such that, if r were not be the case, I would not know that p. It is enough, then, for my knowing that p that r obtains – I do not have to know that it obtains. In order to know that I know that p, however, I have to rule out that $\neg r$, at least according to Nozick's definition of knowledge (1.3.2) – but also, if $\neg r$ is sufficiently likely to occur and the demands on justification are sufficiently high, according

[26]Suppose $w \models \mathsf{B}(p \land \neg\mathsf{B}p)$ and $w\mathsf{DA}w$. Because of $w \models \mathsf{B}\neg\mathsf{B}p$, $w \models \neg\mathsf{B}p$. But the first conjunct becomes $w \models \mathsf{B}p$, so w is contradictory.

[27]Something along this line was Hintikka's original justification for the transitivity of his belief relation: "If something is compatible with everything you believe, then it must be possible for this something to turn out to be the case together with everything you believe without making it necessary for you to give up any of your beliefs. If your beliefs are to be consistent, it must also be possible for all your beliefs to turn out to be true without forcing you to give up any of them." (Hintikka 1962: 24) Williamson takes this robustness of knowledge to be one of its distinctive differences from true belief: "Knowledge is superior to mere true belief because, being more robust in the face of new evidence, it better facilitates action at a temporal distance." (Williamson 2000: 101).

[28]Dretske even denies (S4) even for the case where the first modal operator is weakened to mere belief: "We naturally expect of one who knows that P that he believes that he knows, just as we expect of someone who is riding a bicycle that he believe he is riding one, but in neither case is the belief a *necessary* accompaniment." (Dretske 1971: 21)

to the classical definition (1.3.1). To have knowledge that I know that p, I have to rule out more ways the world might be, and thus there are situations in which, lacking this knowledge, (S4) is false. An especially vivid case is where I even lack the concepts or linguistic resources to express r: it seems not clear, in this case, how I might be able to rule out that $\neg r$, even while I do not have to articulate the background conditions on which may ordinary, first-level knowledge claims depend.

1.3.5 Introspection

In the context of epistemic modal logic, (S4) has been interpreted as the axiom of *positive introspection*. The following axiom may then be aptly be called *negative introspection*:

(S5) $\vdash \neg \mathsf{K}p \to \mathsf{K}\neg \mathsf{K}p$

Even researches who subscribed to (S4) have heavily criticised (S5), most notably Wolfgang Lenzen (cf. e.g. Lenzen (1978: 79) and Lenzen (1979: 35)) and Jaakko Hintikka (Hintikka 1962: 106).

(S5) ascribes to agents the capacity to gain introspective knowledge just in virtue of lacking another piece of knowledge. While this principle may be acceptable if $\neg \mathsf{K}p$ is given some "positive", "substantial" interpretation (not just lacking knowledge that p but positively disbelieving p or doubting that p), it is inacceptable iff $\neg \mathsf{K}p$ just means that p is false in some of the agent's epistemic alternatives. For, as we will see in sct. 4.4, such epistemic alternatives are just total states of the world *not excluded* by the epistemic state of an agent at a given time. The agent does not have to be able to survey the range of these alternatives and to know, e.g., whether there is some $\neg p$ world among them: he has, in other words, epistemic access to them only under the description "worlds where all propositions I know are true" – he does not have to know of every proposition whether it is true or false in his epistemic alternatives. Whether or not p fails in some epistemic alternative cannot be decided by the agent on the sole basis of what he knows; he needs, in addition, some grasp of what he could but does not know.

The acceptance of (S5) in our epistemic logic forces us to ascribe knowledge to epistemic agents on the sole basis of their *lack* of knowledge of some other proposition. Whenever someone does not know that p, e.g. because he has never heard of p or because he even lacks the concepts to formulate a sentence or to entertain a thought expressing it, (S5) (together with Modus Ponens), forces us to ascribe him knowledge of his lack of knowledge. The world would be very different if people were always that knowledgeable about their ignorance. Moreover, (S5) meshes badly with the confidence we naturally have in our knowledge claims. Suppose I am subjectively certain that a false proposition p is true. Given (T), I do not know it. (S5) then ascribes to me knowledge that I do not know that p, which (given that I proportion my degree of confidence in p to my knowledge) seems to rule out my being subjectively certain that p. Unless "knowledge" is understood in an idealised sense, (S5) seems to me out of question.[29]

1.3.6 Epistemic modal logic

Given the six principles above, it is now possible to define different modal logics. The weakest normal modal logic **K** is given by any complete axiomatisation of the propositional calculus, (P), (K) and the necessitation rule (Nec). We add (D) or (T) to get **D** and the stronger **T** respectively, and add (S4), (S5) and (D) to get **K45**, **KD45**. Adding (S4) to **T** gives us a system called **S4**, while adding (S5) to **T** gives us **S5**.

(G) $\vdash \neg\Box p \to \Box\neg\Box\neg p$

which corresponds to confluence of the accessibility relation.[30]

Up to now, we only considered propositional epistemic logics. Space constraints will not allow a discussion of epistemic predicate logics, however.[31] Suffice it to say that problems multiply with the interpretation of bound variables and quantifiers. Jaakko Hintikka, in Hintikka (1962), restricts the rules of existential generalisation and universal specialisation to individual constants a such that the agent in question knows who or what the referent of a is. Accordingly, $\forall x \mathsf{K}_a Fx$ reads "everything such that a knows who/what it is is known by a to be F" and "everything is known by a to be F" is formalised by $\forall x \exists y(x = y \land \mathsf{K}_a Fy)$. This double reading of the quantifiers has been heavily criticised in the literature (see Stine 1974: 128–129 for references). Another problem is that, according to Hintikka, knowing who someone is requires the ability to refer to him by some

[29] It may thus be argued that (S5) forces us to ignore the possibility of error, i.e. to assume that agents are always right in their knowledge claims: this, as Stalnaker (1999b: 258) noticed, blurs the distinction between knowledge and belief.

[30] A relation R is confluent iff it satisfies the following: $\forall s\ \forall t\ \forall u\ ((sRt \land sRu) \to \exists v(tRv \land uRv))$, i.e. iff every branching is undone at a later stage.

[31] Cf. (Gochet and Gribomont 2002: 32–45) for a discussion of some of their problems, especially the interaction of the Barcan formula and the converse Barcan formula with the cardinality of the individual domains associated with the possible worlds.

proper name or singular description. In Stine's system, this drawback is avoided by an interpretation of the quantifiers which takes them to range only over individuals known to the agent in question. So the reading of $\forall x \mathsf{K}_a Fx$ is "everything is known by a to be an F" and thus requires that every individual is known to a. Even though this does not, as Stine (1974: 135) correctly observes, require that every individual is known to a under all its designations, the requirement seems to strong. For we should allow for genuine general knowledge, especially when the open sentences quantified over is a conditional. I may perfectly know that everything taller than 2 m is taller than 1 m without knowing everything in the universe.

1.3.7 Knowledge and belief

As remarked on p. 12, there are obvious structural connections between knowledge and belief, even if neither of them is taken to be reducible to the other. The most evident of these connections is that knowledge implies belief. It is because of this interconnection, that many have tried to analyse knowledge as some kind of successful or epistemically good belief (e.g., like (Gochet and Gribomont 2002: 13), by changing (D) for the stronger (T)). One problem with the principle that knowledge implies belief, however, is that beliefs are more closely tied up with awareness than is knowledge. It seems self-contradictory to speak of unconscious belief, while it is common practice to speak of 'implicit' or 'tacit' knowledge, knowledge which is operative in our behaviour without being explicitly represented in our heads.[32] Knowledge, it seems, is more easily ascribed from an external perspective than belief.

It might seem an attractive option to reverse the direction of explanation and to give an account of belief in terms of knowledge. Rather than considering knowledge to be successful belief, one would then see belief as failed knowledge. This is the route taken by Dretske (cf. 62) and in situation theory (cf. 83), AI[33] and is the one advocated by Williamson:

> "If believing p is, roughly, treating p as if one knew that p, then knowing is in that sense central to believing. Knowledge sets the standard of appropriateness for belief.

[32] Williamson gives the following example, while still sticking to the entailment thesis: "When the unconfident examinee, taking herself to be guessing, reliably gives correct dates as a result of forgotten history lessons, it is not an obvious misuse of English to classify her as knowing that the battle of Agincourt was in 1415 without believing that it was." (Williamson 2000: 42)

[33] Shoham and Moses take belief to be defeasible knowledge, explicate "ϕ is believed" as "it is know that either ϕ holds or else that the assumption is violated", where the assumption ϕ_{ass} is a parameter of the belief relation: $\mathsf{B}(\phi, \phi_{\text{ass}}) := \mathsf{K}(\phi_{ass} \rightarrow \phi) \wedge (\mathsf{K}\neg\phi_{ass} \rightarrow \mathsf{K}\phi)$ (Shoham and Moses 1989: 1169)

That does not imply that all cases of knowing are paradigmatic cases of believing, for one might know p while in a sense treating p as if one did not know p – that is, while treating p in ways untypical of those in which subjects treat what they know. Nevertheless, as a crude generalization, the further one is from knowing p, the less appropriate it is to believe p. Knowing is in that sense the best kind of believing. Mere believing is a kind of botched knowing." (Williamson 2000: 47)

This view of belief as failed knowledge fits surprisingly well with the ordinary strategy in epistemic modal logic, where, as observed in sct. 1.3.3, the step from knowledge to belief is usually taken to consist in the weakening of the truth-axiom (T) to a mere consistency requirement (D). The logic of knowledge then becomes a subsystem of the logic of belief.

The interplay between knowledge and belief, however, creates problems for epistemic modal logic. Given that people have inconsistent beliefs, the logic of belief, if there is some such thing, will have to be normative: it will be able to tell us not which beliefs we have, but which beliefs we should have, given some subset of beliefs we have. It will be concerned with the best use we make of our limited resources, describe in what way we should adjust our belief sets to new information and so on. None of these issues, however, seems immediately relevant to a logic of knowledge. It is not a matter of logic what knowledge is most useful to have, nor is it of immediate logical concern what knowledge we should have. In order for there to be knowledge, there has to be justification and truth – and both of these are not up for grasps; they are up to the world, not to the way we adapt ourselves to it. There is therefore some danger that the joint logical study of knowledge and belief incorporates elements which are extraneous to the former.

There are other problems for such an enterprise as well. There is a well-known argument, originating with Lenzen (1979), that **S5** as the logic of knowledge K and **KD45** as the logic of belief B[34] give us the contra-intuitive result that no false beliefs are possible when conjoined with the following two theses:[35]

(15) $\vdash \mathsf{K}p \to \mathsf{B}p$ entailment property

(16) $\vdash \mathsf{B}p \to \mathsf{BK}p$ (positive) certainty property

This principles, however, look plausible. What to do? Halpern recently suggested to restrict (15) to *objective* (non-modal) formulae, i.e. formulae not containing the K operator.

[34] It suffices, in fact, that $\vdash \neg(Bp \wedge B\neg p)$.

[35] The proof is the following: Assume, for reductio, $\vdash \neg p \wedge \mathsf{B}p$. By (16), $\vdash \mathsf{BK}p$. By $\vdash \neg p$ and (T), $\vdash \neg\mathsf{K}p$. By (S5), $\vdash \mathsf{K}\neg\mathsf{K}p$. By (15), $\vdash \mathsf{B}\neg\mathsf{K}p$. Hence $\vdash \mathsf{BK}p \wedge \mathsf{B}\neg\mathsf{K}p$.

In particular, he questions the desirability of having formulae like

(17) $\vdash \mathsf{K}\neg\mathsf{K}p \rightarrow \mathsf{B}\neg\mathsf{K}p$

To my eye, this just shows that K is not adequately interpreted as being a *knowledge* operator. For certainly knowledge, in the ordinary sense, implies belief. We will later see (in sct. 1.4.3) that the failure of the 'knowledge entails belief' principle gives us some reason to talk not about knowledge, but about information.[36]

1.3.8 Knowledge as a kind of information

Jaakko Hintikka (1962) seems to conceive of knowledge as awareness based on the right kind of information.[37] This comes close to Dretske's famous informational account of knowledge (cf. def. 2.2.5) and will be discussed in much detail later.

A conceptual assimilation on a different level, however, has some attractive features as well. Knowledge not only presupposes information or is caused by it, but seems to be a kind of information itself. Whenever I ask for information about x, I ask for what is known about x.

In what sense, if at all, may knowledge be taken to be a kind of information? I think this connection is best brought out by considering Williamson's arguments for identifying an agent's knowledge with his evidence (Williamson 2000: ch. 9).[38] Equating knowledge with evidence means to reverse the traditional order of explanation (where a definition of knowledge in terms of justification is sought); this, however, is not a fatal drawback, as we saw independent reasons for taking knowledge to be primitive (cf. sct. 1.3.7). Evidence,

[36]Even if we try to get along without (16) and choose – together with **S5** for knowledge – **D** for belief, we get the so-called "paradox of the perfect believer" (i.e. the contra-intuitive result $\vdash \mathsf{BK}p \rightarrow \mathsf{K}p$) from adding the following to (15) (cf. Gochet and Gribomont 2002: 16):

(18) $\vdash \mathsf{B}p \rightarrow \mathsf{KB}p$ introspection for beliefs

Frans Voorbraak's system, as reported in Gochet and Gribomont (2002), tries to circumvent the paradox by changing entailment (15) for positive certainty and $\vdash \mathsf{B}p \rightarrow \mathsf{BK}p$. As shown by Hoek (1993) and reported by (Gochet and Gribomont 2002: 19–20), (18), together with (15, (D) for belief and (S5) for knowledge leads to the catastrophic result that $\mathsf{B}p \leftrightarrow \mathsf{K}p$.

[37]On p. 120, he says: "he does not really know it, that is, his awareness is not based on information which would justify a claim to knowledge." He does not, however, say much more.

[38]Another information-theoretic perspective on knowledge and belief has been put forward by Fagin and Halpern (1992) who compare a conception of belief as generalised subjective probability with another tradition, conceiving of beliefs as representations of evidence.

for Williamson, is what we are justified to rely on.[39] What we are justified to rely on is what we know.

More generally, we are justified to rely on whatever information we have and it is rational to proportion our beliefs to the information available. Knowledge may therefore plausibly be taken to be *paradigmatic* information, information that is, so to say, in the best of its possible forms.

1.4 Information

1.4.1 What information might be

Information is certainly the fuzziest and least worked on of our three notions. It is incredibly popular and one is likely to find at least a passing reference to it in almost every book or article in the area. As Dretske remarked more than twenty years ago, however, there is surprisingly little conceptual work done on such fundamental questions as what information is, how it has to be modelled and so on.[40] It seems just a fancy catchword to wrap up whatever one happens to have to sell.

I will not offer a remedy to this predicament. What I will try to do, rather, is to investigate in what respect information may not only be seen as more fundamental as meaning (cf. 1.2.3) and knowledge (cf. 1.3.8) but in what ways it may be plausible taken to differ from both. On the basis of such an account, it will be possible to approach the vexed issue of the alleged veracity of information and the crucial question how information relates to possibility, a subject taken up again in sct. 4.4. The main focus of this thesis, however, is to present and evaluate a recently suggested 'logic of information' Barwise and Seligman (1997). While this chapter prepares the groundwork, the next two (ch. 2 and 3) will discuss two major influences, both from the theory of knowledge and the theory of meaning.

[39]Cf.: "When we prefer an hypothesis h to an hypothesis $h*$ because h explains our evidence e better than $h*$ does, we are standardly assuming e to be known; if we do not know e, why should h's capacity to explain e confirm h for us?" (Williamson 2000: 200)

[40]Cf. Dretke's apt remarks in the preface to his 1981 book: "A surprising number of books, and this includes textbooks, have the word *information* in their title without bothering to include it in their index." (Dretske 1981: ix) This situation did not much change in the following five years: "While we process information all the time, personally and with the aid of computers, there is no semblance of agreement as to the basic nature of information and information processing. The basic notions are lacking any commonly accepted philosophical and mathematical foundations." (Barwise 1986b: 137) It think – pace Devlin who claims that situation theory has found out what information is (Devlin 2001: 25) – that the same can still be said today.

I am far from wanting to suggest a new (or contribute to an already existing) 'philosophy of information'. Despite the apparently growing popularity of this label (as evidenced by the forthcoming publication of *The Blackwell Guide to the Philosophy of Computing and Information*) it is far from clear to me that there is such a thing, i.e. that it is possible to isolate enough fruitful philosophical problems, tied together by some sufficiently strong family resemblance, to stand up to expectations raised by such a presumptuous label. "Philosophy of information" is a term combining two notoriously ill-defined concepts. Vagueness being additive, one is naturally led to doubts whether the "new computational and information-theoretic approach in philosophy" (Floridi 2002a: 2) is more than hot air. This initial suspicion is vindicated, it seems to me, by a look at the literature.

Luciano Floridi, who, in recent years, has been one of the most active figures in the field, takes the importance of information to reside in what he takes to be the gradual virtualisation of society.[41] He emphatically joins others in what he sees as the new paradigm which will soon be established as a "mature field" (Floridi 2002a: 19) and even as a/the "forthcoming *philosophia prima*" (2002a: 20).[42] This enthusiasm may be partly due to his conviction that professional philosophers work on idle pseudo-problems (Floridi 2002a: 9) (because they have been "technically trained to work only in the narrow field in which they happen to find themselves" (Floridi 2002a: 10)) and to his conception of philosophy as "the last stage of reflection, where the semanticisation of being is pursued and kept open" (2002a: 12).

However, his proposal for a historical account of the philosophy of information before the 'digital revolution' (he cites Plato's *Phaedrus*, Descartes' *Meditations*, Nietzsche's *On the Use and Disadvantage of History for Life* and Popper's third world) may lead one to expect old wine in new bottles. And indeed he claims that "virtually any issue can be rephrased in informational terms" (Floridi 2002a: 18). Such a rephrasing, however, will only prove fruitful if it is based on a clear idea of what information is, which is exactly the issue. While he certainly thinks it is very important, it is rather unclear, however, what Floridi takes information to be.[43] It therefore seems to me that there is little to be learned

[41]He describes the latter, not very helpfully, as "metasemanticisation of narratives", "de-limitation of culture", "de-physicalisation of nature" and "hypostatisation (embodiment) of the conceptual environment designed and inhabited by the mind" in (Floridi 2002a: 7–8).

[42]Bynum and Moor write: "From time to time, major movements occur in philosophy. These movements begin with a few simple, but very fertile, ideas – ideas that provide philosophers with a new prism through which to view philosophical issues. Gradually, philosophical methods and problems are refined and understood in terms of these new notions. As novel and interesting philosophical results are obtained, the movement grows into an intellectual wave that travels throughout the discipline. A new philosophical paradigm emerges." (Bynum and Moor 1998: 1)

[43]Floridi (2002b) seems to think of information as purely mental: "The real pawn is an 'information

from the so-called "philosophy of information". Rather than inquiring into a supposedly fixed and known phenomenon, a more promising approach to our question seems to have a look at the ways in which talking of information may prove helpful in matters of epistemic logic.

1.4.2 Aboutness

Informational contexts[44] are neither truth-functional nor referentially transparent, that is they neither allow for truth-preserving substitution of equivalent sentences nor of coreferential singular terms. Unlike modal contexts, they do not even allow truth-preserving substitution of necessarily equivalent sentences. In this, they side both with epistemic and with meaning-ascription contexts.[45] They differ from meaning contexts and side with epistemic contexts, however, in the specific character of their referential opacity: they seem to allow for truth-preserving substitution of *necessarily coreferential* terms. If you tell me that the president of the United States is Texan, you do not tell me that Bush junior is Texan, for I might well ignore that Bush junior is the president of the United States and you do not tell me so. If I learn, however, that Cicero is a Roman orator, I thereby learn that Tully is a Roman orator.[46] The case for predicates is somehow intermediate. It is, of course, weird to say that by telling you:

> Cicero is Cicero
>
> or
>
> All equiangular triangles are equiangular.

object'. It is not a material thing but a mental entity ..." (Floridi 2002b: 4). However, he suggests in the very same article that information objects might have a moral status and even can be moral *agents* (Floridi 2002b: 11).

[44]As we will see later (in sct. 4.2), informational contexts are usefully taken to be of the following canonical form: a's being F carries the information that b is G.

[45]By the former, I mean sentences of the form "a knows that p", by the latter sentences of the forms ""p" means that p", ""a" expresses the individual concept a" and ""F" expresses the property F"(if the latter is based on a theory of predicates which allows fro necessarily coextensional predicates which are not identical).

[46]I am here assuming that proper names are rigid designators, referring to the same individual in all those worlds where this individual exists.

I give you the information that

> Cicero is Tully
>
> and
>
> All equiangular triangles are equilateral.

On the other hand, the following consideration seems to point towards extensionality: By telling you that all equiangular triangles are F, I also tell you something about all equilateral triangles. For sentences, the situation is clear: even necessary equivalence does not preserve informational value.[47]

These observations may be summed up in the following way: informational contexts are extensional with respect to the position which names the entity the information in question is *about*, but opaque with respect to the position where the information conveyed is represented.

Another connection between informational content and aboutness became important in the inductive logic tradition championed by Rudolf Carnap and Karl Popper. In Carnap (1942), Carnap offered two reconstructions of Popper's original idea in Popper (1935): we might identify the informational content of a statement either with the set of its logical consequences or with the set of its possible falsifiers. One of the problems with the first reconstruction is to prevent an explosion of the informational content due to the validity of $p \to p \vee q$, $p \to (q \to p)$ and the like. Taking a clue from Goodman (1961), Howard Smokler proposed to exclude sentences q from the informational content which are such that p implies the universal generalisation of one of the 'transforms' q' of q (q' being q with one individual or predicate constant replaced by a variable of corresponding degree) (Smokler 1966: 207). As S.G. Hair pointed out, however, this definition has some implausible consequences: it is not the case, e.g., that the informational content of a conjunction includes its conjuncts[48] and that informational content is transitive for languages including polyadic predicates or identity Hair (1969).[49]

Even if these problems could be fixed, a more general problem with this kind of

[47] You told me something about Bush junior, but nothing about Fermat's Last Theorem, though "The president of the United States is Texan" and "The president of the United States is Texan and $\forall n \in \mathbb{N}$, there are no $x, y, z \in \mathbb{N} \setminus \{0\}$ such that $n > 2 \wedge x^n + y^n = z^n$" are necessarily equivalent.

[48] To see this, note that $(\forall x Fx) \wedge Fa$ implies the universal generalisation of its conjunct Fa.

[49] The informational content of $\forall x(R(a,x) \vee R(x,a)) \wedge R(a,b)$ includes $R(a,b)$, the informational content of which contains $R(a,b) \vee R(b,a)$; the informational content of $\forall x(F(x) \vee G(x)) \wedge (F(a) \vee G(b)) \wedge a = b$ includes $(F(a) \vee G(b)) \wedge a = b$, the informational content of which includes $F(a) \vee G(a)$.

approach remains: if the informational content of p is given as a set of *sentences*, the question naturally arises what the informational content of these sentences is and whether it would not be more appropriate (in order to avoid a regress) to include their informational content, rather than the sentences themselves into the informational content of p. This move, however, would make the account immediately circular, the informational content of p being identified with (the fusion, union of) the informational contents of those sentences q such that $\vdash p \rightarrow q$. We will therefore not pursue further this route, coming back to the second of Carnap's reconstructions (informational content as set of possible falsifiers) in sct. 1.4.8.

The connection between aboutness and situation theory, mediated by the relation of carrying information, runs even deeper than that. The basic notion of a situation is that of the part of the world some set of statements is about.[50] Situations, in turn, may plausible be taken to be the entities any act of conveying information aims to describe. We will later see (in ch. 3), how this crucial linkage may be spelled out in more detail.

1.4.3 Information and knowledge

In this section, I want to explore further the similarities and differences between information and knowledge and to discuss the proposal to re-direct epistemic modal logic towards a logic of information, a topic taken up again in sct. 4.1.

We have seen, in sct. 1.3, that almost all of the traditional principles of epistemic modal logic are questionable if K is interpreted as modelling the actual use we make of "knowledge" in ascriptions of mental states to ourselves and to others. If I know that Sam knows that p, e.g., I am justified in believing that he will act on the basis of p, i.e. that he will pursue his goals and intend to satisfy his desires on the assumption that p. Given (Nec), e.g., this link is broken: I will misdescribe Sam's proof-searching behaviour by imputing on him knowledge of all logical truths. The same happens with (S4): I may truly ascribe him knowledge that p without being justified in expecting from him to be able to rule out all only potentially relevant background conditions on which his knowledge that p depends. The clivage between the notion modelled by the formal principles and the way they are used in ordinary life is most visible in the case of (S5): If I am justified in

[50] This becomes particularly clear in Devlin's account of the theory: "Without the situations to refer to, we would be forced to make theorist's assertions of the form "The agent acquires additional information about whatever in the world it is that the agent acquires that additional information about." By putting *situations* into our theorist's ontology, we avoid such awkward circumlocutions..." (1991: 71) Seligman puts the point succinctly: "On our account, a situation just is the sort of thing in which information can be said to be located." (1991: 290)

believing that Sam does not know that the watch he bought is a fake, I am precisely *not* justified in believing that he will act on knowledge of his ignorance: quite on the contrary, I may justifiably think, observing his behaviour, that he not only does not know that it is a fake but positively believes that it is original, hence that he does not know that he is ignorant about the real value of the watch.

Given this situation, it may therefore seem a good idea to conceive of epistemic modal not as a logic of knowledge, but as a logic of *information*.[51] Information, after all, exhibits some structural and inferential similarities to information. It is thus natural to begin the search for a logic of information by an examination of epistemic modal logic. This will be done on a more formal scale in sct. 4.5, but some preliminary remarks are in order.

A first, and cheap, advantage of changing "knowledge" for "information" is that the latter is much less entrenched than the first. It has become common practice to introduce the epistemically interpreted box operator with a whole bunch of caveats: we are told to understand "knowledge" in a wide sense, ascribable to non-human and even inanimate agents,[52] who are at the same time highly idealised and perfectly rational. "$\Box_a p$", epistemically interpreted, is then read as "'agent' a is ideally in a position to know". Sometimes it is claimed that "knowledge" is outright ambiguous and that the logic to be developed is intended to capture only one of its meanings:

> "We do not believe that there is a "right" model of knowledge. Different notions of knowledge are appropriate for different applications." (Fagin et al. 1995: 8)

The obvious difficulty with such a claim, not tackled by Fagin et al., is to distinguish these different notions, perhaps along contextualist lines, and to identify the one that is supposed to satisfy the formal conditions laid down in the subsequent 450 pages of the mentioned book.

Speaking of information rather than knowledge may in this situation be an attractive option.[53] For what information an epistemic agent has at his disposal depends on his

[51] Sam the proof-searcher may be said to have the information that some given set of sentences are logical truths. He just does not know how to make use of this information in an efficient and goal-guided way. Sam the car-driver has the information that the speedometer is working properly if he receives from it the information at what speed he is running. Sam the watch-buyer, lacking the information that it is not a fake, may still be said to have the information that he lacks this crucial piece of information about the watch. That he does not use it, while it is at his disposal, explains why he is acting foolishly.

[52] Cf. e.g.: "Our agents may be negotiators in a bargaining situation, communicating robots, or even components such as wires or message buffers in a complicated computer system. It may seem strange to think of wires as agents who know facts; however, as we shall see, it is useful to ascribe knowledge even to wires." (Fagin et al. 1995: 2)

[53] Not for everyone, I suppose. Keith Devlin has recently retracted from his former, 1991, enthusiasm

objective epistemic situation, rather than his beliefs about it (cf. p. 4). What information something carries, and, a fortiori, what information something carries for some given agent, can be described in impersonal terms, from an outside perspective. For this reason, "information" seems more suitable to idealisation than "knowledge", being less entangled with standards of reasoning and subjective deliberation.[54]

In its beginnings, epistemic modal logic was seen by its protagonists as a conceptual and philosophical investigation into the logical behaviour of a key notion in epistemology, which had resisted clear-cut definition for about two thousand years (cf. sct. 1.3). It soon became clear, however, that the logical apparatus used in modelling what was supposed to be the epistemic behaviour of real-life agents exaggerated their actual epistemic capacities. It is less than clear to me that the remedy to this empirical inadequacy is to claim for the system in question the status of a 'useful fiction' or idealisation. Idealisation, as Stalnaker remarked (Stalnaker 1991: 242), may mean two things: either epistemic logic is held to be a logic of knowledge in the ordinary sense, but applicable only to beings of an idealised kind (agents having unlimited memory capacity, infinite computational power and speed and so on); or epistemic logic does not model ordinary knowledge but knowledge in a special sense. In the first case, the obvious problem is to specify the characteristics we share with these idealised beings and thus the degree to which what is true of them carries over to us; in the second case, the difficulty is to spell out the relations of this idealised concept with our ordinary notion and to justify the claim to call it *knowledge*, albeit in an idealised sense. As it has turned out very difficult to isolate a plausible such notion of knowledge-in-an-idealised sense,[55] I will concentrate on the first horn.

for information and now goes for knowledge: "Though we often speak of information being "a valuable commodity", the value of information lies in its potential to be turned into knowledge. For, ultimately, knowledge is what makes the difference in what we can do, and the value of information depends upon the value of the knowledge to which it can lead." (Devlin 2001: 4) It is, however, very difficult to judge on the basis of his new book *InfoSense* how serious he is about this – leaving out the details, limiting himself to an "intuitive acquaintance with the underlying theory", which he calls "infosense" (Devlin 2001: 5) (which is still Barwise and Perry (1983)) and interested mainly in showing that "scientific understanding of information can be used to improve the way businesses – and individuals – manage information." (Devlin 2001: 3)

[54]In the section (1.3.7), we saw another reason to change "knowledge" for "information": Taking epistemic modal logics as logics of *knowledge*, makes it very difficult to spell out the connections between the latter and belief. Information does not have any such link with belief or awareness; there is no, not even defeasible, presumption that epistemic agents are aware of the information at their disposal, which is the information they may be said to *have*.

[55]A distinction between implicit and explicit knowledge and belief – and the corresponding 'logics of awareness' – seem of little help, for I may even explicitly believe the Peano axioms and still lack any cognitive attitude, be it implicit or explicit belief, to many of the theorems of arithmetic (cf. Gochet and Gribomont (2002: 46) for further criticisms). The general problem, as Stalnaker (1991: 249) remarked, is

Idealising epistemic agents is to misdescribe them.[56] If the misdescription is to be useful, it has to be accurate in some respects or accurate enough for practical purposes. An idealisation may be useful if it correctly represents the *structural* features of the phenomenon in question, while abstracting from their exact nature.[57] Epistemic logic, however, is already situated at the level of structural features: an 'idealising' logical principle (like (S4), for example) is not a straightening out, a simplification of or abstraction from a messy complicated reality which it describes correctly at a structural level, but imputes knowledge on agents they do not in fact have. It is not just a less-than-accurate or simplified description of reality, but a false one.

What then about the other possibility – that epistemic logic is accurate enough a description of the epistemic behaviour of our agents? Read in the straightforward way, this claim is untenable; for knowledge is an *absolute* concept which does not admit of degrees. One might want to say, however, is that epistemic logic describes not the actual epistemic behaviour of the agents in questions, but the behaviour they would have in an epistemically ideal world, a world, e.g., where they are free of contingent limitations of attention and other mental capacities. Such a strategy might be implemented in two ways: one possibility would be to read the knowledge operator "$\mathsf{K}_a p$" as "under idealised circumstances, agent a would know that p". The problem then is about iteration: the intended reading of "$\mathsf{K}_a \mathsf{K}_a p$" is "under idealised circumstances, a would know that he knows that p" rather than "under idealised circumstances, a would know that he would know that p under idealised circumstances". Are we then claiming that "K" is ambiguous, having different senses in embedded and unembedded contexts? Such an ambiguity would invalidate clearly valid inference patterns, however. To argue from $\vdash \mathsf{K}_a(\mathsf{K}_a p)$ to $\vdash \mathsf{K}_a p$, using (T), would be to commit a fallacy of equivocation. Another possibility would be to

that idealisation goes in the wrong direction: the more idealised the notion of knowledge, the *less* need for logical omniscience. Another general problem is that resource-boundedness draws a distinction between accessible and inaccessible beliefs which is askew to the one between explicit and implicit beliefs, at least if the latter distinction is understood in some realistic way: explicit beliefs may still be inaccessible, while I may access, by easy derivation, some implicit beliefs of mine as well.

[56]Stalnaker made the observation that logically omniscient agents are not justifiable taken to be *reasoners* at all: "Any context where an agent engages in reasoning is a context that is distorted by the assumption of deductive omniscience, since reasoning (at least deductive reasoning) is an activity that deductively omniscient agents have no use for. Deliberation, to the extent that it is thought of as a rational process of figuring out what one should do given one's priorities and expectations is an activity that is unnecessary for the deductively omniscient." (Stalnaker 1991: 244)

[57]A good example of this is given by decision theory, representing the subjective certainties of epistemic agents, their degrees of belief, by exact numerical values. There is, on the one hand, a useful heuristics, in terms of dispositions to accept (or consider favourable) certain bets, to determine these values. On the other hand, the precise values are almost never important: all that matters is their respective place in an ordering.

claim that epistemic modal logic, only tells us what our agents would know if they were fully rational etc. Given that a knows that p, the claim would then be that (S4), e.g., tells us that a, if fully rational etc., would know that a knows that p. The transition from $\mathsf{K}_a p$ to $\mathsf{K}_a \mathsf{K}_a p$, so to say, would take place in an epistemically ideal world. The problem with this manoeuvre is that we cannot, as theorists, really provide the premiss necessary to get the result $\mathsf{K}_a \mathsf{K}_a p$, so interpreted. For this would be to make a claim about an epistemically ideal world (that a knows that p in such a world) of which we know nothing.

Another way in which epistemic modal logic might be a description of an idealised reality is normative (deontic or axiological). We could take $\mathsf{K}_a p$ to mean that a should know that p (deontic) or that it would be good if a knew that p (axiological). (S4) would then tell us that a should know that a knows that p if a were as a should be. Apart from the difficulty of providing the relevant premiss, that a should know that p in the first place, we encounter here a second difficulty which brings us back to the difference between knowledge and belief (cf. sct. 1.3.7). In the case of belief, it seems to me, a counterfactual or normative reading of the operator seems possible. Doxastic logic may thus be interpreted as telling us what we should be believe or rather what we should be believe if we believe some other things. Epistemic logic, however, cannot do such a thing – not only (as it has been remarked in sct. 1.3.7) because justification and truth are objective and normatively neutral, but also because knowledge is not volitional: we cannot decide to know.[58] Epistemic logic is thus incapable of giving us norms of behaviour, in the sense doxastic logic may tell us what is rational to believe.

There is a further problem about idealisation in epistemic logic, one not encountered by otherwise comparable idealisations in, say, mechanics or thermodynamics. Whenever we dealing with multi-agent systems, e.g., we are not simply assuming that our agents are omniscient, perfectly rational, have unbounded resources and so on. We furthermore assume that they *know* this of each other. This means that we are not only modelling their behaviour only under somehow idealised circumstances, but that we are actually imputing false beliefs on them and falsely classify such beliefs as knowledge.

I conclude, then, that there is no clear sense in which epistemic modal logic may be said to deal with an idealised notion of knowledge. The problem, in a nutshell, is that

[58]This holds even if knowledge, as Williamson believes, is a mental state – not all mental states are volitional, e.g. fearing is not. It is, of course, a debated question whether belief is volitional, whether we may *decide* to believe. But even if it is not, however, there remains a contrast. We may avoid to believe, e.g. by staying at home or by adopting very sceptical standards. We may not in the same way, however, avoid to know (except, of course, by avoiding to belief in the first place). Whether a belief qualifies as knowledge is an objective matter, about which we cannot do anything. We will come back to this externalist aspect of knowledge in our discussion of Dretske in ch. 2).

knowledge and belief are mental states, regularities among the actual epistemic behaviour of human agents, and not the objective basis of such mental states:

> "[...] the problem is that we need to understand knowledge and belief as capacities and dispositions – states that involve the capacity to access information, and not just its storage – in order to distinguish what we actually know and believe, in the ordinary sense, from what we know and believe only implicitly. We can do this only by bringing the uses to which knowledge and belief are put into the concepts of knowledge and belief themselves, but, on the face of it, it does not seem that when we attribute knowledge or belief to someone we are making any claims about what the agent plans to do with that information." (Stalnaker 1991: 253)

Given this entanglement of knowledge with actual information processing behaviour of human agents, it seems natural and promising to focus instead on the underlying 'objective commodity', the information *available* to the agent.

The issue of transitivity may illustrate this general point. (S4), generating an infinite hierarchy of knowledge claims, clearly demands for some kind of anti-realistic reading. Jaakko Hintikka, in his ground-breaking work *Knowledge and Belief* (1962), gave the semantic conditions on the belief and knowledge operators he introduced in terms of the indefensibility of certain systems of model sets, which he took to be partial descriptions of some possible state of affairs. (S4) is then just built into the very framework, for, as it has been correctly noted by E.J. Lemmon, "indefensibility" is an elaborate term of art:

> "Defensibility is immunity to a certain *kind* of criticism: namely, the criticism that you can be shown wrong on the basis of the consequences of what you already know." (Lemmon 1965: 382)

While the issue of whether knowledge is transitive has been subject of an intense debate (cf. sct. 1.3.4), information flow is undoubtedly – even *essentially* – transitive.[59] In that perspective, information may be seen as idealised or potential knowledge:

> "I take this [the Xerox principle, i.e. transitivity of information flow] to be a regulative principle, something inherent in and essential to the ordinary idea of information, something that any *theory* of information should preserve. For if one can learn from A that B, and one can learn from B that C, then one should be *able* to learn from A that C (whether one *does* learn it or not is another matter)." (Dretske 1981: 57)

The notion of information in use here is an idealised one: what matters is the information carried by the signal, not the information a suitably placed agent might learn from it. It is informativity *in principle* that matters:

[59]This corresponds to the requirement that information flow obeys the Xerox principle (2.2.1 in ch. 2), acknowledged by Barwise, Perry, Seligman, Devlin and, of course, Dretske himself.

"What information a signal carries is what it is capable of "telling" us, telling us *truly*, about another state of affairs. Roughly speaking, information is that commodity capable of yielding knowledge, and what information a signal carries is what we can learn from it. [...] A state of affairs contains information about X to just that extent to which a suitably placed observer could learn something about X by consulting it." (Dretske 1981: 44-45)

We already noted (in sct. 1.4.1), however, that "information" may legitimately be taken to be a weaker notion than "knowledge". Knowledge, but not information, depends on the agent being responsive to what his environment affords him:[60]

"What is the difference between an agent p knowing some fact σ, and the agent simply carrying the information σ? [...] An agent knows σ if he not only carries the information σ, but moreover, the information is carried in a way that is tied up with the agent's abilities to act. [...] Information travels at the speed of logic, genuine knowledge only travels at the speed of cognition and inference." (Barwise 1988: 204)
"If an agent $\mathcal{A}$ is in a state that is linked to an informational item X by some prevailing constraint C, then we say that $\mathcal{A}$ *has* (or *encodes*) that information X. For $\mathcal{A}$ to *know* X, on the other hand, involves considerably more. In particular, it requires that $\mathcal{A}$ be 'aware' of the possession of this information, and takes cogniscience of this information in determining or guiding future activity." (Devlin 1991: 107)

There is another important connection between knowledge and information. It is not just that both presuppose the existence of connections between the representing system (the receiver of the signal, the belief-forming individual) and the state of affairs the system represents (the signal, the situation known). The important point is that both require these connections to be reliable or non-accidental. In the knowledge case, this requirement corresponds to the crucial 'third component' in (1.3.1), turning true beliefs into knowledge, i.e. the reliability of the belief-forming mechanism, its responsiveness to counterfactual error possibilities or the stability of the underlying causal link. In the case of information, this is built into the prior possibility space wherein the signal locates the state of affairs it is about. This has been stressed by Barwise and Seligman:

"Despite the wide range of regularities that permit information flow, it is important to distinguish genuine regularities from merely accidental, or "statistical" regularities. Accidents are not sufficient for information flow." (1997: 9)

It is crucial, then, for a theory of information – as it is for a theory of knowledge – to have the resources to distinguish accidental correlations from real informational linkages.

[60] This means that "information" is most plausibly taken to denote the *available* information, not the information that is actually used or processed and that may constitute knowledge. "Information" thus is just the notion ascribers of externalist knowledge in computer science are in need of (cf. Stalnaker 1999b: 261).

1.4.4 Information and meaning

We already saw, in sct. 1.2.3, that meaning may be conceived of as a kind of information. Information, however, embraces both non-natural and natural meaning and it is not limited to cases where we can speak of meaning at all. Even the information conveyed by an utterance is not exhausted by what may, even in an extended way including reflexive content, be called its meaning or semantic content. Information not only is a more general and widespread phenomenon than meaning, it is also what underlies it.[61]

Another crucial difference between meaning and information is that the first is in a sense "autonomous" while the latter depends on what is already known, in particular on what is known about the relevant alternative possibilities. The muddy children puzzle may illustrate this:

> n children, after having played around in the garden, are sitting in a circle, every one seeing all the others. The mother announces a game: everyone who rightly claims to have some mud on his forehead gets a cookie. Suppose only two children have a muddy forehead. The mother enters the circle and makes the following public announcement: "At least one of you has a dirty forehead." Nothing happens. So she repeats the announcement. Now two children raise their hands and correctly claim to have mud on their forehead. In general, if there are m muddy children, it takes m repetitions of the same announcement to enable the muddy children to legitimately claim their cookies. How can the mere repetition of the same sentence have this effect?

Supposing there are m muddy children, the proof is as follows by induction on m:

> If $m = 1$, the child seeing no muddy child will correctly identify himself as the dirty one after the first announcement. Suppose there are $n + 1$ muddy children, the announcement has made n times and no child has yet said anything. Any one of the $n+1$ muddy children sees n others who are muddy and will thus conclude from their silence after the n^{th} announcement that they see at least n muddy children as well and from this that he or she has to have some mud on his forehead.

This piece of reasoning not only exploits what the sentence "At least one of you has mud on his forehead" *means*, but also what one can conclude from the facts that the others heard and understood it, that they act rationally, and that they above all want a cookie. While the sentence "At least one of you has mud on his forehead" keeps its constant meaning from utterance to utterance (to suppose otherwise, would be to posit gratuitous

[61] This is apparent in the operational definition of information predominant in Information Systems Theory, as reported in Floridi (2001: 3), where information is taken to be (something relevantly similar to) data plus meaning.

indexicality where there does not seem any), the *information* one gains by understanding it increases with the times it already has been repeated. This information conveyed by an utterance of this sentence after m repetitions is that there are at least $m+1$ muddy children: this is what the sentence tells the children given what they know already. This is why repeating the same utterance can be making a point.

We thus have a concrete instance of the utility of distinguishing sharply between what a sentence means (as used at a given occasion and in a certain context) and the information conveyed by it (in that context).

1.4.5 Veracity

Another crucial difference between meaning and information, according to some (e.g. to Dretske 1981: 44), is that information is veridical: Whenever it is right to say that s carries the information that p, or that a got the information that p or that I would like to have that information that p, p must be true. In this view, it is strictly speaking false to talk of misinformation: misinformation is not, as its name suggests, a kind of information, but not information at all:[62]

> "... misinformation is not a species of information any more than belief is a species of knowledge." (Dretske 1990: 115)
> "False information is not an inferior kind of information; it just is not information." (Grice 1989a: 371)

Others, however, have argued to the contrary:

> "'x misinforms y that p' entails $\neg p$, but 'x informs y that p' does not entail that p ... information does not require truth, and information need not be true; but misinforming requires falsehood, and misinformation must be false." (Fox 1983: 160–161, 193)

The issue of veracity, whether information is, should be or is best modelled as being veridical, i.e. whether pieces of information are always, should be or are best seen as being true, is at bottom terminological (as has been remarked in sct. 1.1): there is just no settled semantic core to this notion that could possibly settle it. Everything, it seems, depend on the ultimate usefulness of modelling it in one or rather the other way.

[62]This difference is that between a Dutch duck, which is a kind of duck, and a rubber or decoy duck, which is not (cf. Dretske 1981: 45). We use "false" in this way, though only in special contexts, e.g. when we speak of a false friend or a false alarm. I do not think, however, that this is the sense of "false" used in "false information" (cf. below).

That there is some space of manoeuvre here, i.e. that the core meaning does not force upon us a particular decision on the veracity issue, does not entail that "information" is ambiguous, as it has been claimed to be e.g. by Seligman and Moss

> "Notoriously, the word 'information' is ambiguous. Traditionally, the possession of information that Smith is an anarchist like the knowledge of the same implies that Smith *is* an anarchist. With the advent of mechanisms for the gathering, storing and retrieval of large amounts of information, the word has taken on a more neutral meaning. According to more modern usage, to be adopted here, the information that Smith is an anarchist may be stored on a computer file in ASCII code, transmitted across the globe, converted to speech and heard over a loud speaker, without Smith's ever having had a subversive political thought in his life." (Seligman and Moss 1997: 278)

The existence of such a minimal sense not entailing veracity does not, by itself, establish ambiguity. It only testifies to the fact that, as with many other factive concepts, there may be a pragmatical use in cancelling out the 'success component'.[63] Exactly the same could be said of knowledge: If I say, for example, that most people know what they want, I do not want to claim that whatever they would express in the form of "I want it to be the case that p" is made true by what the best psychological theory would attribute them. I just want to say that they are prepared to make some judgements and that I have no particular reason not to take their words at face value. It is hence useful – even if cumbersome – to have the concept of *potential knowledge*, true of justified beliefs that would qualify as knowledge if only they were true.

Analogously, the useful distinction brought to the fore by the veracity discussion is not the one between information and misinformation (false contents of messages, possibly produced with an intention to deceive) but the one between information and potential information, where the last term just indicates that the modeller does not want to take a stance on the ultimate truth of what plays the behavioural rôle of information.[64]

[63] I fully agree with Hintikka that the apparent need to postulate ambiguity is often a symptom that the core sense has not yet been correctly accounted for: "In fact, what first gives the appearance of several unrelated senses of a word or an expression is often a symptom of the presence of nothing more than different uses [...], accountable for in terms of one basic meaning." (Hintikka 1968: 12)

[64] The distinction between information and potential information is closely connected to Israel's and Perry's distinction between information and informational content: "Informational content is only information when the constraints and connecting facts are factual. If a signal carries the information that P, then P is true. But a signal can have the *informational* content that P relative to a constraint and a connecting fact, even though P is not true. This happens when the constraint or connecting "fact" or both are not factual." (Israel and Perry 1991: 147) Again, this has a parallel in the knowledge case: many epistemic logicians have characterised the state of mind modelled by "$\Box_a p$" as "potential knowledge" (cf. sct. 1.4.3).

1.4.6 Information and data

Given that information is veridical, how do we call potential information, items that might be information if they are true, that purport to inform us about their subject matter but of which we do not (yet) know whether they are linked to it in the way required to qualify as information? I suggest that we call such pieces of potential information "data". "Data", like "information", is a notion whose core meaning may be captured by different – incompatible though equally acceptable – definitions. I propose to use it for "information minus truth" both because it is useful to have a short term for this and because it seems (to me) that this matches to a sufficient degree the use made of "data" in computer science.

Data are not just uninterpreted syntactical items.[65] As the word is standardly used, data may be useful, complex, encoded, compressed and so on, and thus have properties which cannot meaningfully be attributed to strings of letters. It thus corresponds to what is called a "message" in communication theory (cf. sct. 2.1.1), something that represents a bit of information as obtaining or non-obtaining.

I think "data", used in this sense, is rather close in meaning to the use of "information" noted above, to designate something that may be meaning (if conveyed in an utterance) or knowledge (if believed in a justifiable way), but yet is neither and more like the raw material that may be processed in different ways.[66] Devlin seems to use "information" in this way when he writes:

> "There are many cases where information is conveyed without the listener, or indeed the speaker, either knowing or believing that information. For example, the speaker or listener might be a computer, which can acquire and dispense vast amounts of information but which neither believes nor knows anything. Or again, one suspects that a great many television newsreaders neither know nor believe all the information they read to camera. Conveying information does not require belief or knowledge of that information..." (Devlin 1991: 242)

[65]Something along these lines has been claimed by Floridi: "... since MTC [the mathematical theory of communication] is a theory of information without meaning, and information - meaning = data, *mathematical theory of data communication* is a far more appropriate description than *information theory*." (Floridi 2002c: 20) Floridi (2003: 23) identifies data with "completely uninterpreted information". Put in slogan, Devlin believes that "Information = Data + Meaning", "Knowledge = Internalized Information + Ability to utilize the information" (Devlin 2001: 14-15) and explicates the former as "Information = Representation + Constraint" (Devlin 2001: 34). Even if this is certainly a possible way of using words, it is not very illuminating. Though you may, by saying "Data is what you get when your computer prints out a table of figures or a list of names and addresses" (Devlin 2001: 16)), teach the word to a non-native speaker, you will hardly give him a firm grasp on the concept (you may, after all, equally get information or even knowledge that way).

[66]The distinction between information and data is what makes satirical Hamlet's answer to Polonius' question "What do you read, my lord?": "Words, words, words" (*Hamlet*, Act II, Scene 2).

I think that the use of "data" would be preferable here and suggest that we use it in this way in the following.

1.4.7 Relativity

Another largely terminological issue centres around the question whether information is relative. While most participants in the discussion seem to agree that information can be plausibly taken to be an objective (i.e. not man-made) feature of the world, while how much and what information a given signal carries depends on constraints which may be man-made, different authors have emphasised different parts of this two-sided picture.

Israel and Perry (1990: 4), along with Fred Dretske[67] and Jon Barwise,[68] have repeatedly emphasised that information is not an intrinsic property of a signal, but that a signal carries information only relative to some constraint or other. Keith Devlin, on the other hand, seems to ascribe to Shannon's communication theory (to be discussed in sct. 2.1.1 below) a rather different picture:

> "Perhaps *information* should be regarded as (or maybe *is*) a basic property of the universe, alongside matter and energy (and being ultimately interconvertible with them). In such a theory (or suggestion for a theory, to be more precise), information would be an intrinsic measure of the structure and order in parts or all of the universe, being closely related to entropy (and in some sense its inverse). This approach would, of course, fit in quite well with the classic work of Shannon on *amounts* of information." (Devlin 1991: 2)

As with the veracity issue, however, the quarrel is merely apparent.[69] Devlin does not want to suggest that the information carried is intrinsic to the signal, but only that it is

[67]This feature of Dretske's position is perhaps less explicit that one might wish. Different interpreters have thought that the relevant passage in *Knowledge and the Flow of Information* (Dretske 1981: 80) was some sort of slip: "With a minor and eliminable exception (p. 78–80), Dretske defines an *objective* concept of information."(Sterelny 1983: 208) This is exegetically false, as evidenced e.g. by (Dretske 1990: 117). We will see later, in sct. 2.3, that Dretske went into the right direction, but did not go far enough.

[68]"Information is almost always, perhaps always, dependent on unarticulated background conditions. That is, the relations that allow one situation s to contain information about another situation s' are conditional on certain background conditions C being met by the environment." (Barwise 1986b: 151)

[69]In fact, Devlin perfectly acknowledges the kind of relativity of information stressed by Perry and Israel. Cf. "... different agents are capable of extracting different information from the same source (situation)." (Devlin 1991: 14) In recent work, however, he seems to have completely changed his mind about the issue. Imagining the case of having bought a computer that comes with a Japanese instruction manual, he writes: "One one level, you could be said to "have" all the information required to operate the machine. But unless you understand Japanese, that information is not going to do you any good. In fact, for all practical purposes, you *don't* have the information." (Devlin 2001: 32) He completes the volte-face by adding in a footnote: "In fact, the information management techniques described in this book, being motivated by highly practical purposes, will tell you that you *really* don't have the information –

intrinsic to the whole consisting of the signal and whatever constraints and relations that signal has to the rest of the world. This, however, is both unproblematic and uninteresting: every property is intrinsic to the whole consisting not only of its bearer but of anything to which the bearer could possibly be related in virtue of having this property. The important point, much stressed by Dretske (cf. his externalist definitions of information content (2.2.3) and knowledge (2.2.5) to be discussed below), is that an agent receiving information does not have to be reflectively aware of this fact nor of the preconditions on which his receiving *information* depends.[70] Information is *objective* in the sense that it is independent of representation by the observer, but *relational* in the sense that it exists in virtue of there being constraints or informational dependencies between different parts of our world (cf. Barwise 1986b: 139).[71]

There is another aspect to the issue of the relativity of information: in some cases, the most notable being the transmission of information by use of a symbolic, e.g. linguistic, system, the constraints in virtue of which some item carries information are themselves mind-dependent. What information a word carries, e.g., is obviously relative to what that word can justifiably taken to mean, which in turn depends on what languages and uses of languages there are. This phenomenon is nicely brought out by the notion of reflexive content which, as we saw in sct. 1.4.4, may be plausibly taken to capture a part of the informational content of an utterance left out by the traditional notion of meaning:

> "One should think of a belief as having a hierarchy of contents, as more and more is taken as *given* and detached from the truth-conditions, culminating in the referential content. The other contents, the attributive and reflexive contents, are not different beliefs but different aspects of the same belief that can be characterized as attunements to more reflexive contents." (Perry 2001a: 138)

What John Perry here says about beliefs applies equally to utterances considered as signals, i.e. items carrying information. Relative to the conventional constraints we consider, these may be said to carry different pieces of information. Even if we have settled, e.g., on the constraints of English spoken in such-and-such a region, however, an utterance may be said to carry different pieces of information relative to different ways of taking it.[72]

period." To me, this seems to get him from the problematic "p, but for all practical purposes, $\neg p$" to the unacceptable "p but *really* $\neg p$".

[70]This is also emphasised by Barwise and Seligman: "... the definition of information content (and hence of knowledge) only depends on the conditions being met, not on our knowing that they are met." (Barwise and Seligman 1997: 21)

[71]We will later see how this relativity of information reappears in the quantitative theory of information (sct. 2.1.2) and underlies the crucial importance of constraints in the logic of information to be developed (sct. 4.3).

[72]As we discussed this problematic already (in sct. 1.2.1), we may leave the matter here.

Even if the question *whether* information is relative is terminological, the issue of *how* and *to what* it is relative is a vivid one, to which we shall come back in due term (e.g. in the evaluation of Dretske's theory in sct. 2.3).

1.4.8 Information and possibilities

Central to the link between information and knowledge is the following principle, which we might call the "Inverse Relationship Principle":

1.4.1 (Inverse Relationship Principle). *Whenever there is an increase in information there is a decrease in possibilities, and vice versa.*

Jon Barwise (1997: 491) calls (1.4.1) the "main idea of informationalism" (his label for the "pragmatic theory of modality" he wants to develop).[73]

John Perry has even argued that a version of (1.4.1) gives us a *definition* of information (Perry 2001a: 42):

1.4.2 (Perry's account of information). *An event e carries the information that p (relative to some constraint and background) if e could not occur unless it was the case that p (assuming the constraints and background).* [74]

We have to be careful here, however. It is one thing to say that information and decrease in possibilities are correlated – nobody, I think, will dispute that. It is a quite different and much stronger claim that one of them can be used to define the other or to give an account, a model or an explanation of it. And even if this stronger claim is made, everything hinges on the order of explanation. Normally, the direction goes from information to possibilities, the latter being assumed as given. Barwise's aim, however, is to give an "informational accountof possibilities" (Barwise 1997: 491) – as opposed to an possibilistic account of information.

In early work on information, (1.4.1) has usually be interpreted probabilistically. Bar-Hillel and Carnap have identified the informativeness inf of some proposition p with the negative logarithm of its probability:

(19) $\mathsf{Inf}(p) = -\log P(p)$

[73]He actually operates with a stronger principle than (1.4.1), restricted to *available* information. His reason for this is that he wants to countenance impossible worlds, a proposal we will discuss again in sct. 4.8.2 and 7.1.1 and 7.2.1.

[74]Perry (2001b: 22) gives a slightly different account of the informational content of an event as "what the world must be like in order for the event to have occurred".

Bar-Hillel and Carnap take the semantic content of some proposition p to be the set of all state-descriptions which are inconsistent with it. The greater this set, the greater the reduction in uncertainty when we learn that p is true. A paradoxical feature of this approach is that the contradictory proposition $q \wedge \neg q$ has the most inclusive semantic content:

> "It might perhaps, at first, seem strange that a self-contradictory sentence, hence one which no ideal receiver would accept, is regarded as carrying with it the most inclusive information. It should, however, be emphasised that semantic information is here not meant as implying truth. A false sentence which happens to say much is thereby highly informative in our sense. Whether the information it carries is true or false, scientifically valuable or not, and so forth, does not concern us. A self-contradictory sentence asserts too much; it is too informative to be true." (Bar-Hillel and Carnap 1953: 229)

Even if it is granted – contrary to what has been argued in sct. 1.4.5 – that information does not have to be true, the paradox still remains. Intuitively, learning the inconsistent proposition does not reduce our uncertainty in any way: it does not eliminate any *real* possibilities. We will later see how (variants of) this problem re-appear in Dretske's theory (sct. 2.3) and how it and they are best approached (sct. 4.8.2).

In communication theory, the statistical theory of information we will discuss in connection with Dretske's theory in chapter (2), (1.4.1) underlies the crucial identification of informational value with *entropy*, taken as a measure of the degree of 'randomness' of a physical system. Though often obscured by an anthropomorphic manner of speaking, this may be seen to be the main idea in communication theory:

> "Information is, we must steadily remember, a measure of one's freedom of choice in selecting a message." (Weaver 1949: 18)

In other words: The signal's being in a certain way carries information because and insofar as it could be otherwise. To give information is to exclude possibilities. These possibilities, as Dretske repeatedly emphasis, have to be real, in the sense that they represent relevant alternatives, ways the signal *actually* might be.[75] This is embodied in Dretske's notion

[75] He introduced the notion of relevant alternatives already in his (1970): "What is explained is a function of two things – not only the fact [...], but also the range of relevant alternatives. A relevant alternative is an alternative that might have been realized *in the existing circumstances* if the actual state of affairs had not materialized." (Dretske 1970: 44, emphasis mine). I emphasised "in the existing circumstances" to prevent the misreading of Stine who interprets the quoted sentence as "an alternative is relevant only if there is some reason to think that it is true" and the "might" as indicating physical possibility (Stine 1976: 253–254). Dretske's idea is not at all to just to take all physical possibilities as relevant alternatives, but to restrict them to those that might be alternatives *to a given situation*, i.e. sharing its background features.

of channel conditions (def. 2.2.6), discussed in detail later. Dretske maintains against the skeptic that the question of when a channel is secure or what are the relevant possibilities is an empirical one which cannot be settled by imagining or ignoring possibilities:

> "The fact that we can imagine circumstances in which a signal would be equivocal, the fact that we can imagine possibilities that a signal does not eliminate, does not, by itself, show that the signal is equivocal. What has to be shown is that these imagined circumstances *can* occur, that these imagined possibilities do obtain." (Dretske 1981: 131)

This, in Dretske's view, is just the upshot of information's being an "objective commodity", i.e. a commodity which

> "... though we speak of it as being *in* a signal or *at* a receiver, is constituted by the network of relationships existing between a signal and a source. This is objective in the sense that the amount of information transmitted is independent of its potential use, interpretation, or even recognition." (Dretske 1981: 82, cf. also the quote on p. 4)

It is clear, however, that just treating the issue as an empirical one does not settle all relevant theoretical questions.

The most notable difficulty of the relevant alternatives approach is that the set of alternatives relevant to the evaluation of some given kind of knowledge claim is not constant but varies with the broader setting in which this claim is mounted. Let us agree that the identification of something as a barn on the basis of visual appearances obtained under optimum conditions while driving through the country-side justifies an agent in the belief that there is a barn out there, even if his evidence does not distinguish between real and fake barns. Intuitively, however, the situation changes if we add the additional information that the agent in question is driving through a region which is full of expertly made papier-maché facsimiles of barns. Dretske seems right in claiming that this additional piece of information in itself changes the picture, independent of whether it is known by the agent to obtain. In this case, we might say, that the visual impression of our agent carries the information that the object seen is not a fake barn. The facsimile case, then, is a relevant alternative to consider – if only in this, and not in other settings.

Let us now have a look at how Dretske exploits the alleged objectivity of information to give an account of knowledge, belief and information flow.

Chapter 2

Knowledge: Dretske's information-theoretic account

2.1 Basic ideas and notions

2.1.1 Communication theory

Dretske starts his investigation of the nature of information discussing a theory which does not, according to his own judgement (Dretske 1981: x), even address the question of interest to him: he calls this theory, launched by Claude Shannon's article "The Mathematical Theory of Communication" Shannon (1948), *communication theory.* Communication theory is concerned with the amount of information transmitted by a given signal. Communication theory takes the amount of information generated by the occurrence of an event (or the realization of a state of affairs) to be the *reduction in uncertainty*, the elimination of (relevant) alternatives to that event or state of affairs. This reduction is measured by the number of binary decisions necessary to uniquely specify the signal among all the relevant possibilities, i.e. the depth of a binary tree representing uniquely the one that actually occurred. Therefore, the *average amount of information (or entropy)* $I(s)$ *generated by the source* s (i.e. some process reducing n equally likely possibilities to one) is the following:

(1) $I(s) = log_2(n)$

In the general case of n events s_i with different probabilities $P(s_i)$, we get the entropy by adding the surprisal values of the different possibilities $I(s_i) = log_2(\frac{1}{P(s_i)})$ and simplifying by $log(1/x) = -log(x)$:

$$(2) \quad I(s) = -\sum_{i=1}^{n} P(s_i) log_2(P(s_i))$$

To be passed along, information must not only be generated, but also received. On the way, some of the information generated at s may be lost. Not all of the possible ways the signal may be at s correspond to possible ways the signal may arrive at another situation r. In addition, among the relevant possibilities at r may be some where the signal was distorted or augmented with irrelevant extra information on its way.

We therefore need a measure for the informational dependency between a source situation s and a receiver situation r, a measure indicating how much of the information generated at s is received at r. This is the amount of information $I_s(r)$ received from or about s in situation r (how much of $I(r)$ is received from s). To calculate $I_s(r)$, we either can take $I(r)$ (the information of the signal at r) and subtract from it the *noise* at r with respect to s (the information at r that is independent of what happened at s), or, alternatively, $I(s)$ (the information of the signal at s) and subtract from it the *equivocation* at s with respect to r (the amount of information generated at s but not transmitted to r). If we have a case of information transmission from s to r with n possibilities $s_1, \dots s_n$ at s and m possibilities $r_1, \dots r_m$ at r, we compute the average noise $N(s)$ at s and the average equivocation $E(r)$ at r as following:

$$(3) \quad N(s_i) = -\sum_{j=1}^{m} P(r_j|s_i) log(P(r_j|s_i)) \qquad N(s) = \sum_{i=1}^{n} N(s_i)P(s_i)$$

$$(4) \quad E(r_j) = -\sum_{i=1}^{n} P(s_i|r_j) log(P(s_i|r_j)) \qquad E(r) = \sum_{j=1}^{m} E(r_j)P(r_j)$$

Equivocation and noise, being measures of the independence of what happens at s and at r, make clear that information flow, i.e. the 'successful' transmission of a signal (i.e. of a signal with positive entropy), basically is just lawful or at least regular dependence of what happens at two different places. This dependence may, but does not have to be, causal.

Even in cases where informational dependencies are underwritten by causal relations, however, the two relationships have to be distinguished: the flow of information depends not only on actual, but also on merely possible causal relations, those that would have occurred if their causal antecedent had existed (Dretske 1981: 28, 34). The informational value of an event thus depends on the range of causes that could have produced it. Information can also be transmitted across 'ghost channels', i.e. channels existing in virtue of the common dependency (causal or otherwise) of what happens at two places on a source: in this case, there is flow of information between the two 'receivers', in virtue of their being linked to a common source (Dretske 1981: 38).

2.1.2 Quantitative and qualitative theories of information

Communication theory falls short of a qualitative or conceptual analysis of information in several respects. It is concerned with averages, e.g. the average information generated by a source or transmitted along a channel, and has no use for the notion of the surprisal value of a single signal.[1] In principle, there is no limit to what can be learned from a particular signal, because its informational content depends on a prior partition of the possibility space (Dretske 1981: 51): the finer this partition, the more informational content a signal coming from that source carries. A qualitative theory of information, therefore, cannot take this partition for granted.

This will be a recurrent theme in the following (cf. sct. 4.2): the dependence of information on notation, a feature Dretske unduly neglects. Suppose the following is the output of a calculation:

$$x := 3,123\ldots$$

How much information do we receive by being told this result? Quite a lot, even indefinitely much, as there is no upper bound to the decimals one may want to take into account. Under another form, e.g.

$$x := \pi$$

we get considerably less information, though it seems reasonable to maintain that, in another sense, we get more information, any decimal expansion like the above just being an approximation to π. This dependence on notation does not make information subjective, however.[2] Informational value is objective in the sense that how much information a

[1] cf. Warren Weaver: "The concept of information applies not to the individual messages (as the concept of meaning would), but rather to the situation as a whole, the unit information indicating that in this situation one has an amount of freedom of choice, in selecting a message, which it is convenient to regard as a standard or unit amount." (Weaver 1949: 9)

[2] Although multiplying the number of possibilities at the source taken into account by more detailed descriptions increases the equivocation of potential signals, $I_s(r)$, the amount of information transmitted,

message carries is not a function of how much information the recipient thinks it carries but only of the actual possibilities at the source and the conditional probabilities of these possibilities after the message has been received (Dretske 1981: 55–56).

Most importantly, however, communication theory does not address the *content* of a given piece of information. Two signals carrying the same *amount* of information, of course, may still tell us different things and thus differ in their value and their causes and effects. So the qualitative side of information, i.e. *what*, as opposed to *how much* information a signal carries, has to be taken into account. This means that the possibilities at the source have to be interpreted differently: we have to think of them not as ways the signal might be (relative to some encoding system) but as ways the signal might represent the world as being. An example might bring this out: Large parts of Shannon's original article (1948) are devoted to establishing the amount of information carried by sequences of letters in English, thereby investigating the engineering question what the best encoding system for such sequences would be. So it may be said, for example, that the sequence of letters "proselyt" carries much more information than "man", being much rarer in common usage. It is clear that this is not the amount of information of interest to Dretske: for him, the important possibilities to exclude are not all the different sequences of letters that might be composed out of the same stock of 26 primitive symbols, but the ways the world might be but is not, if what the signal says is true.

Despite of these fundamental differences there is, in Dretske's view, much to be learnt for semanticists from communication theory:

> "Communication theory does not tell us what information is. It ignores questions having to do with the *content* of signals, what *specific information* they carry, in order to describe *how much* information they carry. [...] Nevertheless, in telling us *how much* information a signal carries, communication theory imposes constraints on what information a signal *can* carry. These constraints, in turn, can be used to develop an account of what information a signal *does* carry." (Dretske 1981: 41)

Warren Weaver, in his *Recent Contributions to the Mathematical Theory of Communication* (1949), made the same point, distinguishing three levels of communications problems, the technical ("How accurately can the symbols of communication be transmitted?"), the semantic ("How precisely do the transmitted symbols convey the desired meaning?") and the effectiveness problem ("How effectively does the received meaning affect conduct in the desired way?"), and claiming that any limitation at the level of the first question automatically applies also to adequate answers to the other two.

remains stable, because the value of $I(s)$, the information generated at the source, is also increased (Dretske 1981: 62).

Taking over only the limitative results, the only quantities of interest for comparisons in information theory are the amount of information generated by a particular event s_a (its surprisal value), $I(s_a) = log_2(1/p(s_a))$, and the amount of information carried by a particular signal r_a about s_a, $I_s(r_a) = I(s_a) - E(r_a)$, where $E(r_a)$, as before, stands for equivocation (Dretske 1981: 52).[3]

In order to calculate the amount of information generated by a particular event or received by a particular signal, one has to know (1) the alternative possibilities to that event, (2) their associated probabilities, and (3) the conditional probabilities of these alternative possibilities, given what we know about (the epistemological situation of) the receiver. Even if such knowledge is unattainable in practice, communication theory may provide us with a heuristics to estimate and compare the informational values of different events and different signals. The most important use of the formulae for surprise value and information carriage, however, is the direct comparison of the two: Dretske wants to require, as a necessary condition for information flow, that the subject receives as much information as is generated by the state of affairs he knows to obtain, i.e. that $I(s_a)$ is less or equal to $I_s(r_a)$.

2.2 A Semantic Theory of Information

2.2.1 Information

Dretske starts developing his so-called "semantic" theory of information by imposing certain constraints on it. The first such regulative principle is that it should make sense to speak of a *flow of information* (Dretske 1981: 57):

2.2.1 (Xerox principle). *If A carries the information that B, and B carries the information that C, then A carries the information that C.*

It follows from this principle that partial transmission of information is not possible: I cannot receive the information that a is F by a signal carrying less information than that generated by a's being F: $I_s(r_a)$ must at least be equal to $I(s_a)$ for information flow to take place. Otherwise some finite chain would serve as a counterexample to transitivity:

[3]So Dretske disagrees with the proposal of Weaver to assimilate the future theory of information to statistical communication theory: "Language must be designed (or developed) with a view to the totality of things that man may wish to say; but not being able to accomplish everything, it too should do as well as possible as often as possible. That is to say, it too should deal with its task statistically." (Weaver 1949: 27)

> "What communication theory (together, of course, with the xerox principle) tells us is that for the communication of *content*, for the transmission of a *message*, not just any amount of information will do. One needs *all* the information associated with the content." (Dretske 1981: 60)

In Dretske's view, the Xerox principle, together with two semantic conditions, determines what information is:

2.2.2 (Communication and semantic conditions). *Any acceptable definition of information must impose the following three conditions on a signal's r carrying the information that s is F:*

> **communication** *r carries as much information about s as would be generated by s's being F.*
>
> **veracity** *s is F.*
>
> **aboutness** *The quantity of information r carries about s is (or includes) the quantity generated by s's being F.*

The veracity principle we already discussed (cf. sct. 1.4.5) clearly is a consequence of Dretske's general approach, for non-obtaining states of affairs do not have a surprisal value at all and therefore do not generate any information. Unfortunately, Dretske does not tell us much about the third condition (which I call the aboutness condition), just remarking that while the first two are individually necessary, they are not jointly sufficient and that the third therefore just marks whatever is needed to capture the relevant notion of a signal carrying the information that s is F (Dretske 1981: 65). It is needed to take care of sources generating pieces of information which have the same amount but are about different properties of the source and hence different. We already met a very similar requirement in Grice's theory of non-natural meaning (see p. 10 and also sct. 1.4.2), where it was required for someone to mean something that he intended to achieve his intention of inducing a belief in his audience *by means of* the audience's recognition of this intention. This bears a strong resemblance to Dretske's second semantic condition: in both cases, there has to be a way back, i.e. the information transmitted must be specific enough to allow the receiver not only to individuate the fact that generated it but also to recognise it as the fact the information is about. We will later see the importance of these two conditions in our discussion of constraints.

Dretske proposes the following definition of the *information content* of a signal, claiming that it meets the three conditions:

Definition 2.2.3 (Informational content). *A signal r* carries the information *that s is F iff the conditional probability of s's being F, given r (and the receiver's prior knowledge k), is 1 (but, given k alone, less than 1).*

The requirement of a conditional probability of 1 means that there is a nomic regularity between the two event types which nomically precludes r's occurrence when s is not F (Dretske 1981: 245, n.1). The informational content of a signal is a de re content expressed by the ascription of an open sentence "... is F" to an individual s: it does not vary with different descriptions of what the information is about. Whenever a signal increases the conditional probability of a possibility A, but not to 1, we may say that it carries the information that s is probably in state A.

Even if we agree that – based on the argument from the Xerox principle sketched below – the conditional probability has to be 1, there is the further question whether (2.2.3) states not only a necessary, but also a sufficient condition. Does it not let too much information in? Dretske evades the problem of omniscience (that any signal informs us of all the necessary truths) by his somehow ad hoc proviso that the conditional probability be less than 1 given the prior knowledge alone. He does not evade a more general problem, however, which is the phenomenon of nesting, for which Dretske cannot claim to have satisfactory solution.[4] We will come back to this problem and the rôle of the background knowledge k in sct. 2.3. Here, then, is the definition:

Definition 2.2.4 (Nesting). *The information that t is G is* nested *in s's being F iff s's being F carries the information that t is G.*

This phenomenon has different consequences. For one, it means that there cannot be asymmetric information about necessary connections: Whenever two states of affairs are connected by a nomic regularity, information that one of them obtains is information that the other obtains (Dretske 1981: 71). Second, it does not, given nesting, make sense to speak of *the* informational content of a signal: not only is the information carried by a signal deductively closed, it is closed even under any necessary connections whatsoever:

> "Generally speaking, a signal carries a great variety of different informational contents, a great variety of different pieces of information, and although these pieces of information may be related to each other (e.g. logically), they are nonetheless *different* pieces of information." (Dretske 1981: 72)

[4] As he himself acknowledges: "Frankly, I do not know what to say about our knowledge of those truths that have an informational measure of zero (i.e. the necessary truths)." (Dretske (1981: 264, n. 3), also cited in Barwise (1997: 491)).

The problem for Dretske, however, is how to distinguish these different pieces of information.[5] It seems that Dretske's theory, making abstraction of the fact that different classificatory systems may allow us to describe in different ways *the* information carried by a signal, is not able to draw this necessary distinctions. We will later see how Barwise's and Seligman's theory fixes this bug (cf. 6.2.1).

Not every regularity, however, gives rise to nesting. Since "F" and "G" can be coextensional even if the probability of s's being F conditional on it's being G is not 1, a signal carrying the information that s is F does not necessarily carry the information that s is G even if every F is a G and vice versa:

> "Correlations are relevant to the determination of informational relationships *only insofar* as these correlations are manifestations of underlying lawful regularities." (Dretske 1981: 247, n.6)

Necessary connections are not the only source of nesting phenomena: the prior knowledge k may equally be important, particularly the receiver's prior knowledge about the possibilities at the source. This background knowledge, however, is relevant only insofar as it affects the value of $I(s)$; in particular, the receiver does not have to know that the signal is reliable (i.e. that $I(s) = I_s(r)$). Dretske thus has an *externalist* theory of information flow: what is important is the factual quality of the channel, not what it is taken to be by the communicating parties (Dretske 1981: 81).

2.2.2 Knowledge

This externalist perspective is brought to the fore in Dretske's account of knowledge:

Definition 2.2.5 (Knowledge). *When there is a positive amount of information associated with s's being F: K* knows that *s is F iff K's belief that s is F is caused (or causally sustained) by the information that s is F.*

(2.2.5) differs in several respects from the classic definition of knowledge (1.3.1). It applies, first, only to empirical knowledge (Dretske 1981: 241, n.2). Second, it is intended to be a characterisation of perceptual knowledge, i.e. of de re knowledge in situations where perceptual (noncognitive) factors fix what is the something known to be F (Dretske 1981: 86). To save (2.2.5) from circularity (arising from the fact that the definition of informational content (2.2.3) uses the notion of (prior) knowledge), it has to be possible to receive information that does not depend on any prior knowledge about the source.

[5] We will come back to this point in sct. 2.3.

"Being caused by the information that s is F" means being caused by that property of the signal in virtue of which it carries the information that s is F. Even if K's belief that s is F has not been caused by that information (if he, say, acquired the belief from a fortune-teller), it still may be knowledge, if it is (now) causally sustained by the information that s is F. It is causally sustained by it iff this piece of information affects the belief in such a way that it would, in the absence of other contributory causes, suffice for the existence of the belief (Dretske 1981: 89).[6] It does not follow from (2.2.5) that the relevant belief that s is F *itself* carries the information that s is F: it may well be such that it would have been caused by many things other than the relevant information, though – in this specific case – it has been caused by that information. It does so only in the absence of other causes of that belief. Therefore, not everyone who knows that s is F is someone from whom one can learn that s is F, i.e. which can cause beliefs in others which qualify as knowledge.

Dretske claims that his analysis can handle Gettier problems (cf. sct. 1.3). Gettier showed that justified true belief is not sufficient for knowledge if one allows the justification in question to be fallible. The additional premiss, that justification is inherited by the known logical consequences of what one is justified in believing, plays no significant role. In Dretske's view, Gettier-type difficulties cannot be raised against his account that makes knowledge conditional on the receipt of information because this is some sort of infallible (objective) 'justification' (Dretske 1981: 97). If I believe things on the basis of false beliefs, which, however, were formed on the basis of knowledge, I may or may not know: it depends on whether the causal influence of the information reaches through the intermediate false belief, making it causally dispensable. In this way, (2.2.5) provides us with a criterion to assess the qualifications needed to uphold the general (and in its generality implausible) "no false lemmas" principle which was taken to be required to handle the Gettier cases and which states, in full generality, that no belief formed on the basis of false beliefs may constitute knowledge.

Dretske defends the necessity of (2.2.5) as follows: If there were a less-than-absolute threshold, some sufficiently high probability (short of 1) sufficing for knowledge, a lottery paradox argument could be brought against it, since equivocation is additive (when the sources are independent of one another) and belief and justification (and hence knowledge) are closed under conjunction (Dretske 1981: 100). If one would not require that knowledge-constituting beliefs have to be caused by the information, i.e. if one would allow for positive equivocation, communication chains of arbitrary length could be construed

[6] This counterfactual, as Keith Devlin (1991: 183) notes, is problematic, for only a minimal change in one's belief set is relevant to its truth.

so that equivocation accumulates and a communication breakdown occurs at an arbitrary point in the chain: we would then have to claim, paradoxically, that at some point in the chain, even if the knowledge of the sender is as good as that of all his predecessors and even if he is communicating over the same channel as all the others before him, the knowledge communicated along the chain becomes, at this point, suddenly incommunicable (Dretske 1981: 104). As this is clearly intolerable, transmission of the full information is required, i.e. no positive equivocation can be allowed.

(2.2.5) may be seen as a refinement of a requirement Dretske defended some ten years earlier: He then considered it both necessary and sufficient for a's knowing that p that a believes that p on the basis of *conclusive reasons* for p, where a conclusive reason for p is an experiential state q of a such that a would not be in q unless it were the case that p (Dretske 1971: 4, 17).[7] He interprets this counterfactual dependency as stronger even than Goldman's causal connection, even though, like the informational account, it is limited to the token events in question (Dretske 1971: 9–10).[8]

2.2.3 Channels

Dretske construes factual knowledge as an absolute, not as a comparative concept (Dretske 1981: 107): although information about s can come in degrees, information that s is F cannot. Admitting that knowledge is an absolute concept is not, however, to surrender to scepticism, for we are still free to regiment what counts as, e.g., equivocation. The irregularities the skeptic purports to find in communication and which in his view are responsible for the (putative) fact that we do not really know what we pretend to know do not count as irregularities for purposes of communication (Dretske 1981: 110-111). To pull the sceptics' teeth, Dretske asks us to distinguish between

(1) the information (about a source) a signal carries and
(2) the channel on which the delivery of information depends.

Dretske's answer to the sceptic, then, is that it is a mistake to suppose that a signal's dependence on factors about which it carries no information makes it carry no or less information. The sceptic is confusing the channel with the source.

[7] This accounts corresponds roughly to what Barwise and Seligman call the "possible-worlds information content" (Barwise and Seligman 1997: 18).

[8] He does not yet claim, in Dretske (1971), that a singular connection strong enough to support counterfactuals is sufficient. He tries rather to capture the particularity by insisting that the general connection holds only relative to some, possibly quite specific, set of circumstances.

Though different things, (1) and (2) stand in a complex relationship. If the communication channel has possible states in which the same type of signal would be transmitted without it's carrying that information, then the signal does not carry the information unless the channel is known to be not in that possible state: this is just to say that known sources of error must be ruled out. If, however, the channel's being not in that deceiving state produces no new information (because it's being in that space is not a relevant alternative scenario), then the information-bearing signal does not have to be self-authenticating in that way. The channel's not being in that state is then called a *channel condition* (Dretske 1981: 116).[9]

Definition 2.2.6 (Communication channel). *The* channel of communication *is that set of existing conditions (on which the signal depends) that either (1) generate no (relevant) information, or (2) generate only redundant information (from the point of view of the receiver).*

We will later see how this notion may help us to understand the importance of channels in Barwise's and Seligman's theory of information flow (cf. sct. 7.2). Dretske is externalist not only about informational content and knowledge, but also about channels: What qualifies a condition as a channel condition is not that it is known to be stable or to lack relevant alternative states, but that it is *in fact* stable and *in fact* generates no new information (Dretske 1981: 119). If information could not be transmitted without simultaneously transmitting information about the channel's reliability, a regress would follow and information transmission would be impossible. Conditions which have no relevant alternative states or whose alternative states have been excluded by prior test and calibration (whether this is known or not) are part of the fixed framework within which equivocation and hence information is reckoned (Dretske 1981: 123).

Even though Dretske is externalist about channels, he does not mean to deny that what qualifies as a channel condition in one setting may become a relevant condition of the signal in another, a condition about which the signal carries information. On the contrary, exactly this is what happens in the calibration of measuring devices.

The notion of relevant alternatives which is interdefinable with that of channel conditions (the signal's being F qualifies as a relevant alternative at the source iff it is not a channel condition) is thus a robustly realistic one: Although there can always be equivo-

[9] Again, this notion may be seen as a refinement and improvement of what Dretske called the "fixed set of circumstances" we have to assume to make sense of subjunctive counterfactuals and which do therefore not enter into their truth-conditions; rather, the scope of application of the respective counterfactuals is restricted to situations in which these circumstances obtain (Dretske 1971: 12).

cation of which we are unaware, taking something to be a (relevant) possibility does not make it so (Dretske 1981: 127). Suspicions that neither affect the cause (information) nor the effect (belief) cannot destroy knowledge but can at best undermine our knowledge (or reasons for believing) that we know. The distinction between relevant and irrelevant alternatives then becomes just the distinction between a source *about* which information is received and a channel *over* which information is received.

2.2.4 Perception, Content, Beliefs and Meaning

Analogously to the distinction between analog and digital ways of representing (continuous) properties, Dretske distinguishes between two ways of representing facts (Dretske 1981: 136):

> "S carries the information that t is F in digital form if and only if (1) S carries the information that t is F, and (2) there is no other piece of information, t is K, which is such that the information that t is F is nested in t's being K, but not vice versa." (Dretske 1981: 260, n. 2)

A signal thus carries the information that s is F *in digital form* iff the signal carries no information about s that is not already nested in s's being F. Otherwise (i.e. iff it carries more specific information about s), it carries the information that s is F *in analog form.* What a signal carries in digital form is the most specific piece of information it carries about s.[10]

While sensation and sensory representation are essentially analog, cognition is digital.[11] Dretske thus takes the difference between perceptual experience and the knowledge or belief that is normally consequent upon that experience to be a difference in coding, but not in subject matter or information carrying (Dretske 1981: 143).[12]

[10] A statement of p typically carries the information that p in digital form, whereas a picture carries it in analog form (unless p is maximally specific). For pattern recognition, e.g., we need a device that extracts the information a picture carries in analog form, digitalises it and thus makes a distinction between relevant and irrelevant features of the picture.

[11] Sensory processes are means to deliver information within a richer matrix of information (hence in analog form) to the cognitive centres for their selective use (and for digital storage): "It is the successful conversion of information into (appropriate) digital form that constitutes the essence of cognitive activity." (Dretske 1981: 142)

[12] In an analogous way, he explains the difference between non-propositional and propositional seeing, which is of great importance to situation theory. It is by learning how to extract information that is already available and recoding it in digital form, that children pass from seeing to seeing that, from experience to belief (Dretske 1981: 144)

Belief, according to Dretske, is that capacity that distinguishes genuine cognitive systems from such conduits of information as thermostats, voltmeters, and tape recorders. All information-processing systems, however, occupy intentional states by carrying information:

> "To describe a physical state as carrying information about a source is to describe it as occupying a certain intentional state relative to that source." (Dretske 1981: 172)

Given that carrying information qualifies for Dretske some given system already for occupying intentional states, one is naturally led to ask how he draws the important distinction between simple measuring devices like thermostats and voltmeters and cognitive systems like people, capable of higher-order intentional states. Dretske does so by distinguishing different orders of intentionality and different corresponding kinds of content: these different kinds of content differ in that they are closed to different degrees under nesting (cf. def. 2.2.4). The first kind, what he calls *informational content* (i.e. content exhibiting the first order of intentionality) may fail to be closed under material implication.[13] Content of the second order is neither closed under material implication nor under implications which hold by natural law.[14] *Semantic content* (i.e. content of the third order of intentionality), finally, does not even have to be closed under analytically (nor logically) necessary implication.[15] Different kinds of content, then, are distinguished with respect to their specifity, i.e. according to the strength of the connections between items of information they may tell apart.

The distinction is helpful, I think, to capture another crucial difference between meaning and information (cf. sct. 1.2.3). That mere information content fails to exhibit the second order of intentionality is just to say that it may contain nested information.[16] It is therefore in general not possible to ascribe to physical structures a determinate or exclusive informational content.

Meaning and belief, however, are both situated at the level of semantic content.

[13] This means that a signal has informational content iff it is consistent to assume that it has the content that t is F, that all Fs are Gs and that it does not have the content that t is G.

[14] This means that a signal has content exhibiting the second order of intentionality iff it is consistent to assume that it has the content that t is F, that it is a natural law that all Fs are Gs and that it does not have the content that t is G.

[15] This means that a signal has semantic content iff it is consistent to assume that it has the content that t is F, that it is analytically necessary that all Fs are Gs and that it does not have the content that t is G.

[16] Cf.: "Information processing systems are incapable of separating the information that t is F from the information that is nested in t's being F. It is impossible to design a filter, for example, that will pass the information that t is F while blocking the information that t is G." (Dretske 1981: 171)

They not even nest analytically equivalent pieces of information. So while my speedometer gives me information not only about the speed of my car, but about all factors nomically connected with it, my belief that this glass contains water differs in content not only from a belief that this glass contains H_2O but does not even entail a belief that this glass contains something.

The distinction between mere informational and full-blown semantic content is related to the analog/digital distinction (cf. p. 58) in the following way: While analog information is nested and thus qualifies at best for the first order of intentionality, information carried in digital form may exhibit the third order of intentionality (Dretske 1981: 260, n. 3). Dretske identifies the semantic content of a signal with the information it carries in digital form (1981: 177), the piece of information in which all the other information it carries is nested. For Dretske, it is the capacity for digitalising that makes systems cognitive and distinguishes, say, human minds from speedo-meters.[17]

Semantic contents hence differ from mere informational contents in that they may be specified uniquely. Any given system exhibiting the third order of intentionality, however, may encode different semantic contents *about different objects* in digital form. Dretske identifies the semantic content of a signal with its *completely digitalised content*, the information it carries in its *outermost informational shell*, i.e. as that piece of information in which all other information it carries is (either nomically or analytically) nested:

Definition 2.2.7 (Semantic content). *Structure S has the fact that t is F as its* semantic content *iff:*

- *S carries the information that t is F, and*
- *S carries no other piece of information, r is G, which is such that the information that t is F is nested (nomically or analytically) in r's being G.*[18]

The main difference, according to Dretske, between a belief-forming system and a simple measuring device is that the former, but not the latter, is capable of completely digitalising the information it carries; in the latter case, this information is always embedded in larger

[17] Cf. also: "If a system is to display genuine cognitive properties, it must assign a *sameness of output* to *differences of input*." (Dretske 1981: 183) By digitalising a particular piece of information, the system abstracts, generalises and categorises.

[18] It is because of this sensitivity and counterfactual responsiveness to a *particular* piece of information, namely their semantic content, that semantic structures may be said to *highlight* one particular one of the components of the information carried by the incoming signal: "In virtue of the singular way it codes information, in virtue of its selectivity, a semantic structure may be viewed as a system's *interpretation* of incoming, information-bearing, signals." (Dretske 1981: 181)

informational shells depicting the state of those more proximal events on which the delivery of information depends.[19]

There are two main problems with (2.2.7): on the one hand, it makes having semantic content rather difficult; on the other hand, it does not address the problem of accounting for error. To see the first problem, consider the fact that for a to know that p is for a to be in a state which has the semantic content that p. The problem then is the following:

> "[...] suppose it is possible for s to be F without x knowing it (even implicitly). Then any state of x that carries the information that s is F will carry more information about s: that s is (implicitly) known by x to be F. Perhaps the account of aboutness will say that this is not really information *about* s. But however aboutness is explained, if there are facts about s that are nomically or conceptually necessary for knowledge that s is F, but not necessary for s to be F, then knowledge that s is F in digital form, in Dretske's sense, will not be possible." (Stalnaker 1999b: 264, n. 6)

The solution, as Stalnaker recognises,[20] is to make room for local information flow, underwritten by conditional constraints, i.e. to distinguish between knowledge, under condition ψ, that ϕ and knowledge that $\phi \wedge \psi$. To do this, we need a way to model constraints and information channels as objects, first-order citizens, and thus to reify the distinction between source and channel conditions. We will see later (in sct. 4.8.3) that Barwise and Seligman (1997) give us a way to do this.

The information carried by a signal and a fortiori the information is carries in completely digitalised form (which, according to def. 2.2.7, is its semantic content) cannot be false. How then can Dretske hope to identify beliefs with semantic structures of one kind or other? By treating misinformation as some kind of exception. Here is what he says about misrepresentation by maps:

> "A particular configuration of marks can *say* (mean) that there is a lake in the park without there actually being a lake in the park (without actually carrying this piece of information) because this particular configuration of marks is an instance (token) of a general type of configuration which *does* have this information-carrying function." (Dretske 1981: 192)

[19]Such simple measuring devices, therefore, do not have an outermost informational shell (Dretske 1981: 260, n. 6) and therefore lack semantic content.

[20]Cf.: "The problem is that it is too much to ask that the agent have the capacity to make its actions depend on whether ϕ in all cases in order for it to know whether ϕ in some particular case. Suppose I can make my actions depend on whether ϕ [and hence, according to Stalnaker's view, that I know that ϕ] only under condition ψ. Then provided that condition ψ obtains, won't this be enough to make available my knowledge whether ϕ? But the problem is to distinguish the case for which this description seems appropriate from the general case in which one's knowledge that ϕ is implicit in one's knowledge of something that entails it." (Stalnaker 1999b: 265)

The picture emerging, then, is the following: Given some information about our surroundings, we manufacture meanings or concepts by some sort of abstraction. These concepts once in place, they can be triggered by signals which lack the relevant information[21] and we thus produce tokens of them which have meaning (being tokens of meaningful types) without carrying information (and therefore without constituting knowledge). Belief, in Dretske's view, is basically failed knowledge.[22]

To be able to ascribe meaning to internal states which do not carry information, Dretske is forced to depart from his general information-theoretic outlook and to adopt a form of behaviourism: the meaning of an internal state is determined by the effects such a state has on the system's behaviour:

> "What is believed is determined by the etiology of those structures that are manifested in behavior." (Dretske 1981: 202)

Dretske turns to behaviourism also to mend another shortcut of his theory, namely its inability to ascribe to signals *specific* pieces of information, a is F rather than b is G when Fs are (either by natural law, by logic or language) Gs (Dretske 1981: 215).[23] One of the concepts being necessarily complex, he argues, there will be enough 'cognitive' (his name for 'behaviouristic') differences between them to distinguish the different pieces of information according to their influence on behaviour:

> "Having identical semantic contents is not enough to make two structures the same *cognitive* structure. They must, in addition, be *functionally* indistinguishable for the system of which they are a part." (Dretske 1981: 217–218)

This seems to get things wrong, however. The fact that I may believe that my glass contains water without believing that it contains H_2O is not adequately explained by the latter being a composite and the former being a 'primitive' concept. For all I know, both or none of these concepts might be composite. This brings us now to an evaluation of Dretske's theory.

[21] It is not exactly correct to speak, as Dretske does (Dretske 1981: 193) of triggering in this context, for concepts are *types* of internal states which are not triggered but tokened.

[22] This seems the correct interpretation to me, though he presumably would not put it that way. The distinction between knowledge as successful belief and belief as failed knowledge is Barwise's and Perry's. We will discuss it further on p. 83.

[23] This corresponds to his third desideratum for a theory of information noted in (2.2.2) on p. 52.

2.3 Criticism and later developments

Dretske's theory, though brilliantly suggestive, interconnecting domains which are at first sight very different, has some severe limitations. One of the most important shortcomings is its inability to account for the information carried by necessary truths (cf. sct. 4.8.2 for a general discussion). Another problem, due to its abstraction from the problem of coding, which is central to the engineering point of view, is its neglect of notation, i.e. the fact that it applies only to *de re* (notationally invariant) propositional content (Dretske 1981: 67). It is therefore unable to account for the similarities information has with meaning (cf. sct. 1.4.4). The two problems are interconnected, e.g. in his hopeless account of explaining the informativity of necessary truths: a necessary truth, according to him, may be informative because we do not know by which cognitive structures it is encoded (though we know – even a priori! – *that* it is encoded):

> "The investigation [whether robins are birds] may be necessary, *not* in order to acquire information in excess of what one has when one has the information that the objects under observation are robins (for one cannot *get* the information that s is a robin *without* getting the information that s is a bird), but in order to bring about a consolidation of one's representational structures. The investigation may be necessary, that is, in order to solve what is basically a *coding* problem: the problem, namely of determining which cognitive structures (if any) encode information that is already encoded, either in whole or in part, by other cognitive structures." (Dretske 1981: 219)

This seems just wrong. I may sincerely doubt whether penguins are birds or falsely believe that all (non-mutilated, adult etc.) birds can fly. I may very well receive the information that some animal in front of me is a penguin without thereby coming to know that it is a bird. When I investigate into the truth about this (for me) open question I learn something – not about my encoding of information I possess in any case but about the world. It just seems trivially true that necessary truths can be learned and that belief in them can fall short of knowledge for other reasons than coding problems.[24]

Another, but related, family of problems pertains to nesting. Though nesting is important, Dretske's account of it (2.2.4) fuses two very different phenomena. It is one thing to allow for pieces of information *contained* in one another and another, quite different

[24] One of Dretske's moves in response to this problem, i.e. the ad hoc presumption that one may fail to learn nested pieces of information not by lacking the concepts, but by lacking the beliefs (Dretske 1981: 218), is efficiently blocked by the kind of counterexample provided e.g. by Christopher Maloney (Maloney 1983: 27), where the cognitive agent, due to the antecedent belief that water is H_2O, acquires the belief that a certain raindrop is H_2O caused by that information, without, however, thereby coming to know this.

thing, to make every piece of information explode by including in it everything that on its basis has probability 1. Harmless cases of nesting are ready at hand. Information, by its relation to knowledge, is e.g. closed under conjunction: a signal that carries both the information that p and the information that q may unproblematically be said to carry the information that $p \wedge q$; the fact that $p \wedge q$ carries the information that p and it carries the information that q. If I tell you that the thief was 2 meter tall, I tell you that he was taller than 1.5 meter. If you learn that my cat Bruce is pregnant, you learn that I have a cat and that it is called Bruce. Information is not, however, closed under disjunction. If I tell you that I am sick, I am not telling you that either I am thick or the moon is made of green cheese. Neither is information flow transmitted over any kind of connection. Even though he successfully distinguishes information flow from causal dependence, Dretske does not manage to isolate broad enough a notion of informational connection.[25]

As Dretske acknowledges (Dretske 1981: 179), his distinction between different orders of intentionality does not solve the most difficult problems, which are not problems of nesting or implication but of equivalence. Nomically or analytically coextensional properties cannot be distinguished even by semantic content. What is more, his definition (2.2.7) makes semantic content cognitively intransparent in an extreme way. To decide whether p is part of the semantic content of a source s, we have to overview all relevant possibilities at s to rule out that there is another, more general, piece of information in which the information that p is nested. So Dretske's notion of semantic content cannot do enough and what it can do it achieves only at the price of cognitive inaccessibility.

Another weakness of Dretske's theory is his behaviouristic account of belief. It is very questionable whether behavioural effects are varied and fine-grained enough to bestow on cognitive structures the specific contents Dretske wants them to.[26] But even if this is granted, the behaviouristic account of belief leads Dretske to a misconstrual of the relation between knowledge and belief. Even if we agree (as I argued we should in sct. 1.3.7) that neither belief nor knowledge is straightforwardly definable in terms of the other, it is quite another thing to make, like Dretske, belief a *harder* thing to achieve than knowledge. It is, according to Dretske (1981: 218), because I have to form the corresponding *belief* that robins are birds, that not every information about a robin makes me know that it is a bird (given that robins are essentially birds).

A fifth and last general problem, which we will have to face also in connection with

[25]The problem has been known as the "disjunction problem" (cf. Sturdee (1997: 90) and Stalnaker (1999b: 267, n. 8)), though the diagnosis given is mine.

[26]This is, at least, the upshot of the famous arguments for the indeterminacy of translation and the inscrutability of reference first exposed and argued for by Quine in *Word and Object* Quine (1950).

Barwise's and Perry's situation theory and which will turn out to be one of the main motivations for the account of information to be presented in the second half of this work is Dretske's account of error (and the account of belief motivated by it). Dretske is forced, by taking information as basic and construing it as veridical, to interpret error (false belief, misrepresentation) as an occasional malfunctioning of an otherwise properly working cognitive system. The semantics is done for the standard case: knowledge, true belief, representation based on information. The structures invoked in this account are thereafter 'reified', taken to have an existence (and semantic content) on their own, independent of their functioning as links in an informational chain. Error is then explained as deployment of these mechanisms under non-normal circumstances.

It straightforwardly follows from such a strategy that error – by the theory's lights – can be but a relatively rare, local phenomenon. Global or even massive error is excluded by definition. Though there have been philosophers trying to argue positively for such a view – be it with transcendental arguments like Kant or from an interpretationist perspective like Davidson –, this consequence seems to undermine the form of realism Dretske otherwise advocates. Realists like Dretske should not take the frequency of error to be constrained by the very theory explaining what error is. Whether or not we are right in all, most or even some of our beliefs cannot be decided on philosophical grounds alone.[27]

The general diagnosis seems to me the following: all five problems with Dretske's theory are due to his construal of the space of possibilities, his realistic conception of what the relevant alternatives for a given signal are. His realism and the correlated insistence that information is an "objective commodity" (Dretske 1981: vii) has led him to misrepresent the relativity of information (as we noted in sct. 1.4.7). It seems that background knowledge can not only narrow down the space of possibilities to be considered at the source (and thus make, for some receiver, part of the information objectively encoded useless), but that (lack of) background knowledge can also expand that space of possibilities. We already noticed (on p. 50) such a case while discussing the importance of notation – not only for coding, as Dretske assumes, but also for the determination of the relevant possibilities.

It seems to be necessary, then, to allow a second and different rôle for the background knowledge k in Dretske's definition of informational content (2.2.3). Apart from narrowing down the realm of relevant possibilities (thus increasing the amount of information nested in the content), it can also widen this realm, even beyond the realm of real physical

[27] Another drawback of Dretske's line, as it has been correctly emphasised by Sterelny (1983: 210), is his commitment to the dubious claim that all empty (non-referring) concepts are complex.

possibility.[28] Dretske tries to save his definition of knowledge (2.2.5) from vulnerability to sceptical scenarios by the notion of a channel condition (2.2.6): he thereby excludes some abstract possibilities from being relevant possibilities.[29] There seems to be a need for a dual: not only antecedent knowledge making more things knowable, but antecedent beliefs making less things knowable (this would also offer a promising route to solve the problem mentioned in n. 24).[30]

Intuitively, when I am told that Fermat's Theorem is true, that George W. Bush is the present president of the US or that robins are birds, I learn some new piece of information, a piece of information I could have lacked even while possessing the relevant concepts. This gain in information is most straightforwardly described as the exclusion of possibilities which – while not real possibilities given the way the world really is – were possibilities *for me*. This notion of subjective possibilities is elusive, to be sure, but it deserves more consideration that Dretske was willing to give it.[31]

Construing possibilities as subjective in this sense brings them back into cognitive reach, thereby alleviating the fourth worry noted above. It also allows us to recognise a form of error different from mistaken reliance on non-informational connections, namely false belief that some (merely possible) alternative is relevant and/or still open. We will take up these matters again in ch. 4, after having discussed the second main inspiration of the theory of information flow to be discussed later, namely Jon Barwise's and John Perry's situation theory.

[28]It would, in any case, be a bad idea just to drop the reference to k and to go for "absolute informational content", as Savitt (1987: 188) suggests. Not only would this severely limit the applicability of Dretske's theory, it also would leave an important cognitive factor out of the picture. As Barwise and Seligman note, it would neither help with the general problem of necessary truths' being nested in any information whatsoever (Barwise and Seligman 1997: 17)

[29]This move has earned him the charge that he rules out scepticism *by fiat*, by "conflat[ing] the merely possible with the unlikely" (Sterelny 1983: 209) As it has since then become clear, however, he is ready to plead guilty to this charge (cf. Dretske 2001ba).

[30]Richard Foley Foley (1987: 182) seems to make a related point, though he suggests an additional, coherentist, requirement on the total set of the agents' belief (roughly, that they be not too irrational), while I would prefer inclusion only of the agent's belief *about* the source. Adopting this proposal would not, therefore, mean to forsake the whole of Dretske's realism, but just mean to extend it to the agent.

[31]All the more because it is not $I(s)$ nor $I_s(r)$ but conditional probability, and thus a prior event space, that is doing the work in the crucial definitions of informational content (2.2.3) and knowledge (2.2.5). It may even be misleading, as Foley (1987: 165) suggests, to describe Dretske's account of knowledge as "information-theoretic" at all. W. E. Morris has noted (Morris 1990: 382) that Dretske is not altogether consistent on this point. Cf. e.g. the following passage: "Whether a signal carries the information that s is F does depend, among other things, on what the speaker knows about the object s." (Dretske 1983: 84)

Chapter 3

Meaning: Situation Theory

3.1 General Outlook

Situation theory, as described in Barwise's *Scenes and Other Situations* (1981), grew out of the investigation of naked infinitive perceptual reports like "Sam saw Austin get shaved in Oxford" (Barwise 1981: 7). The semantics of such reports seemed to demand a departure from the classical Fregean doctrine that the reference of a sentence (and of a naked infinitive in particular) is a truth-value. Instead, the embedded infinitives seemed rather to refer to *situations*, i.e. more precisely to scenes, which are perceptually registered situations.[1] Naked infinitive reports differ from those embedding that-clauses in that they are "epistemically neutral" (their truth demands no command of concepts and no knowledge on the part of the perceiver), veridical and extensional. Most important of all, they seem to be in need of a 'logic' of their own, for they intuitively behave differently than those embedding that-clauses e.g. with respect to negation: While

$$a \text{ sees that } \neg\phi \iff \neg\,(a \text{ sees that } \phi)$$

seems plausible, the right-to-left direction of

$$a \text{ sees } \neg\phi \iff \neg\,(a \text{ sees } \phi)$$

[1]Devlin, otherwise relying heavily on Barwise's and Perry's 1983 theory, has surprisingly denied this: "... I am taking the visual content of any visual experience to be a *proposition*, and not a situation or an object." (Devlin 1991: 191) He goes on to distinguish perceiving from seeing and construes the latter as conceptual. He thereby severely limits the applicability of his analysis.

is out of the question. Even if it may be said that I see that Sam is not at home iff I do not see that he is at home (at least under a reading of the latter where it is understood that I carefully looked), it does not follow from my not seeing Sam being at home that I see him not being at home. It is not even clear what the latter means, for seeing a not doing F seems to presuppose some perceptual contact with a. But even if seeing Sam not being at home is understood as meaning to see Sam being somewhere else than at home, the implication does not hold. For I may carefully look at the house and fail to see him there, without seeing him anywhere else (I may have no clue where he is). It seems, therefore, that rather different principles are at work in these two examples.

This contrast may be compared with the one between epistemic and non-epistemic seeing, made famous by Dretske (1989): In the non-epistemic sense of "seeing", I may be said to see that Sam is at home iff I am visually presented with a scene (if I have the visual impression described by others as one) in which Sam is at home. I might, e.g., be looking out of my window and see my neighbour sitting in his kitchen. In the same sense in which I see Sam at home, my cat may be said to see Sam at home: both of us are looking in the same direction, the lighting is good and we have similar visual impressions and both our impressions are impressions *of* Sam (in Dretske's terminology, they are signals carrying the information that Sam is at home). In another, the epistemic, sense of "seeing", however, seeing that Sam is at home requires knowing who Sam is, being able to recognise Sam and to know, e.g., that he is a human being. In this sense, my cat does not see anything (except, perhaps, mice and cheese and me), for epistemic seeing presupposes (at least some) conceptualising abilities. This is the sense in which the policeman may show me (but not my cat) a picture and ask whether I have seen this person. While there is some controversy whether ordinary uses of "see" followed by that-clauses really are ambiguous between these two senses, it seems clear that naked infinitive reports are most appropriately used to report cases of non-epistemic seeing. They are assertions to the effect that some person was visually presented with a certain scene 'described', in a sense, by the naked infinitive clause.

To investigate such phenomena as naked infinitive perceptual reports, then, is to work on a different project than the ordinary logic of propositions, a project that demands new tools, new ideas and, ultimately, a new ontology. This was the first motivation behind Jon Barwise's and John Perry's "situation theory" (or "situation semantics", as it has been called in its early years).

Apart from that, the motivation for their project is to be found in two rather general considerations: one pertaining to the importance and fruitfulness of partial modelling in general and the focus on situations (local, agent-individuated parts of objective reality and their 'abstract' counterparts) as opposed to models (maximally specific ways the world might be), the other to what they call the "Relation Theory of Meaning" (Barwise and Perry 1983: 6): *Meaning*, and linguistic meaning in particular, rather than being reified as "proposition", is conceived of as a relation between situations of different types but sharing some configuration of uniformities. *Linguistic meaning* is a relation between utterance situations and the things the sentence uttered are about (described situations), which is determined by linguistic conventions (conventional constraints) (Barwise and Perry 1983: 17). Meaning, in short, is some type of information flow between situations:

> "One situation s can contain information about another situation s' only if there is a systematic relation M that holds between situations sharing some configuration of uniformities with s and situations that share some other configuration of uniformities with s'." (Barwise and Perry 1983: 14)

To make this necessary condition sufficient, Barwise and Perry propose a two-step model in the spirit of Kaplan's work on demonstratives (Kaplan 1989): The interpretation of an utterance u depends on *and* is determined by the meaning of u and the facts (facts about the utterance situation) and is the situation described by u (i.e. the situation u is about). The truth-value of u depends on *and* is determined by the interpretation of u and the facts (facts about the described situation): if the described situation is factual, u is true; if it is merely possible or impossible, u is false. This may be seen as a variant of the two-stage theory pioneered by Stalnaker:

> "... the syntactical and semantic rules for a language determine an interpreted sentence or clause; this, together with some features of the context of use of the sentence determines a proposition; this in turn, together with a possible world, determines a truth-value. An interpreted sentence, then, corresponds to a function from contexts into propositions, and a proposition is a function from possible worlds into truth-values." (Stalnaker 1971: 36)

So Stalnaker's "propositions" are Barwise's and Perry's "interpretations" – conceived of not as sets of possible worlds, but as (sets of) this-worldy situations.[2]

To motivate further their focus on situations, Barwise and Perry provide a list of six properties of natural languages standard model-theoretic semantics has difficulties to deal with:

[2]Though situation theory's situations were first introduced as a replacement for the propositions postulated by ordinary model-theoretic semantics, they were later re-introduced in the theoretical framework (see sct. 4.2).

the external significance of language: To account for the fact that sentences can convey information that does not belong to their meaning (*however* the latter notion is understood), we must have a language-independent way of representing the way the world is or is claimed to be. Situation theory provides such a way.

the productivity of language: The Principle of Compositionality (saying that some kind of feature F of the whole is a function of the F of its parts) holds for meaning (else productivity, our ability to form and understand sentences never heard before by putting them together out of known components, could not be accounted for), but not for interpretations. So they have to be distinguished, in the way situation theory does.

the efficiency of language: Expressions can be recycled to mean different things when used in different ways, at different locations, by different people. Expressions can have different interpretations, even though they retain the same linguistic meaning. The discourse situation (indexicality), the speaker's connections to objects, properties and locations (use of names, deictic use of pronouns, referential use of tense) or her resource situation (knowledge of facts) can all be exploited to counter-balance the underdetermination of interpretation by meaning and hence to achieve the efficiency of language. These factors have to be represented in an account of what is said.

the perspectival relativity of language: Each of these three forms of exploitation (of discourse situations, speaker's connections and resource situations) gives rise to its own form of perspectival relativity: the speaker has to adopt its utterance to the discourse situation, his referential connections and his resource situation to get others at the right interpretation of what he says, i.e. to describe the situation he wants to describe.

the ambiguity of language: Semantic and syntactic ambiguity – a rule and not an exception – should not be regimented away (as it is in the fiction of completely disambiguated propositions), but seen as just another aspect of the efficiency of language and calls for a classification taking into account how a particular sentence is used in a particular utterance.

the mental significance of language: Utterances convey information not only about the world or the described situation, but also about the speaker's state of mind. In normal cases, we learn what he believes and we therefrom infer what he fears, hopes etc. and predict what he will do. This mental significance of language has to be explained by its external significance (thereby excluding a recourse on, e.g., Fregean senses located in a disconnected third realm).

In their first two papers on situation semantics, *Situations and attitudes* (Barwise and Perry 1981b)and *Semantic innocence and uncompromising situations* (Barwise and Perry 1981a), both published in 1981, Jon Barwise and John Perry take it to be one of the principal virtues of situation semantics to preserve what they call, following Davidson (1968), *semantic innocence*, while at the same time being able to offer a clear diagnosis of the fallacy in the famous *slingshot argument*, mounted by Church (cf. Church (1943 1950)) against Carnap (Carnap 1942) and designed to show that the referent of a sentence cannot be but its truth-value.

A semantic theory is semantically innocent iff it attributes the same semantic values to sentences embedded into intensional contexts and to unembedded ones. A semantically innocent theory thus gives the same analysis of occurrences of "Mary is a dog" preceded by "Sam believes" than to such following "John is a horse and". Situation semantics interprets all such occurrences as referring to the situation of Mary's being a dog and is thus semantically innocent. To maintain semantic innocence in problematic 'Frege cases', where substitutivity of coreferential singular descriptions within so-called 'opaque contexts' (belief ascriptions, modal contexts and quotation), Barwise and Perry draw an (ultimately ontological) distinction between what they call value-loaded and value-free occurrences of definite descriptions (see below, p. 79). This also helps them to avoid the slingshot, for logically equivalent sentences do no longer have the same referent (Barwise and Perry 1983: 24-25).[3]

In the following, I will first sketch and discuss the pertinent feature of the original theory expounded in *Situation and Attitudes* Barwise and Perry (1983), with the aim both to give the flavour of the ambitious project and to identify critical assumptions operative in much of the subsequent work and influential on the theory of information flow presented in Barwise's and Seligman's *Information Flow* Barwise and Seligman (1997), to be discussed at length in ch. 5 to 8. A final section will then try to give a preliminary assessment (sct. 3.3).

[3]As Dagfinn Føollesdal has pointed out Føllesdal (1983), this will not help against a closely related argument Quine gives in (Quine 1960: 197–198). To counter this, Barwise and Perry would have to adopt a rigid designator conception of proper names as expounded by Kripke (1972).

3.2 The 1983 theory

3.2.1 Situations and courses of events

In building their situation semantics for natural languages, Barwise and Perry take the notion of a situation as a primitive. Individuals, properties, relations, and space-time locations are uniformities across situations and themselves taken as primitive. They are the referents (interpretations) of nouns and predicates respectively, while the interpretation of a statement is the situation described. *Situation-types* are partial functions from n-ary relations r and n individuals $a_1, \ldots a_n$ to truth-values, i.e. relations between a constituent sequence $\langle r, a_1, \ldots a_n \rangle$ and 0 or 1 (standing for truth-values). A *course of events* is a partial function from locations to situation-types, i.e. a set of triples consisting of a location, a constituent sequence and a truth-value. A *state of affairs* is a course of events defined on just one location, i.e. a pair of a location and a situation-type.

Situations are either real or abstract. Abstract situations are actual, factual or (merely) possible (non-factual). The *real* situations are what the world (the real world, which is the only world in the existence of which Barwise and Perry believe) consists of (Barwise and Perry 1983: 50). Abstract situations are just singleton subsets of courses of events or states of affairs. Such a state of affairs is *factual* iff it correctly classifies a real situation, i.e. iff, whenever the state of affairs represents a location t as being such that $a_1, \ldots a_n$ stand there in the relation r, they really stand in that relation, and, whenever it represents them as not standing in such a relation, they really do not stand in that relation there. The state of affairs is (not only factual but) *actual* iff the biconditionals hold, i.e. iff it not only correctly but exhaustively classifies what is going on in a real situation.

More formally, a *structure $\mathcal{M}$ of situations* is a collection (not necessarily a set) M of courses of events with a non-empty subcollection M_0 (the actual courses of events) satisfying the following conditions:

- Every $e \in M_0$ is coherent.[4]
- Parts of courses of events in M_0 are also in M_0.[5]

[4]A course of events is coherent iff it assigns only coherent situation-types to locations. A situation-type is coherent iff it neither assigns two values to a constituent sequence, nor represents two different objects as being the same nor does represent anything as being different from itself.

[5]A course of events e_1 (a partial function from locations to situation-types) is a part of a course of events e_2 iff every member of e_0 (every state of affairs in e_0) is a part of e_1, i.e. iff has the same location as some such part e_2 and assigns to that location a situation-type which is a subset of the situation-type of e_2.

- For every subset of M there is a course of events in M_0 which has all the members of this subset as parts.
- M respects all constraints in M (cf. sct. 3.2.2).

Courses of events are the basic building blocks of Barwise's and Perry's semantic theory. To get a grip on enough of the uniformities holding between them, they introduce (basic) indeterminates, abstract objects without interfering structure, standing for individuals $(\dot{a}, \dot{b}, \dots)$, relations $(\dot{r}, \dot{s}, \dots)$ and locations $(\dot{l}, \dot{l}', \dots)$. An *event-type* then is a set of pairs of locations and situation-types, in which individual, relation and location indeterminates are allowed (a course of events is just an event-type without indeterminates). An *anchor* for an event-type E is a function assigning individuals, relations or locations to some of the indeterminates of E, and thus defining a new event-type $E[f]$. Such an anchor is *total* iff $E[f]$ is a course of events. A course of events e is *is of type* E if there is an (necessarily total) anchor f such that $E[f]$ is part of e. Partial anchors give us object types $E(\dot{a})$ (event-types with exactly one individual indeterminate) and types of complex properties $E(\dot{a}, \dot{l})$.

Complex indeterminates (i.e. *roles*) are given by the following recursive definition of indeterminates: Every basic indeterminate is an indeterminate; if $\dot{x}$ is an indeterminate and $E(\dot{x}, \dots)$ is an event-type, then $\langle \dot{x}, E \rangle$ (abbreviated by $\dot{x}_E$) is an indeterminate. Indeterminates of the latter sort are roles. A partial function f from indeterminates to individuals, relations and locations then is an anchor iff it is an anchor for the basic indeterminates and, if $\langle \dot{x}, E \rangle$ is in its domain, an anchor for each indeterminate in E and such that $f(\langle \dot{x}, E \rangle) = f(\dot{x})$. An object x *has the role* $\dot{r} = \langle \dot{x}, E \rangle$ *in a course of events* e iff e is of type E and every total anchor f for E such that $E[f]$ is part of e maps $\dot{x}$ to x. A course of events e is a *context for an event-type* $E(\dot{r_1}, \dots, \dot{r_n})$ iff there is an anchor for E in e with which every other anchor agrees on $\dot{r_1}, \dots, \dot{r_n}$ in e.

A *schema* S is a set of event-types. If f is an anchor, then $S[f]$ is the set of all event-types $E[f]$ for $E \in S$. An event e is of type S iff it is of type $E[f]$ for some $E \in S$.

3.2.2 Constraints

Chapter 5 of *Situations and Attitudes* discusses constraints, the basic regularities to be found in the world which make it possible that some situations carry information about other situations, with which they are linked. The theory of constraints is both the most important and the least developed part of the 1983 theory and it is a safe guess that further reflection on constraints led to most later developments of the theory. A late descendent of the constraints of *Situations and Attitudes* may be seen in the notion of an information channel in Barwise's and Seligman's theory of information flow, which will be discussed in great detail later (cf. ch. 5, in particular sct. 5.2.2).

Constraints are states of affairs relating event-types. Constraints may be classified with respect to their sources. Some constraints arise from the metatheory, in *Situations and Attitudes* Kripke-Platek set theory with urelements (ruling out, e.g., courses of events containing themselves). Others are of a metaphysical nature, e.g., the principle of non-contradiction (incorporated as fourth condition into the definition of a situation structure), ruling out incoherent factual courses of events. Necessary constraints arise from relations between properties and relations (every woman is a human) and mathematical and logical laws. Nomic constraints arise from natural laws and account for the type of information flow Grice (Grice 1948) has called "natural meaning" (as when we say that smoke means fire or that the footprints in the snow mean that the cat has come home). Conventional constraints arise from explicit or implicit conventions within a community. Such constraints account for our knowledge of language and the transmission of information by linguistic or, in general, symbolic means.

Unconditional constraints may be classified by a primitive relation of *involving* between types of events. If every event of type E_1 is of type E_2, the following state of affairs is factual:

$$\mathsf{at}\ l_u : \mathsf{involves}, E_1, E_2 : \mathsf{yes}$$

where l_u stands for the universal location and yes for truth. A *simple constraint* is a state of affairs of the form

$$\mathsf{at}\ l_u : \mathsf{involves}, E, S : \mathsf{yes}$$

where S stands for a schema. A *constraint* is a course of events which has a simple constraint as a part. If C is a constraint, a course of events e is said to be *meaningful with respect to* C ($e \in \mathsf{MF}_C$) if e is of type E. If e is meaningful with respect to C, then e' *is a meaningful option from* e *with respect to* C ($e\mathsf{MO}_C e'$) iff for every total anchor f for exactly the indeterminates in E, if e is of type $E[f]$, then e' is of type $S[f]$.

An course of events e which is meaningful with respect to C *precludes an event* e' *with respect to* C (e precludes$_C$ e') if e' is incompatible with e or if there is an anchor f for the indeterminates in E such that e is of type $E[f]$ but for every extension of f to an anchor g for the indeterminates in S, e' conflicts with every course of events of type $S[g]$.[6]

A structure of situations $\mathcal{M}$ *respects a simple constraint* C iff for every actual e_0 in $\mathcal{M}$ the following holds: if e_0 is meaningful with respect to C then some factual situation e_1 of $\mathcal{M}$ is a meaningful option from e_0 with respect to C. $\mathcal{M}$ *respects a constraint* C if it respects each simple constraint which is part of C. The *interpretation of an event* e_0 *with respect to a constraint* C is the set $[\![e_0]\!]_C := \{e_1 \mid e_0 MO_C e_1$ and e_0 does not preclude$_C$ $e_1\}$, i.e. the set those of courses of events which are open possibilities if e_0 and C are factual.

3.2.3 Situation semantics

One of the principal aims of Barwise and Perry in *Situations and Attitudes* is to provide us with (the beginnings of) a semantic theory for natural languages. The attribution of meanings to sentences, in their view, is just a special case of the account of informational regularities arising from constraints:

> "The linguistic meaning of expressions in a language are conventional constraints on utterances. To study semantics is to attempt to spell our these constraints, to spell out what it is the native speaker knows in knowing what utterances of his language mean." (Barwise and Perry 1983: 119)

The *meaning* of an indicative sentence ϕ is a relation $u[\![\phi]\!]e$ between utterance situations u and described situations e which constrains both u and e. The constraints of $[\![\phi]\!]$ on e are usually studied in model-theoretic semantics, but those on u (e.g. that one is talking about someone in particular, etc.) are considered equally important in situation semantics.

From the generic event-type

$$\begin{aligned} \mathsf{D} := \quad \mathsf{at}\ \dot{l} : \quad & \mathsf{speaking}, \dot{a}; \mathsf{yes}; \\ & \mathsf{addressing}, \dot{a}, \dot{b}; \mathsf{yes}; \\ & \mathsf{saying}, \dot{a}, \dot{\alpha}; \mathsf{yes}; \end{aligned}$$

we get linguistic roles of (i) the speaker ($\langle \dot{a}, \mathsf{D} \rangle$), (ii) the addressee ($\langle \dot{b}, \mathsf{D} \rangle$), (iii) the discourse location ($\langle \dot{l}, \mathsf{D} \rangle$) and (iv) the uttered expression ($\langle \dot{\alpha}, \mathsf{D} \rangle$). A *discourse situation* d is a situation where these roles are uniquely filled (speaker, addressee, discourse location and expression uttered are then denoted by "a_d", "b_d", "l_d" and "α_d" respectively). A discourse

[6] Two courses of events conflict (or are incompatible) with each other iff they assign to one location situation-types which do only differ in their truth-values.

situation of the above type augmented by

$$\mathsf{referring-to}, \dot{a}, \alpha, \dot{a}_1; \mathsf{yes};$$

is a *referring situation* c and we write $c(\alpha)$ for the referent of α in c. The *speaker's connections* c in an utterance is the partial function from expressions α to their referents $c(\alpha)$. Utterances in which names do not refer or do refer to two things are meaningful, but do not describe any actual situation. As c and d are the most influential factors in the utterance situation, but we need, in addition, a (linguistic) setting σ provided by other parts of the utterance (e.g. to determine the referents of anaphora), we write $d, c[\![\alpha]\!]\sigma, e$ for the meaning of an expression α in an assertive utterance.

So meanings of linguistic expressions are, in general, four-place relations between discourse situations, referential connections, linguistic settings and described situations. They are constraints: given a discourse situation with an utterance, the connections and the setting, only some of all the possible situations are candidates for the situation to be described by the utterance.

This general framework now allows Barwise and Perry to state the meanings of many English verb and noun phrases. The *meaning of a tensed VP*, e.g., is a relation between discourse situations, connections, individuals and courses of events:

$$\begin{aligned} d, c[\![\text{“is biting Molly”}]\!]a, e \iff & \ \mathsf{in}\ e : \mathsf{at}\ l : \mathsf{biting}, a, \mathsf{Molly}; \mathsf{yes} \\ & \text{where } \mathsf{Molly} = c(\text{"Molly"}), l = c(\text{"is"}) \text{ such that } l \circ l_d{}^7 \end{aligned}$$

The *meaning of a tenseless VP* is a relation between discourse situations, connections, settings (providing for an individual, a location and a truth-value) and courses of events which holds under the following condition:

$$\begin{aligned} d, c[\![\text{“biting Molly”}]\!]\sigma, e \iff & \ \mathsf{in}\ e : \mathsf{at}\ l_\sigma : \mathsf{biting}, a_\sigma, \mathsf{Molly}; tv_\sigma{}^8 \\ & \text{where } \mathsf{Molly} = c(\text{"Molly"}) \end{aligned}$$

The *meaning of a tenseless sentence* is a relation between discourse situations, connections,

[7] a here replaces σ because the only rôle of the setting is to provide an individual biting Molly. “in e” means that the following event-type has to be a *part* of the described situation e. “$l \circ l_d$” means that the location l referred to by the tensed verb “is” is overlapping the location of utterance l_d.

[8] “tv_σ” is the truth-value the statement has as being evaluated evaluated in setting σ, that is with respect to the individual a_σ.

its setting in some larger utterance and described situations:

$$d, c[\![\text{“Jackie biting Molly”}]\!] l_\sigma, e \iff \textsf{in } e : \textsf{at } l_\sigma : \textsf{biting, Jackie, Molly; yes}$$
$$\text{where } \textsf{Molly} = c(\text{”Molly”}), \textsf{Jackie} = c(\text{”Jackie”})$$

One of the advantages of this framework is that it allows for a straightforward representation of the meaning of indexicals. The *meaning of "I"*, e.g., is a relation between discourse situations, connections, the speaker and a described situation, which holds iff the speaker determined by the setting is the speaker of the utterance:

$$d, c[\![\text{“I”}]\!] a_\sigma, e \iff a_\sigma = a_d$$

Analogously, the *meaning of "you"* is a relation between discourse situations, connections, the addressee and a described situation:

$$d, c[\![\text{“you”}]\!] a_\sigma, e \iff a_\sigma = b_d$$

The *meaning of a name* is a relation between discourse situations, connections, an individual and a described situation:

$$d, c[\![\text{“}\beta\text{”}]\!] a_\sigma, e \iff c(\text{“}\beta\text{”}) = a_\sigma$$

The *meaning of a third-person pronoun like "it"* is a relation between discourse situations, connections, an individual and a described situation:

$$d, c[\![\text{“it}_1\text{”}]\!] a_\sigma, e \iff c(\text{“it}_1\text{”}) = a_\sigma$$

For the *meanings of definite descriptions*, we have to make what has been called *Donnellan's distinction*, i.e. distinguish between their referential and attributive use.[9] A referentially

[9] This is the distinction between two uses or readings of definite descriptions made by Keith Donnellan in a series of papers, beginning with *Reference and Definite Descriptions* Donnellan (1966) and continuing with (1968) and (1979). It is may be introduced with something like the following thought experiment: I am with you at a party, watching a man at the bar holding a glass with a transparent liquid in his hands and talking to the waiter. If I tell you that *the man drinking Martini over there* is my daughter's new boyfriend, it is clear that you may understand me and get the meaning I intended to convey *even if the man is actually drinking water.* It does not seem to matter whether there is, unnoticed by the two of us, another man standing roughly in the corner and actually drinking Martini. The description, in these cases, is just a means to get you to pick out the individual I intend to speak about. It does not have to be true to achieve that purpose. This is what Donnellan calls the "referential use" of the definite description

used definite descriptions is used to pick out a specific individual, which is believed to satisfy the description. In their attributive use, on the other hand, definite descriptions serve to talk about whatever individual happens to satisfy a given criterion. In general, the meaning of a definite description is a relation between discourse situations, connections, situations and settings:

$$d, c[\![\text{“the } \pi\text{”}]\!]a_\sigma, e \iff d, c[\![\text{“}\pi\text{”}]\!]a_\sigma, e$$
$$\text{and there is at most one } b \text{ such that } d, c[\![\text{“}\pi\text{”}]\!]b_\sigma, e$$

The difference between referential and attributive uses is that the latter are just relations or partial functions from courses of events (resource situations) to individuals (those picked out by the descriptions), while the former are specific values of the function in question when applied to some particular resource situation. The same distinction applies in the case of the *meaning of singular NP*, with the sole difference that their meaning will be a relation and not, in general, a function.

The *interpretation* $[\![\phi]\!]$ of a statement (an utterance of an indicative sentence) ϕ is the collection of events e such that $d, c[\![\phi]\!]e$.[10] A statement is *true* if there is a factual situation e in its interpretation. A statement ψ is a *strong consequence* of a statement ϕ ($\phi \vdash \psi$) iff $[\![\phi]\!]$ is a subcollection of $[\![\psi]\!]$. If two statements have the same interpretation, they are called strongly equivalent. A statement ψ is a *weak consequence* of another statement ϕ iff ψ is true in any situation structure in which ϕ is true. If two statements are true in the same situation structures, they are called weakly equivalent.[11] Every statement is a weak consequence of statements not true in any situation structure (e.g. statements violating factual constraints), though for example self-contradictory utterances, describing certain impossible situations, do not have any non-tautological strong consequences whatsoever.

Sentences like "x means that p" describe constraints: they are true iff the described constraint is factual. In sentences like "Jackie's biting Molly (always) means that Jackie is scared" we have to distinguish between *event-type meaning* and *event meaning*.[12] The

"the man over there drinking Martini". In its attributive use, I use it to speak about whatever individual it is true of, as in "the man drinking Martini would better take a taxi to get home".

[10]Against this, Sten Lindström argued that the semantic value of a statement should be identified with the ordered pair of its extension (situations of which it is true) and its anti-extension (situations where it is false), allowing, e.g., for a distinction between strong and weak negation (Lindström 1991: 761).

[11]Strong consequences of a given statement are also weak consequences of it, though the converse is not true. The distinction is modelled on a corresponding one in model-theoretic semantic, where weak consequence is transferal of truth-in-a-model (validity), while strong consequence is transferal of truth-in-a-world.

[12]This distinction is related, though not exactly identical to Devlin's less clear-cut distinction between

event-type meaning of the former sentence is the following constraint:

$$\text{at } l_u : \text{involves}, E(l), E'(l); \text{yes}$$

where the event-types involved are the following two:

$$\begin{aligned} E &:= \text{at } l : \text{biting, Jackie, Molly; yes} \\ E' &:= \text{at } l : \text{scared, Jackie; yes} \end{aligned}$$

Event meaning, on the other hand, e.g. the meaning of "that means that ϕ" is a relation which holds if the event referred to by "that" is part of the described situation and every meaningful option from that event is such that ϕ is true of it. In other words, a situation s has the event meaning that there is a situation of type S' iff there is a factual constraint

$$\text{at } l_u : \text{involves}, S, S'; \text{yes}$$

such that s is of type S.

In subsequent work (e.g. Barwise 1986c), Barwise rephrased this distinction as the one between situation meaning and situation-type meaning. The first is of the order of Gricean natural meaning, the latter is a constraint (in the case of human languages a conventional constraint) which we can become attuned to:

> "Attunement to meaning$_t$ [situation-type meaning] is what allows an agent with information about the first [situation] to soundly infer what it means$_s$ [its situation meaning]." (Barwise 1986c: 51)

A *constituent* of an interpretation is an object, property, relation or location that is a constituent of every course of events in the interpretation of a statement. Donnellan's distinction can be rephrased thus: when descriptions are used attributively, the describing condition is a constituent of the interpretation; if they are used referentially, the described individual itself is such a constituent. In the latter, but not the former case, the individual is *loaded into* the interpretation. Definite descriptions are interpreted as exploiting a resource situation to describe another situation. In this respect, definite descriptions are similar to indexicals. They are different, however, in that they reach further (they do not even presuppose perceptual accessibility of the referents) and one has more freedom in the choice of the resource situation to exploit.

"abstract meaning" and "meaning-in-use" (Devlin 1991: 222).

Barwise and Perry make a similar distinction in the case of proper names. According to the Russell-Frege-Searle theory of names, some describing conditions are constituents of the interpretation of the statement in which the name occurs. According to the Donnellan-Kripke theory, the individual named is the only constituent. Barwise and Perry adopt the latter, with the amendment that names need not be unique and can also serve as common nouns (as when I say that there are more than one Peter Smith in the room). Because individuals could have been named otherwise, the property of being a Jackie or being named "Jackie" is not part of what I assert by a statement about Jackie. Nevertheless, hearers get the information that I think that what I call "Jackie" is named "Jackie". The information they get that way is called *inverse information.*[13] To account for the fact that it is part of the discourse situation that the speaker has the information that the individual to whom he refers has the name the speaker uses, we need to amend our definition of the *meaning of names* (inf means the relation of having information that):

$$
\begin{aligned}
d, c[\![\text{``}\beta\text{''}]\!]a_\sigma, e \iff\; & c(\text{``}\beta\text{''}) = a_\sigma \\
& \text{in } d\text{: at } l_d\text{: } \mathsf{inf}, a_d, E; yes; of, \dot{b}, a_\sigma; \text{yes} \\
& \text{where } E{:=} \text{ at } l\text{, being-}a\text{-}\beta\text{, } \dot{b}\text{; yes}
\end{aligned}
$$

3.2.4 Epistemic attitudes

Barwise's and Perry's theory of epistemic attitudes like seeing, believing, thinking and knowing is guided by the desire to preserve what has been, after Davidson (1968), been called "semantic innocence", i.e. the belief that sentence tokens embedded in such "intensional" constructions like believing, knowing or saying mean the same as their unembedded tokens (Barwise and Perry 1983: 174). They take as their paradigmatic epistemic attitude a kind of seeing, "seeing$_n$", reports of which are epistemically neutral and whose grammatical objects are gerunds or naked infinitives (as in "I saw Jon walking") as opposed to "seeing$_t$" whose reports are epistemically positive and whose grammatical objects are that-clauses (as in "I saw that John walked"). Reports of the first kind of seeing, but not of the second, preserve their truth-value under substitution of coreferential terms (and are thus extensional). Both are veridical.[14]

The main difference between "sees$_n$" and "sees$_t$", in Barwise and Perry's view, is that

[13] This is what Perry now prefers to call "reflexive content" (cf. sct. 1.2.3).

[14] This means that the following constraint is factual: at l_u : involves, $E(l), E'(l)$; yes where the event-types involved are $E :=$ at $\dot{l}$: seeing, a, e; yes and $E' :=$ at $\dot{l}$: actual, e; yes.

the former describes just the relation between a perceiver and one scene (visible situation), while the latter describes a relation between a perceiver and a whole range of situations, i.e. his *visual alternatives.* To give the semantics of attitude reports involving "sees that", they start with the following observation: while visual information at a time may provide us clues about a variety of situations and exclude certain others, it remains mute about most of all the possible situations. So while some situations are classifiable as compatible with what an agent a sees or knows at a given time and some as incompatible with, i.e. ruled out by, what he sees or knows, many can be classified neither in this nor the other way. We thus have, for any given situation, a tripartite distinction. Some situations are compatible with what a knows in s (hence classified as visual options SO for s), some are incompatible with what a knows in s (and thus classified as visual non-options for s), and some are neither. If the first is the case, we say that the agent's situation s includes

$$\mathsf{in}\ s : \mathsf{at}\ l : \mathsf{SO}, a, e; \mathsf{yes}$$

and we call e a *visual option for a in s.* In the second case, we have no in place of yes. If the latter is *not* the case, i.e. if e is *not* ruled out by a's perceptual evidence (not classified as non-option in the situation in which a is), that is if we do *not* have

$$\mathsf{in}\ s : \mathsf{at}\ l : \mathsf{SO}, a, e; \mathsf{no}$$

we call e a *visual alternative for a in s.* Visual alternatives for a in s are thus all those situations which are not classified as non-options, i.e. the visual options and those situations which are neither visual options nor non-options.

To determine whether a sees that ϕ, it is not enough to know that ϕ holds in all of a's visual options, because a might not have bothered to classify all the relevant options. The only way we can tell for sure is to require that ϕ holds in all of a's visual alternatives:

$$\begin{array}{lcl} d, c[\![\text{“ sees that } \phi\text{”}]\!]a, e & \Longleftrightarrow & \text{for every } e' \text{ either} \\ & & \qquad d, c[\![\text{“ } \phi\text{”}]\!]e' \\ & & \text{or} \\ & & \qquad \mathsf{in}\ s : \mathsf{at}\ l = c(\text{“sees”}) : \mathsf{SO}, a, e'; \mathsf{no} \end{array}$$

So the meaning of "seeing that ϕ" is a relation between discourse situations and described situations which holds iff ϕ holds in all visual alternatives of the described situation, i.e. in all those situations not ruled out by a's perceptual evidence (not classified as non-options).

Barwise and Perry then give a dual account of "not seeing that":

$$
\begin{aligned}
d,c[\![\text{"does not see that } \phi\text{"}]\!]a,e \iff\ & \text{there is an } e' \text{ such that} \\
& \qquad \text{it is not the case that: } \ d,c[\![\text{" } \phi\text{"}]\!]e' \\
& \text{and} \\
& \qquad \mathsf{in}\ s : \mathsf{at}\ l = c(\text{"sees"}) : \mathsf{SO}, a, e'; \mathsf{yes}
\end{aligned}
$$

So I do not see that ϕ iff there is a visual option from my present situation where ϕ does not hold. Seeing and not seeing are thus not construed as contradictories, but only as contraries. In many cases I neither see nor do I not see that ϕ (namely in those cases where ϕ fails in some of my visual alternatives, but only in visual alternatives which are not visual options for me), but it is never the case that I both see and do not see that ϕ.

In a first step towards an adequate analysis of the epistemic attitudes, Barwise and Perry tentatively extend this analysis to all other epistemic attitudes, in every case identifying the relevant correlates of visual options and alternatives.

The meanings of "knows that" and "does not know that" are analysed in a completely parallel way, subject to a veridicality constraint (as explained in fn. 14 above) and to the following additional constraint connecting perception and knowledge: any situation that is precluded by what you see, on the basis of what you know, is precluded by what you know. a thus knows that ϕ if ϕ holds in all epistemic options, i.e. all situations not incompatible with what a knows. a does not know that ϕ if ϕ fails in some epistemic alternative.

This is an internalist picture of knowledge: in order to truly ascribe lack of knowledge concerning ϕ to some agent a, ϕ has to fail in some scenario that a has epistemic access to, some way the world could be that is a "live option" for the epistemic agent a. It is not enough, in particular, that ϕ is false in the actual world, which is only guaranteed to be an epistemic alternative (by the veridicality constraint), but does not have to be an epistemic option. This certainly reflects a valid intuition: if ϕ is a sentence a could not even formulate, involving concepts a does not possess and uniquely concerning things a has no beliefs about, it certainly seems weird to ascribe lack of knowledge that ϕ to a. The requirement, however, may seem too strong (and the corresponding requirement for knowledge too weak): not every failure to know is tantamount to some active epistemic access to a possible situation of which the claim in question does not hold. Sometimes we are just at a loss and have no clue as to why and under what circumstances a putative item of knowledge may be false: but we may still realise that this our being in such a situation precludes an entitlement to knowledge, for knowledge requires some active and positively qualified epistemic stance to what is known, if not actually an ability to demonstrate or make plausible the truth of what is known, so at least the capacity to access "in principle" such an epistemically authoritative stance.

Barwise's and Perry's putative reply to such worries is the following: an epistemic alternative, a situation not ruled out by what is known, does not have to be an epistemic option, i.e. a situation compatible with what is known. This, however, just pushes the distinction back to the original classification – made, if subconsciously, by the agent, we are to suppose – into epistemic options and non-options. The layman reading a scientific measurement device may not thereby gain knowledge, even if all $\neg\phi$ situations are actually incompatible with what he sees (and thus, by the bridging constraint, with what he knows). To withhold ascription of knowledge, we have to say that he fails to *classify* at least one such situation as non-option (i.e. that one such situation is an epistemic alternative for him). This, however, is just to say that he fails to *read* or *understand* what he sees, which is to say that he does not, by perception alone, gain (the relevant kind of) knowledge. In order to ascribe him lack of knowledge, which we certainly should,[15] we have to say not only that he fails to classify such situations correctly, but that he actually *misclassifies* them, i.e. treats them as epistemic options when they are actually non-options. This puts too heavy a weight on him: how could he misclassify such situations if he has no clue about their existence? how could he misclassify them if he even lacks the concepts to describe their crucial features? These were exactly the considerations motivating partial modelling in the first case.[16]

We will return to these questions later (cf. sct. 4.8.2). Let us proceed with Barwise's and Perry's account of the other epistemic attitudes. *Belief*, according to them, can either be viewed *as failed knowledge* or as a *probabilistic strategy*. In the first case, one would define *doxastic alternatives* in terms of a relation BO, a subset of the epistemic alternatives but not necessarily conforming to the veridicality constraint. In the second case, where knowledge is modelled as successful belief, one would again define doxastic alternatives (and BO) and then assign weights to them and not require that ϕ does hold in all alternatives, but require that the combined weight exceeded some (normally agent-relative) threshold. At least in this second perspective, belief in ϕ and ψ does not entail belief in $\phi \wedge \psi$. *Doubting*

[15]This may perhaps be doubted. It seems to me crucial, however, to be able to describe the progress the layman makes by learning to read the instrument as acquisition of a new means to get knowledge, i.e. as turning non-knowledge about the measured processes into knowledge.

[16]It is instructive to observe that these worries have their exact parallel in the case of Dretske's, very similar, definition of knowledge (2.2.5): my belief that p may well be caused by a signal carrying the information that p – as long as I have no clue about this latter fact and the causal antecedents of my belief, it certainly seems weird to ascribe knowledge to me.

that ϕ, understood as lack of belief in ϕ (in contrast to actual disbelief), is a relation which holds between utterance situations and situations with at least one doxastic option where ϕ is not true. So a does not doubt that ϕ iff ϕ holds in all of a's doxastic options, while a merely believes that ϕ iff ϕ holds in all of a's doxastic *alternatives*. Though Barwise and Perry consider this approach promising (Barwise and Perry 1983: 177), they do not work it out but instead sketch another theory which should overcome the main deficiencies of the first shot.[17]

The second approach, aiming at 'representing the mental', indirectly classifies mental states by the situations they are about or represent. To get a grasp on the uniformities among the contents of different mental states (which are crucial for folk psychological explanation), Barwise and Perry propose to classify mental states by indexed event-types, i.e. event-types in which abstraction is made over certain roles, the so-called *frames of mind*:

> "A person has an attitude (believes, knows, doubts, sees that, imagines, etc.) by being in a certain frame of mind, with various ideas and concepts anchored to the world in various ways." (Barwise and Perry 1985: lxxv)

So perceptions, e.g., are classified by the event-types of the perceived situation e: If and only if an agent a is perceptually aware of an event e_1, E is an event-type such that f is a total anchor for E in the situation e_0, in which the perception takes place, and $E[f]$ is part of e_1 we have a special kind of event, a *(represented) perception* e_0 of e_1, i.e. a situation with at least the following two features:

in e_0 : at l :	S_r, a, E; yes		frame of mind, perceptual condition
	of, $\dot{x}, f(\dot{x})$; yes	for all $\dot{x}$ in E	setting of the R-perception

The indeterminates in E which have to be anchored in e_0 are the roles of the perceiver, of his or her location and special roles called *images* (or *ideas* in the case of the attitude of believing to be discussed later), anchored to suitable (perceived) entities in the environment of the perceiver.

[17]According to Barwise and Perry (1983: 220f.), these are the following: (i) As we cannot explain the difference between you and me knowing that I am hungry we cannot explain why mental states are of use to predict behaviour. (ii) Mental states and activities play no role in our account. (iii) Constraints about some or all courses of events of a certain sort cannot be stated in the 'involve'-form. (iv) Because there is a proper class of events in which ϕ does not hold, the course of events required to classify a visual state (relying on all visual alternatives not ruled out) is a proper class. It can thus be no constituent of another situation and iterated attitude reports are impossible. Additionally, actual doxastic alternatives have to be situations where what is believed is true, because otherwise e would be not a doxastic alternative to itself and so a constituent of itself, which violates the axiom of foundation.

The relation S_r between a perceiver a and a type E of the perceived course of events "is one the observer brings to the situation when she characterises the agent a as having a certain type of perception" (Barwise and Perry 1983: 234). So here we have a case of indirect, so to say external, classification of the situation e_0 of a's perceiving e_1: the perception is classified in terms of what is perceived, i.e. in terms of the event-type E which, as anchored in the perceiving situation, is part of the situation perceived. This – and the corresponding new level of analysis, the 'frames of mind' – gives us the required uniformities across (perceiving) individuals captured by the constraints and patterns of folk psychology which operate at the level of frames of mind – but at a price. The price to pay, it seems, is that perception is defined as being veridical (in accordance to fn. 14 above).

This price is somehow mitigated by two distinctions, one between event and event-type meaning, the other between conditional and unconditional constraints. Perceptual frames of mind have event-type meaning: this much is ensured by the account above. For actual perceptual events to have event meaning, however, they have to be normal. The veridicality constrained is thus formulated conditionally: *in normal conditions* (barring perceptual illusions, hallucinations, drugs etc.), what we perceive is true. *Perceptual constancies* are then explained by the fact that, in normal circumstances, visually linked images are of the same thing. The evident problem, to which we shall return (in sct. 3.3 and 4.8.3) is how to non question-beggingly characterise normality conditions.

The generalisation to belief is then again straightforward: *Beliefs* are taken to be dispositional states, i.e. states known through their effects, and modelled by complex event-types, the components of which are called *ideas.* Beliefs are factored into efficient doxastic conditions and settings and classified with a relation B_r (represented belief) between an agent a and a belief schema S (a set of indexed event-types). A (represented) belief S of a is a situation e_0 which has constituents at various locations l which are of the following form:

in e_0 : at l :	B_r, a, S; yes		the doxastic condition
	of, $\dot{x}, b$; yes	for some ideas $\dot{x}$ in S	the setting

On this account, belief conditions can be thought of as strategies, suitable to some environments (partly specified by the embedded statement in the belief report) while unsuitable to others. They are suitable in environments in which their ideas get suitably anchored, but may fail to be about real things in other environments (due to the lack even of a conditional veridicality constraints, these environments are not therefore classified as abnormal).[18]

If $\dot{o}$ is an indeterminate in a belief schema S (an idea), $\langle \dot{o}, S(\dot{o}, \dots) \rangle$ is called a *concept*. A concept $\langle \dot{o}, S(\dot{o}, \dots) \rangle$ is *applied* to an individual b in e if the agent has the following belief (stands in the belief relation to the following event-type):

$$\text{at } \dot{h} : \text{same}, \dot{o}, \dot{t}; \text{yes}$$

where $\dot{t}$ is a visual image in e that is anchored to b. Using a's concept of b, the perceptual mode of recognition of b used by a can be modelled by a constraint-type. Although failure to apply concepts one possesses accounts for cases of the 'essential indexical' and for the non-substitutivity of coreferential descriptions,[19] it seems too strong a condition, in that it leads to the ascription of too many "sameness" beliefs to ordinary agents: we very rarely are aware of explicit identifications. Even if these beliefs are supposed to be held in some implicit, normally subconscious way, they still require a cognitive command of the abstract (and complicated) relation of sameness. Again, the problem is with cognitive command: knowledge is ascribed to agents which may not meet the most elementary preconditions even for conceptualising the issue at hand.

Barwise and Perry sketch two ways in which this picture of belief might be brought to bear on the semantics of belief ascriptions. A *first approach* would be to redefine the relation of being a doxastic option (BO) in terms of the new relation B_r:

$$\begin{array}{lcl} \text{in } e_0 : \text{at } l : \text{BO}, a, e; \text{yes} & \Leftrightarrow & \text{for every } S \text{ and } f \text{ such that:} \\ & & \qquad \text{in } e_0 : \text{at } l : B_r, a, E; \text{yes} \\ & & \qquad \text{of}, \dot{x}, f(\dot{x}); \text{yes} \\ & & \text{there is a } g \text{ extending } f \text{ so that } e \text{ is of type } S[a, l, g] \end{array}$$

Otherwise, i.e. whenever there are such S and f not extended by any such g, we have:

$$\text{in } e_0 : \text{at } l : \text{BO}, a, e; \text{no}$$

[18] This is somehow imprecise, for ungrounded beliefs (beliefs containing notions which are not anchored in any individuals, say about a particular unicorn, may still have some efficacy (cf. Devlin 1991: 178).

[19] Every one of my beliefs S contains information about me and my location, because $\langle \dot{i}, S \rangle$ is a concept of me and $\langle \dot{h}, S \rangle$ is a concept of my location. When I know something about myself but do not know that I am the person about whom I know this, I have two concepts of myself without realizing that they are concepts of the same individual and thereby without merging the two schemata into one (this would be a case of the *application of beliefs*). The same thing happens when I, while knowing a lot about Cicero, am unable to answer any question about Tully: Beliefs explain actions only when they are applied.

Adopting this redefinition of BO, we get:

$$d, c[\![\text{"believes that } \phi\text{"}]\!]a, e_0 \quad \Leftrightarrow \quad \begin{array}{l} \text{for each alternative } e \text{ compatible with all of } a\text{'s beliefs} \\ \text{as anchored at } c(\text{"believes"}) \text{ according to } e_0 \\ \; d, c[\![\phi]\!]e \end{array}$$

A *second approach* requires that the embedded statement in the attitude report captures the interpretation of some of the agent's beliefs. $\langle f, S\rangle$ is *a way of believing* ${}_{d,c}[\![\phi]\!]$ for a at l iff $d, c[\![\phi]\!]e$, whenever $a, l, f[\![S]\!]e$, i.e. iff the interpretation of S given a, l, f strongly implies (is a subcollection of) the interpretation ${}_{d,c}[\![\phi]\!]$. Then we define:

$$d, c[\![\text{"believes that } \phi\text{"}]\!]a, e_0 \quad \Leftrightarrow \quad \begin{array}{l} \text{there are } S,\ f \text{ such that } \langle S, f\rangle \text{ is a way of} \\ \text{believing } {}_{d,c}[\![\phi]\!] \text{ for } a \text{ at } l = c(\text{"believes"}) \text{ and} \\ \qquad \mathsf{in}\ e_0 : \mathsf{at}\ l : \ B_r, a, S; \mathsf{yes} \\ \qquad \mathsf{in}\ e_0 : \mathsf{at}\ l : \ \mathsf{of}, \dot{x}, f(\dot{x}); \mathsf{yes} \end{array}$$

$\langle f, S\rangle$ is *a way of failing to believe* ${}_{d,c}[\![\phi]\!]$ for a at l iff for every e: if $d, c[\![\phi]\!]e$ then $a, l, f[\![S]\!]e$, i.e. iff the interpretation ${}_{d,c}[\![\phi]\!]$ strongly implies (is a subcollection of) the interpretation of S given a, l, f. Then we define:

$$d, c[\![\text{"does not believe that } \phi\text{"}]\!]a, e_0 \quad \Leftrightarrow \quad \begin{array}{l} \text{there are } S,\ f \text{ such that } \langle S, f\rangle \text{ is a way of} \\ \text{failing to believe } {}_{d,c}[\![\phi]\!] \text{ for } a \\ \text{at } l = c(\text{"believes"}) \text{ and} \\ \qquad \mathsf{in}\ e_0 : \mathsf{at}\ l : \ B_r, a, S; \mathsf{no} \\ \qquad \mathsf{in}\ e_0 : \mathsf{at}\ l : \ \mathsf{of}, \dot{x}, f(\dot{x}); \mathsf{yes} \end{array}$$

Defined in this way, both belief and lack of belief are closed under strong consequence and subject to very weak rationality principles.[20]

Instead of specialising their account of belief to knowledge (e.g. by adding a veracity constraint), Barwise and Perry make a fresh start from Dretske's account (as explained in ch. 2), taking knowledge as primitive. Knowledge, according to Barwise and Perry, is not just belief caused by one's having information, but belief that *contains* information.

[20]Belief that $\neg\phi$ does not imply, e.g., according to this account, that one does not believe that ϕ. It may still be possible both to believe and to disbelieve ϕ, though *not in the same way*, e.g. by having both a positive "Cicero" and a negative "Tully"-belief.

They do not, however, spell out their notion of containment but just say that belief contains information iff it carries that information (Barwise and Perry 1983: 266). So they distinguish between three informational mental states which have the following form:

$$\mathsf{in}\ e_0 : \mathsf{at}\ l : \ \mathsf{F}, a, S; \mathsf{yes}$$
$$\mathsf{in}\ e_0 : \mathsf{at}\ l : \ \mathsf{of}, \dot{x}, f(\dot{x}); \mathsf{yes}$$

where F stands for inf (having information), B_r (believing) or K_r (knowing) respectively. Such mental states are informational in virtue of involving *attunement to constraints.* If "$S \Rightarrow S'$" denotes a constraint C, i.e. a situation of the form:

$$\mathsf{at}\ l_u : \ \mathsf{involves}, S, S'; \mathsf{yes}$$

cognitive agents are said to be *attuned* to C iff they have frames of mind which are of corresponding types $\#S$ and $\#S'$ such that the constraint $\#S \Rightarrow \#S'$ is actual. If a cognitive agent is moreover able to detect or discriminate situations of the S and of the S' type,[21] we get the following commutative diagram of 'internal' and 'external' constraints:

$$\begin{tikzcd}
S \arrow[r] & S' & \text{External significance} \\
\#S \arrow[u] \arrow[r] \arrow[ur, phantom, "\circlearrowleft"] & \#S' \arrow[u] & \text{Efficient mental states}
\end{tikzcd}$$

It is in virtue of being attuned to constraints, e.g. constraints linking perception and knowledge or the constraints of a common language, that we receive and transmit information. Barwise and Perry boldly conjecture (Barwise and Perry 1983: 292) that all natural language sentences have meanings which may be represented as constraints between types of events, utterances and described situations. The account of constraints they give in 1983, however, is far too crude. This is not the only problem, however. Almost all of the constraints we are attuned to are conditional: they only work properly under certain circumstances. In order to explain when and why information flow takes place, Barwise and Perry would have to spell out these conditions. This, however, would render the theory implausible, for, by their account of attunement, these extra-conditions would have to be mirrored in the efficient mental states, which they are not. That is why constraints became the "emergent theme"[22] of situation semantics and, later, situation theory in general.

[21] This requirement provides us with the vertical linkages in the diagram. The vertical constraints are what makes the mental states meaningful with respect to the environment (Devlin 1991: 101).

[22] "Constraints: the emergent theme" is a subtitle in the 1999 Preface to the CSLI reissue edition of

3.3 Criticism and later developments

One of the major later developments in situation theory was the development of a *logic* of infons (cf. sct. 4.2). In order to completely describe a possible state of the world (later taken to be a way of resolving all issues (cf. Barwise 1997: 494), we do not have to assign truth-values to all sentences that may be used to describe such issues. The representational system used is assumed to have a logical structure. In traditional modal logic, this is captured by the distinction between a set of propositional variables $\mathbb{P}$ and the well-formed formulae generated by a recursive set of rules. Much the same is done in situation theory, where infons are "intended to capture the basic informational items about the world" (Devlin 1990: 84). Types are then obtained by abstraction and identified with sets of situations. By union and intersection, we get disjunction and conjunction of types and so a logic of situation types begins to take shape. A basic problem, however, comes with negation (cf. e.g. Cooper and Kamp 1991). In the original theory, infons have been taken to be persistent, i.e. such that if an infon holds in some situation s, then it also holds in any situation of which s is a part. This rules out negative existential infons. Later, situation theorists have given up this persistence principle by allowing for negative infons in general (cf. Barwise 1989d: 235). Another problem, namely that simple complementation is not well-defined if the collection of all situations does not form a set, persisted (cf. sct. 4.8.5).

The emphasis on logic led to another crucial feature of many of the subsequent developments of situation theory. Based on situations rather than propositions, logic became all-pervasive and no longer bound to language:

> "... an understanding of logic and the related notions of meaning and inference requires us to look beyond language. Inference, for example, is an activity that attempts to use facts about the world to extract additional information, information implicit in the facts. A sound inference [...] is one that has the logical structure necessary to serve as a link in an informational chain but that need not use language at all." (Barwise 1986c: 39)

Such informational chains relate to another big issue in post-1983 situation theory: the theory of constraints, in particular the problem of how to characterise conditional constraints. Jerry Seligman (1990) proposed to make the notions of a situation-type and a constraint relative to what he called "perspectives", rather than, as with the conditional constraints in the 1983 theory, to some background type. Perspectives "limit the applicability of a collection of constraints to a specific domain of situations over which they are ubiquitous"

Situation and Attitudes (Barwise and Perry 1999: xxvii).

(Seligman 1990: 151). This specific domain became later the collection of normal situations. The biggest advantage of this move is that we do no longer have to require that our language is expressively rich enough to contain a background type for every constraint uniquely specifying situations in which the constraint applies.

Barwise even went as far as claiming that giving the logic of some activity is spelling out the constraints governing it:

> "External states, processes, and events have their logic. The latter processes are, in general, "inferences" about the former. Sound inferences are those that meet the conditions necessary to ensure that the resulting mental states contain information concerning the external situations they are about. The problem for our subject, then, [is] to get clear about these conditions necessary to preserve information." (Barwise 1986c: 53)

The idea expressed here is certainly appealing: For every constraint $S \Rightarrow S'$ there is a logic, i.e. some general conditions ensuring information flow along the corresponding constraint $\#S \Rightarrow \#S'$. Abstracting common features of such constraints on efficient mental states, we get a general theory of sound reasoning.

The emphasis on constraints made the distinction between situation meaning and situation-type meaning (cf. p. 79) more and more important.[23] It became clear that this distinction was needed not only to account for indexicals, but that it was a pervasive feature limiting the applicability of constraints working at the level of situation-types:

> "... the meaning of a sentence [that is, the situation-type meaning of any of its utterances] underdetermines the content of any particular statement made with it [i.e. the situation meaning of such utterances]." (Barwise 1988: 64)

We will see later (in sct. 4.8.3) that this distinction became one of the main motivations to depart from ordinary model-theoretic semantics in the description of information flow:

> "They [possible world semantics and situation semantics] attach differing significance to the gap between a sentence and an assertion or statement made with it. Possible world semantics treats the gap as an annoyance, whereas situation semantics takes it to be at the heart of meaning." (Barwise 1987: 79)

This quote somehow misrepresents at least the early stage of situation semantics, where the dependence of propositional content of utterances on the perspective (utterance situation) of the speaker was tried to be captured by postulating unarticulated constituents of the

[23]Devlin, however, boldly identifies situation-type meaning with meaning tout court: "... the meaning itself is a linkage between *types*, not between situations." (Devlin 1991: 90)

descriptive content of the utterance. Only in 1989 and in particular with Seligman's work on perspectives came a different model to be adopted, where different agents may have different points of view – and hence different focus situations – even with respect to what is, "from an unsituated perspective" one situation, i.e. the situation that makes both their utterances "the salt is to the left of the pepper" and "the salt is to the right of the pepper" true (cf. Barwise (1989d: 238) and also Barwise (1989a: 271)).

Chapter 4

Information: first steps towards a logic of information

4.1 What a logic of information might be

Given our lack of clear-cut intuitions about information (as observed in sct. 1.1) and the empirical and conceptual implausibility of the picture of knowledge emerging from epistemic modal logic (cf. sct. 1.3), one might be tempted to define "information" to be that item of the objective world, whatever it is, that is adequately modelled by epistemic modal logic. Many authors have, at least tentatively, succumbed to this temptation. Indeed, it has become common practice to append disclaimers to the introduction of epistemic modalities, claiming that they are not supposed to model the actual knowledge some agents really have, but the knowledge they had if they were fully rational, the knowledge they are (ideally) in a position to acquire, the knowledge we can, under certain idealisations, externally ascribe to them etc. A handy way to summarise such qualifications, I argued in sct. 1.4.1, could seem to be that $\mathsf{K}_a p$ is not to be read as "a knows that p" but as "a has the information that p".

The logic of information some are looking for would then appear to be just what logic, first-order, classical, but especially modal logic, has been all along: a study of the interrelationships of information-carrying items, no matter whether they are called "infons" or "propositions". The notion of information at hand, some theorists have thus tried to redefine the subject-matter of epistemic modal logic:

> "One may very well doubt that there is an intuitively acceptable notion of knowledge for which **S5** is appropriate. We argue that it may be appropriate, however, for a

certain notion of *information*." (Israel 1989: 3)

Such a reinterpretation seems to be also one of the guiding ideas of situation theory.[1] In a 1985 comment on the criticisms of *Situations and Attitudes*, cast in interview form, we find the following passage:

> "**B[arwise]**: But maybe this is just what propositions are, an invariant across information and misinformation.
> **I[interviewer]**: So true propositions are information; false ones are misinformation?
> **P[erry]**: Something like that sounds pretty plausible to me." (Barwise and Perry 1985: lxxii)

If propositions are just informational invariants, a theory of information will be a theory of propositions and so include at least ordinary propositional logic.

As late as in 1990, John Perry still believed that "the informational content of a fact is a true proposition" (Israel and Perry 1990: 3). So it seems that in the early days of situation semantics, pieces of information and true propositions have been regarded as closely related and inter-convertible. It was only in 1991 that Devlin suggested a crisp contrast between the two notions:

> "For purposes such as these [design of information processing devices and analysis of ordinary language use], a 'logic' based on truth (such as classical logic) is not appropriate; what is required is a 'logic' based on *information*. It is the aim of this book [i.e. *Logic and Information*] to develop such a 'logic'." (Devlin 1991: 10)

Barwise adopted a different conception of propositions from 1987 on, one that conceives of propositions as of the form $s \models \sigma$, where s is the focus situation of some agent (the smallest portion of the world which contains all that is within focus for the agent, without necessarily being itself an object of focus) and σ is a state of affairs used by the agent to classify s (an infon, in the later terminology) (Barwise 1989d: 228).

The information carried by the utterance of a true proposition, then, is split up into an external component, the situation talked about, and an internal component, a property ascribed to (and true of) that situation. In this chapter, I would like to explore some of the consequences of this move, present in detail the alternative picture emerging from epistemic modal logic based on Kripke semantics and finally identifying the desiderata for a theory of information flow taking into account the guiding ideas of situation semantics, improving on its deficiencies, while avoiding the weaknesses of the classic Kripke-style approach. Ch. 5 will then move on to an exposition of such a theory of information flow.

[1] Cf. "...I would argue that much of the work in the logic of knowledge is best understood in terms of the logic of information." (Barwise 1988: 204)

4.2 Information bearers

Situation theory takes the basic structure of information flow to be given by the following:

(1) a's being F carries the information that b is G.

We can, but do not have to, conceive of a's being F as a situation.[2] What is important, however, is that a's being F involves a certain particular or *token*, a, being of a certain *type*, namely F. David Israel and John Perry call a's being F, occurring in such informational contexts as (1) an *indicating fact* 1990: 1. What connects a's being F ($a \models F$) and b's being G ($b \models G$) is a constraint of the system, some informational dependency between a and b in virtue of which the former's being F carries the information that the latter is G. Such systems may be seen as *classifications*, ways of carving up situations in the world according to (some of) the types they support.

First order logic represents (1) as follows, for a suitable binary relation $\rightarrow^*$:

(2) $Fa \rightarrow^* Gb$

The crucial disanalogies between (2) and the situation-theoretic account of information flow (1) are: (i) that the situation theoretic account does not give a *local* analysis to the constraint, representing it as a fact *about* a certain point in the system (possible world or whatever), but incorporates it into the structure of the system as a whole; and second (ii), that (2), but not (1), does commit us to a propositional structure of the types; situation theory, but not first order logic, allows, e.g., for subsentential conjunction of the form $a \models F \wedge G$, thereby dodging at least part of what has become known as the variable-binding problem, i.e. the problem of how to represent the fact that both occurrences of the individual constant in "$Fa \wedge Ga$" stand for the same individual.

(i) does not mean that constraints cannot be considered as entities in their own right. Indeed they can and they even have to, as we combine isolated classifications to information channels (cf. def. 5.1.4). The crucial point is rather that the interconnected entities get a representation on their own and may thus independently enter into logical relationships, instead of simply being embedded in a complex representation as in (2).

As we saw in ch. 3, there is another kind of information bearer: parametric states of

[2]Though this is certainly the intended interpretation: "In situation theory we try to account for the flow of information in terms of constraints between types of situations. A situation of such-and-such a type carries the information that there is a situation of so-and-so a type." (Barwise 1989c: 224, n. 1)

affairs, called "Russellian propositions" by Barwise and Etchemendy in 1987 and generally called *infons* from 1990 on (Israel and Perry (1990: 9), Devlin (1991: 22). Devlin takes an infon to be a "discrete item of information" (Devlin 1991: 11) which, when corresponding to the ways things actually are in the world, may also be called a "fact" (1991: 23).[3] Formally, he construes them as equivalence classes of pairs of representations and constraints under the relation of denoting the same item of information (Devlin 1991: 40) or giving rise to the same item of information (1991: 47).[4] He gives them the same ontological status than he assigns to mathematical entities: like the real numbers, they are "abstract objects created by the human mind" (Devlin 1991: 44) but that can be modelled in different ways and are, in this sense, representation-independent:

> "...infons are semantic objects within the framework of our theory. That is, their status is the same as, say, the real numbers within mathematics. This gives them an 'absolute' nature, independent of representation. (Though this absoluteness is relative to an original ontology of individuals, locations, relations, and what have you, that is essentially dependent on the agent or species of agent under consideration.)" (Devlin 1991: 48)

The possibility of compounding infons leads Devlin to develop what he calls an "infon logic" (cf. sct. 3.3),[5] a project (called the development of a "calculus of states of affairs" by Barwise (1989d: 223)) undertaken from a more algebraic perspective by Seligman and Moss (1997).

What then becomes of the ordinary logic of propositions? We saw that in Barwise's and Perry's 1983 theory (ch. 3), the interpretation of a statement was a situation. This situation, however, may carry information which is extraneous to the sentence, e.g. that the sentence is couched in some language or other, that languages, utterances and speakers exist and so on. It seems weird, however, to call something informationally richer than a given statement an interpretation of it. This was why Barwise and Perry re-introduced propositions in 1985. Barwise subsequently identified the propositional content of a statement with the proposition that the type of situation described is realised (Barwise 1986a:

[3]More precisely, he distinguishes between non-parametric infons which may be facts and parametric infons, which are got from the former by abstraction: "Non-parametric infons, or 'states of affairs' are the basic items of information about the world; parametric infons are the basic units our theory uses in order to study the transmission of information." (Devlin 1991: 123)

[4]Seligman and Moss (1997: 2) define infons, relations and roles by two relations Rel and Arg taken as primitives of the theory.

[5]Devlin does not take such a logic to give us new infons but rather something different than infons, something he unfortunately calls "compound infons" (Devlin 1991: 132). But this idiosyncrasy need not concern us further.

119, n. 7). In 1987, this messy situation was finally remedied by the distinction between Russellian and Austinian propositions.

Russellian propositions are just infons.[6] Austinian propositions are abstract entities of the form $s \models \sigma$, where s is a situation and σ an infon, the two connected by a binary relation $\models$. They correspond most closely to what are usually called "propositions".[7] Propositions, in Barwise's view, are somehow derivative entities:

> "What comes first is the characterization of focus situations [what "s" stands for in "$s \models \sigma$"] in terms of facts they support. It is only when we are interested in stepping back and characterizing someone's characterization of their focus situation that we encounter propositions." (Barwise 1989c: 229)

We thus get the following two-level picture. The basic level of information bearers consist of the characterisation of situations in terms of sub-propositional items, so-called 'infons', describing local properties of situations.[8] Because they are properties of situations, they can generally and unproblematically taken to be *about* the situations they are properties of. Only on a second level comes the evaluation of claims for truth and falsity. Such claims are given the canonical form $s \models \sigma$[9] and are true or false as a function of whether the infon σ correctly or incorrectly describes the situation s.

Hence, as we noticed in sct. 1.4.2, information is closely connected to the general

[6]Confusingly, Barwise sometimes assimilates Russellian propositions to Austinian ones: a Russellian proposition then is just an infon made factual by the total situation which is the whole actual world: "If we assume [...] that the actual world is a situation, and we assume [...] that every situation is non-perspectival, then for every infon σ there will be a distinguished Austinian proposition ($W \models \sigma$), the claim that the world makes σ a fact. We call this the ***Russellian proposition*** associated with σ." (Barwise 1989a: 273) As there are good reasons to deny the claim that there is such a total situation (a question left open by the 1983 theory), I will avoid this way of speaking in the following. Cf. also Barwise (1986e)'s comment on "Corollary 3" ("There is no largest situation"): "One might find this a bit odd, since it seems to show that reality, all that is, is not a situation. However, all it really shows is that there is no possibility of having reality be a completed totality that can be treated as a first class citizen, for it always outstrips our attempts to comprehend it as a totality." (Barwise 1986e: 191) Devlin, in turn, argues for the claim that the world is not a situation from his analysis of the Liar sentence (Devlin 1991: 288).

[7]Devlin calls propositions constructs of the form "$x : T$" (object x is of type T). Infonic propositions are the special case where T is a situation-type (Devlin 1991: 62–63, 136–137).

[8]These properties are not to be individuated by necessary coextensiveness – at least not if the realm of situation is not fine-grained enough. This similarity between infons and propositions, traditionally conceived, has been well brought out by Seligman and Moss: "On the one hand, information is *representation-independent*: the same information may be represented in many different ways. On the other hand, information is *fine-grained*: two pieces of information may be logically equivalent without being identical." (Seligman and Moss 1997: 258)

[9]This is a substantial claim, even if situation-theorists sometimes claim to use 'proposition' innocently, as just "a term denoting that by virtue of which a statement has the truth-value it does" (Seligman and Moss 1997: 295)

notion of aboutness. This feature of information is the linkage between infons, the bearers of information, and propositions:

> "...information is intrinsically information *about* some portion of the world, that is to say, it is information about some situation. Thus there is a sense in which information is essentially *propositional.*" (Devlin 1991: 142)

It is propositional in the sense that it provides us with material that may be evaluated for its truth or falsity. The point where it is evaluated, however, the portion of the world it either matches or not, is not given by the item of information itself. This is the point of the relativity of information (cf. sct. 1.4.7). *We* bring information to bear on the world and use it to form propositions, which then in turn are true or false.

4.3 Information and constraints

As it has been remarked in (1.4.4), information is not to be identified with meaning and can plausibly seen to be more fundamental than the latter. Information, but not meaning, is in the world. Seen from this perspective, information is meaning *objectified*:

> "We try to drive a wedge between meaning and information by saying that information depends on the relevant constraints being actual, and the conditions right, while meaning also requires attunement to the constraints. That is, meaning depends on the presence of minds, or information processors more generally." (Barwise and Perry 1985: xlix)

There is a connection, however: information, while objective, is given to us only with respect to some representational system or other (as we already saw for the qualitative concept, in sct. 1.4.7, and for the quantitative concept in sct. 2.1.2).[10] *What* and even, as we saw, *how much* information we receive by a given signal depends on the coding chosen. That is why the main theorems in communication theory are concerned with the possibility of an optimal *coding* under certain circumstances and with respect to the transmission of certain types of signals along certain channels. But this does not mean that what and how much information *there is* so depends on the code. This is a point well brought out by Barwise and Seligman:

[10]This does not mean that we have to have access to or knowledge of some specific representational system in order to take some item to be informational: even before the discovery of the Rosetta Stone Egyptian hieroglyphics were justifiably (and correctly) regarded as informational. The information they carry was inaccessible (then), but still there.

> "Logical and linguistic investigations into the topic of information give the impression that one should be concerned with properties of sentences. Even when it is acknowledged that information is not a syntactic property of sentences and that some system of interpretation is required to determine the information content of a sentence, the role of this system is typically kept firmly in the background." (Barwise and Seligman 1997: 7)

This is just the charge, met above, that the constraints, in virtue of which information flows, do not enter into the traditional picture. This, then, is another motivation for the explicit modelling and thus the reification of constraints (cf. sct. 4.8.3): information is relative to constraints; carrying information is not an intrinsic property of a situation (cf. Israel and Perry 1990: 4). What information a signal carries, then, is not only relative to the background knowledge of the receiver, but also to the representational system in use. The more expressive the language, the more possibilities at the source can be described by it, the more information a signal carries describing one of them (at least if we can take the Brentano Principle (1.4.1) as our guide). The language, then, plays its rôle in defining the *relevant issues*.

In the theory we will develop in much detail later, this is nicely mirrored by the central notion of a classification (def. 5.1.1 in ch. 5). A classification is a classification of things by parts of the language available in that classification. As the expressible issues may or may not coincide with the relevant ones, we may or may not have reason to admit for indistinguishable tokens. Distinguishability becomes a very relative notion. This does not mean that Barwise and Seligman have an "internalist conception of information – one where information is relative to an agent's information processing ability", as Oliver Lemon (Lemon 1998: 398) ascribes to them. What and how much information there is is an entirely objective matter. What and how much of it is decoded (and even decodable) and may be put to work, however, depends on the means, in particular the linguistic means, at hand.

Another important issue is the relation between information and causation, i.e. the question whether there are non-causal constraints supporting information flow. It seems that the answer must be in the affirmative. As we saw on p. 48, informational dependency has to be distinguished from causality. While backwards causation is metaphysically dubious, it is straightforwardly possible for present events to carry information about their past causes.

Informational dependency, however, neither is just the symmetric closure of causal dependency, as shown by another distinction between information flow and causality: Dretske showed that a signal (e.g. a sensory experience) can carry information about the source without carrying information about the more proximal members of a branching causal chain along which this information is communicated (Dretske 1981: 158). The content transmitted in information flow may well be semantic, i.e. not carrying information about its causal origins; the branching is then 'digitalised away'.

There is thus another crucial difference between informational and causal dependence: the former is specific in a way the latter is not. Even in cases where information flow is underwritten by a causal chain, the signal can carries information only about *some*, but not all events and conditions in its causal history. This is partly a matter of coding and depends on range of relevant alternatives, e.g. the expressive resources of the language of which the signal is a sentence. Apart from this, however, it is also a matter of different possible causal chains underlying the same informational dependence: when the causal chains underlying the information flow could be different even while the signal would continue to carry the same information, the information carried does not depend on (and thus not testify about) the events along this contingent causal chain.

Even when the causal antecedents are unique and no branching is taking place, we can distinguish between them. This is done, e.g., by Dretske's notion of *primary representation.*[11]

That such a narrowing down of the range of objects which are such that a given piece of information can legitimately be taken to be about them, is possible clearly shows that information flow is a much more specific and fine-grained phenomenon than causal dependence, even if one adopts a realistic, cause-and-effect account of the difference information makes to the world.[12] Information, but not causality, classifies the world in intentional terms.

[11]Dretske defines it as follows: s gives *primary representation* to the property F of a relative to the property G of b iff s carries information about a and about b but represents a's property F by means of representing b's property G, i.e. iff s's representation of b's property G depends on the informational relationship between a's being F and b's being G but it's representation of a's property F does not depend on that relationship. This definition allows Dretske to define the "object of experience", i.e. the (possibly unique) object the information carried by some signal is about. This is the object which produces information (and hence is distant enough not to depend on contingent causal chains) and to which properties the signal gives primary representation.

[12]Dretske recently emphasised this: "If an event's carrying information doesn't make a difference – and by a difference here I mean a causal difference, a difference in the kind of effects it has – then for all philosophical (not to mention practical) purposes, the even doesn't carry information." (Dretske 1990: 112)

4.4 Epistemic and informational alternatives

We already saw that (in sct. 1.1) that information cannot just be identified with knowledge. We discussed, however, in sct. 1.4.1, 1.4.3 and 4.1 the idea of reinterpreting epistemic modal logic in informational terms, changing a notion of knowledge idealised beyond recognition for a suitably robust and realist notion of information. Knowledge as modelled in epistemic modal logic is truth in all epistemic alternatives. The question therefore arises whether there is a corresponding notion of informational alternatives. In this section, I want to pursue this question and lay out some differences between epistemic and informational alternatives.

A crucial difference between epistemic and informational alternatives concerns the issue of *granularity*. Granularity is the phenomenon that the world and any relational system in particular can be described at various levels of detail and so may be target of richer and poorer representations. The first case may be illustrated as follows: The table I am sitting at is wooden, rectangular and white; but I am equally true to say that I am sitting in front of a complex and highly structured aggregate of molecules, molecules which are neither wooden, nor rectangular, nor white. In describing the thing I am sitting at as a table, I already indicate in what features of it I take interest: that it has a plane surface to write on, is solid, does not move by itself and so on. Though I am not presently interested in the molecular structure of my table, I could become so interested and then would have to choose another more detailed representation of it, one depicting the structure of its molecules. We may thus say that my representation of the thing in front of me, like any other representation, comes with a certain granularity, a grid at the level of which the constraints of interest between its components operate.

Granularity has to be neatly distinguished from *perspectivism*, a doctrine which says that granularity is not only one-dimensional, that different representations of one thing are not only richer or poorer, depending on the grid they impose, but that representations come at *incommensurable* levels of generality. Vision might perhaps provide an example. If both of us are directing our eyes to the same object, what we see may differ in two ways: if we are looking from the same direction, but at different distances, one of us is likely to see *more* than the other. If we are looking from different directions, however, what we see may be different, in a sense, though aspects of one and the same thing we are directing our eyes at. Perspectivism assimilates the difference in granularity to the difference in perspective, claiming that any level of granularity corresponds to a different aspect of the thing represented.

Epistemic and informational alternatives now differ in that the latter, but not the former, allow for the perspectivist account of granularity. Epistemic alternatives, as traditionally construed, are alternative ways the world I take myself to be living in might be, given what I know. These ways, however, are required to be ways *the world* might be, i.e. complete specifications of everything, settling every issue. They are, to be speak metaphorically, worlds of the same kind than the actual world, with the only difference that they are only (epistemically) possible and not actual (nor taken to be actual). I may specify them at different levels of granularity, but this is a modeller's choice, to be made from an external perspective and once for all. It is not open to me, for example, to specify the same world in two different ways without deciding whether I (or the agent whose epistemic alternatives I am modelling) take these two worlds to be identical.

Informational alternatives, on the other hand, seem to allow for perspectivism. If two people tell me the story of a car accident they saw, I learn something from their account even if I do not know and do not even have any beliefs about whether they saw one and the same accident. If I learn this or get information that allows me to form reasonable beliefs about that issue, I learn more and different things from what I already knew before. So informational alternatives are not necessarily transparent. A given epistemic alternative either is or is not the same than any other specified in different terms: it cannot leave any question open.

Another difference is connected with the issue of aboutness. We saw (in sct. 1.4.2) that a given piece of information is essentially about some quite specific aspect of reality. Informational alternatives, then, depict different ways things about which I have information might be. They are essentially local, depicting minor deviances of reality, anchored in real things I have information about. It is not possible to have information about non-existing things, even if those things are possible. Epistemic alternatives, on the other hand, take a stance on everything: they completely specify matters about things the agents have never heard. That is why they may only in a derivative sense be said to be about something: they may perhaps be said to be about the real world, but even this is problematic.

An especially vivid case of this contrast is given by the truth axiom (T) which says that the real actual world is an epistemic or informational alternative respectively. Interpreted in informational terms, the axiom is straightforward: no wonder is the actual world compatible with the information we have it, for we do not have information about anything else. Interpreted epistemically, however, the axiom seems problematic: for agents having false beliefs will not consider the actual world to be a way the world might be. To say that the actual world is not ruled out by what they really know is of little help: for they, short of being omniscient, will not be able to specify one such world as the actual one. To take a world as actual, it seems, more is required than just not being able to rule it out. The externalist picture of knowledge of use in theoretical computer science that privileges the modeller's point of view against those of the (often inanimate) agents is thus more plausibly taken to be a picture of *information*.[13]

This brings us to a third difference between epistemic and informational alternatives, concerning the agents' self-location in epistemic space. If epistemic alternatives are to be alternatives *for* the agent, ways the world might be he considers to be compatible with what he knows about it, they are not alternatives to the actual world but alternatives to the world the agents *take to be* actual. Epistemic agents act on what they believe they know, not just on what they really know.

We have to distinguish between two sorts of epistemic possibility, corresponding to a referential and an attributive reading of the phrase "the actual world" in "a way the actual world could have been (given what we know about it)". According to its referential interpretation, anything which is compatible with our knowledge and hence might be, for all we know, a world we happen to inhabit qualifies as an epistemic possibility. Anything we cannot rule on the basis of what we know about *this* world is possible in this first sense, even if it would preclude this world's being a world we inhabit, being a world where we have the knowledge we have or being a world about which we have any information. An attributive reading, however, demands more: to be epistemically possible in this second way, a possible world not only has to be one of which we do not know that it is not ours, but more, namely one which we could *take to be* our actual world. It seems to me that this second sense is closer to our usual use of "knowledge", while it is clearly the former which is relevant to "information".

To bring out this difference, consider again Moore's paradox (14) already discussed in sct. 1.3.4:

(3) p, but I do not believe that p

Hintikka, we saw, argued from the weirdness of (3) to the transitivity of doxastic and

[13]Stalnaker (1999b: 258) characterises this picture as follows: "As distributed systems theorists have emphasized, their conception of knowledge is an externalist one in the sense that the content of a knowledge claim is characterized from the point of view of the theorist, and not of the knower. The language of the epistemic logic talks about what processors know, but it is not intended to model the knower's way of expressing or representing what it knows."

hence epistemic alternativeness. Against this we argued, that (3) is inappropriate to utter because it would not be uttered in a state of complete information. The contrast may now be deepened. The following cousin of (3) is not only (too often) true, but clearly an appropriate thing to say in many contexts:

(4) p, but I do not know that p

That I may even, on some occasions, be justifiably said to *know* that (4) is true constitutes another argument against (S4). But still it is weird to utter (4): this might explain why it is rarely uttered even on occasions where it should be. It seems, then, that Hintikka's arguments against the appropriateness of (3) carry over to (4), thereby ruling out too much. It seems to me that whatever weirdness (3) or (4) may be said to have derives from their kinship to the following, clearly paradoxical, statement:

(5) p, but I do not have the information that p

(5) seems to me by far the most clearly paradoxical of these three statements: it not only violates pragmatic maxims, but goes against the constitutive rule of assertion which is to provide information.

It seems then, that both belief and knowledge are notions usefully modelled entirely from the agent's viewpoint. It is the agent's perspective, his beliefs and his knowledge, which determines what possibilities or impossibilities have to be taken into account in order to describe his epistemic and non-epistemic behaviour. To limit the range of alternatives is automatically to misconstrue (some would prefer to say: idealise) his epistemic behaviour. The relativity of information (cf. sct. 1.4.7), however, is of an altogether different nature.

Information, as Dretske repeatedly emphasised, is an objective commodity (cf. p. 4). This brings a fourth difference to the fore: informational alternatives may suitably be modelled from an external perspective, while epistemic alternatives are determined by and depend on the beliefs and the knowledge of the agent. Which of them are relevant depends on the world and not on our beliefs about it.

It is precisely on this point that Dretske's theory has met with much criticism, either for being too relativistic (Sterelny 1983: 209) or not flexible enough (Morris 1990: 389) about what counts as a relevant possibility. In fact, however, Dretske makes a perfectly clear and valid point: What is relative on the agent is the *partitioning* of the event space, the possibilities he attributes to the source, both by his prior knowledge about the source and by drawing a pragmatically motivated distinction between source and channel. These factors directly affect the surprisal value of the signal and hence the information the agent is liable to receive. What is not relative on the agent and what is irrelevant for the information-theoretic account of his epistemic behaviour, on the other hand, is what the agent believes or knows about the possibilities at the source or the channel: the agent does not have to know or to believe that alternative states have been excluded or that the channel is secure to receive information through it (Dretske 1981: 123). The agent-relative partitioning of the event space at the source, so to say, is a partitioning of an objectively existing and agent-independent reality.

What then are epistemic alternatives? In the framework to be developed below (sct. 4.5), they are taken to be relations on the set of possible worlds. If w and v are two such possible worlds, wR_iv (where R_i is a binary relation indexed by some $i \in \mathcal{A}$, where $\mathcal{A}$ is a set of epistemic agents) is taken to mean "v is a way the world might be from the point of view of w". As we saw above, we are not entitled to assume that w is taken, by the agent, to be his actual world. It is rather something like an epistemic state, described from the outside. We may perhaps describe the situation most accurately from the modeller's point of view. The modeller, assumed to be omniscient, asks himself the counterfactual question: "What might I take to be my actual world if my epistemic state were that modelled by w?" If he finds out that his epistemic state modelled by w would not rule out the actuality of v, he includes (w, v) into R_i.

There are several problems with this heuristics. One of them, to be discussed in sct. 4.8.1, is the reliance on the unexplained notion of a possible world. Another one (to be discussed in sct. 4.8.4) is that the relations of epistemic alternativeness, indexed by the set of agents and supposed to model their respective epistemic behaviour, are individuated extensionally. Only the agents' numerical multiplicity is relevant, not their actual epistemic practice. It is not possible, therefore, to model the dynamics of the epistemic states of an agent, the way it changes in time and in response to deliberation, changes of mind and the incorporation of new information,

We will see, in sct. (4.6), a first tentative to remedy some of these shortcomings by explicitly coding the relational structure of epistemic alternativeness into the points of our models, thus changing 'possible worlds' for 'epistemic states'. Such an is then characterised by the truth-values the agent a assigns to the propositional variables $p \in \mathbb{P}$ by taking himself to be in w and the *information state* he believes himself to be in, i.e. the set $\{x \in W \mid wR_ax\}$ of those worlds which are epistemic alternatives for a in w, that is worlds that, as far as a can tell, could be what he takes to be his actual world. We will then, from ch. 5, greatly expand and generalise this model to a general theory of information flow.

4.5 Kripke frames and models

In this chapter, I will present and discuss some important notions of epistemic modal logic, both to confront the philosophical remarks made with a clear picture of what is at stake and to be able to draw later on notions presented here, in the discussion of Barwise's and Seligman's logic of information flow. The approach will be semantical: in contrast to the syntactic discussion in sct. (1.3), not the logics themselves but the underlying relational structures will be introduced. Seen from the Amsterdam perspective, mentioned in the preface, these relational structures are considered the main modelling devices, the logics being just a handy way to describe them.

For a start, then, we need the notion of a Kripke frame:

Definition 4.5.1 (Kripke frames). *A (multi-modal)* Kripke frame *$\mathcal{F}$ is a tuple $\mathcal{F} = \langle W, \{R_i\}_{i \in \mathcal{A}} \rangle$ of a nonempty set of worlds W and a finite set of relations $R_i \subset W \times W$, indexed by $i \in \mathcal{A}$.*

We will think of $\mathcal{A}$ as our sets of *agents* and of the relation R_i as some relation of *epistemic alternativeness*. Whatever it is more specificly, this relation is meant to describe the epistemic state of the agent in the following way: to know (or to be capable of knowing or to know ideally etc.) that p is for that agent to have only epistemic alternatives where p. We already discussed some aspects of the crucial question of how to interpret relations of epistemic alternativeness in sct. (4.4).

While our agents are given a set of knowable propositions $\mathbb{P}$, they may also know of themselves and of each other that they know something, they may know logically complex combinations of such propositions and they may also ascribe to themselves and others knowledge of logically complex propositions. This is why Kripke models are defined with respect to the formulae of a multi-modal language $\mathcal{L}$ which we define with respect to a set of propositional variables $\mathbb{P}$:

Definition 4.5.2 (Our standard modal language). *Suppose a finite set of propositional variables $\mathbb{P}$ is given. $\mathcal{L}$ is the smallest class such that: $\top \in \mathcal{L}$; $\mathbb{P} \subset \mathcal{L}$; if $\phi \in \mathcal{L}$, then $\neg\phi \in \mathcal{L}$; if $\phi, \psi \in \mathcal{L}$, then $(\phi \wedge \psi) \in \mathcal{L}$; if $\phi \in \mathcal{L}$ and $a \in \mathcal{A}$, then $[a]\phi \in \mathcal{L}$.*

We abbreviate $\neg\top$ by $\bot$, $\neg(\neg\phi \wedge \neg\psi)$ by $\phi \vee \psi$, $\neg[a]\neg\phi$ by $\langle a \rangle \phi$ and $\neg(\phi \wedge \neg\psi)$ by $\phi \rightarrow \psi$.

We will also have use for a modal language allowing infinitary conjunctions:

Definition 4.5.3 (Infinitary modal languages). *Suppose a finite set of propositional variables* $\mathbb{P}$ *and a cardinal* $\alpha \in \mathbf{On}$ *is given.* $\mathcal{L}^\alpha$ *is the smallest class such that:* $\mathbb{P} \subset \mathcal{L}^\alpha$*; if* $\phi \in \mathcal{L}^\alpha$*, then* $\neg\phi \in \mathcal{L}^\alpha$*; if* $\phi, \psi \in \mathcal{L}^\alpha$*, then* $(\phi \wedge \psi) \in \mathcal{L}^\alpha$*; if* $\phi \in \mathcal{L}^\alpha$ *and* $a \in \mathcal{A}$*, then* $[a]\phi \in \mathcal{L}^\alpha$*; if* $\Phi \subset \mathcal{L}^\alpha$ *is a set and* $|\Phi| < \alpha$*, then* $\bigwedge \Phi \in \mathcal{L}^\alpha$*.*

We will call the formulae of $\mathcal{L}^\alpha$ formulae of *(syntactic) degree* α. $\mathcal{L}$ is the class of formulae with `degree` ω. By the clause for finite conjunction and by $\bigwedge \emptyset = \top$, we have $\mathcal{L} \subset \mathcal{L}^\alpha$ for any $\alpha \in \mathbf{On}$. $\mathcal{L}_\infty$ is the set of all formulae which are formulae of a language $\mathcal{L}^\alpha$ for some cardinal $\alpha \in \mathbf{On}$.

If Φ is a set of formulae indexed by a set, say $\Phi = \{\phi_i \mid i \in \mathcal{I}\}$, we write $\bigwedge_{i \in \mathcal{I}} \phi_i$ for $\bigwedge \Phi$. If Φ is indexed by a $\{1, \ldots, n\}$. we write $(\phi_1 \wedge \phi_2 \wedge \cdots \wedge \phi_n)$ for $\bigwedge \Phi$. We will use the standard abbreviations $\langle a \rangle \phi$ for $\neg[a]\neg\phi$ and $\bigvee \Phi$ for $\neg \bigwedge \{\neg\phi \mid \phi \in \Phi\}$.

To pick out subsets of $\mathcal{L}_\infty$, we define the *modal depth* of a formula of $\mathcal{L}_\infty$ (as in (Milner 1990: 1237), Barwise and Moss (1996: 143), Blackburn et al. (2001: 74)):

Definition 4.5.4 (Depth of a formula). `depth` *is a function from* $\mathcal{L}_\infty$ *to ordinals satisfying the following (for any* $a \in \mathcal{A}$*):*

$$\begin{aligned}
\mathtt{depth}(p) &= 0 \\
\mathtt{depth}(\top) &= 0 \\
\mathtt{depth}(\neg\phi) &= \mathtt{depth}(\phi) \\
\mathtt{depth}(\phi \wedge \psi) &= \mathbf{max}\{\mathtt{depth}(\phi), \mathtt{depth}(\psi)\} \\
\mathtt{depth}([a]\phi) &= \mathtt{depth}(\phi) + 1 \\
\mathtt{depth}(\textstyle\bigwedge \Phi) &= \mathbf{sup}\{\mathtt{depth}(\phi) \mid \phi \in \Phi\}
\end{aligned}$$

For each ordinal α, $\mathcal{L}_\alpha$ is the class of all formulae of $\mathcal{L}_\infty$ of depth $\leq \alpha$. Because the ordinary modal language (4.5.2) consists of formulae ϕ of `degree` ω and hence $\mathtt{depth}(\phi) < \omega$, we have $\mathcal{L} \subset \mathcal{L}_\omega$.

We are now ready to introduce the central notion of a Kripke model, which is a Kripke frame together with a valuation:

Definition 4.5.5 (Kripke models). *Let a set* $\mathbb{P}$ *of propositional variables be given. A* Kripke model M *is a triple* $M = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}} \rangle$*, where* $\langle W, \{R_i\}_{i \in \mathcal{A}} \rangle$ *is a Kripke frame and* $\pi : \mathbb{P} \to \mathcal{P}(W)$ *is called a* valuation *of* $\mathbb{P}$ *(where* $\mathcal{P}(W)$ *is the powerset of* W*). A* Kripke world *is a pair* (M, w) *of a Kripke model* $M = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}} \rangle$ *and a world* $w \in W$*.*

We write w for (M,w) whenever the reference to the underlying Kripke model is clear. Truth of $\mathcal{L}_\infty$-sentences in Kripke worlds is defined in the usual way:

Definition 4.5.6 (Truth in worlds). *We define for a Kripke world (M,w) and $\phi \in \mathcal{L}_\infty$:*

$$\begin{array}{rcl}
(M,w) \models \top & :\Longleftrightarrow & w \in M \\
(M,w) \models p & :\Longleftrightarrow & w \in \pi(p) \\
(M,w) \models \neg\phi & :\Longleftrightarrow & (M,w) \not\models \phi \\
(M,w) \models \phi \wedge \psi & :\Longleftrightarrow & (M,w) \models \phi \;\; \textit{and} \;\; (M,w) \models \psi \\
(M,w) \models \bigwedge \Phi & :\Longleftrightarrow & (M,w) \models \phi \quad \textit{for all } \phi \in \Phi \\
(M,w) \models [a]\phi & :\Longleftrightarrow & (M,w') \models \phi \quad \textit{for all } w' \textit{ such that } wR_a w'
\end{array}$$

Validity in a model is defined as truth in all the worlds of that model.

Definition 4.5.7 (Validity). *A formula $\phi \in \mathcal{L}_\infty$ is* valid *in a Kripke model M, written $M \models \phi$, iff it is true in all worlds in that model, i.e. iff $(M,w) \models \phi$ for all $w \in W$.*

As argued in (4.4) an important species of informational equivalence between Kripke models is bisimilarity. We begin by a notion of partial bisimulation, which we define as a triadic relation between Kripke worlds and von Neumann ordinals $\alpha \in \mathbf{On}$:

Definition 4.5.8 (Ordinal bisimulation). *If $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W'.\pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ are two Kripke models for the same set of agents $\mathcal{A}$, a nonempty relation $\cong_{\mathbf{On}} \subset W \times W' \times \mathbf{On}$ is an* ordinal bisimulation *between M_1 and M_2 iff, for all $w \in W, w' \in W'$ and $\alpha \in \mathbf{On}$, the following three conditions hold if $(w, w', \alpha) \in \cong_{\mathbf{On}}$:*

$$\begin{array}{llll}
(6) & \forall p \in \mathbb{P}: & w \in \pi(p) & \textit{iff} \quad w' \in \pi'(p) \\
(7) & \forall \beta < \alpha \; \forall i \in \mathcal{A} \; \forall v \in W: & \textit{if} \quad wR_i v & \textit{then} \quad \exists v' \in W' : w'R'_i v' \wedge (v, v', \beta) \in \cong_{\mathbf{On}} \\
(8) & \forall \beta < \alpha \; \forall i \in \mathcal{A} \; \forall v' \in W': & \textit{if} \quad w'R'_i v' & \textit{then} \quad \exists v \in W : wR_i v \wedge (v, v', \beta) \in \cong_{\mathbf{On}}
\end{array}$$

*Two Kripke models $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ for the same set of agents $\mathcal{A}$ are α-*bisimilar *($M_1 \cong_\alpha M_2$) iff there are worlds $w \in W$, $w' \in W'$ and an ordinal bisimulation $\cong_{\mathbf{On}}$ such that $(w, w', \alpha) \in \cong_{\mathbf{On}}$.*
*Two Kripke worlds (M_1, w) and (M_2, w') are α-*bisimilar *($(M_1, w) \cong_\alpha (M_2, w')$) iff there is an ordinal bisimulation $\cong_{\mathbf{On}}$ such that $(w, w', \alpha) \in \cong_{\mathbf{On}}$.*

For $\alpha \leq \omega$, the intuitive idea underlying (4.5.8) is the following: A 0-bisimulation forces the worlds it connects to make the same elementary propositions and hence the same formula of `depth` 0 true. A 1-bisimulation between w and w' guarantees that for any world reachable from one of them in one step there is a world reachable in one step from the other which has the same valuation. It thus guarantees that w and w' validate the same modal formulae of `depth` 1. Similarly 2-bisimulation for reachability in two steps etc. A ω-bisimulation forces the worlds it connects to be such that for any world we may reach from one of them, we find a world reachable from the other making the same elementary propositions true. It is in this sense that ordinal bisimulations may be considered partial informational equivalences. We will later, in th. 4.5.23 on p. 116, state a general result validating these claims. Before doing so, however, we will study some other and connected notions.

First, we may also state the condition under which a relation is a bisimulation *tout court*:

Definition 4.5.9 (Bisimilarity). *Two Kripke models $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ for the same set of agents $\mathcal{A}$ are* bisimilar *($M_1 \cong M_2$) iff there is an ordinal bisimulation $\cong_{\mathbf{On}}$ and worlds $w \in W, w' \in W'$ such that $(w, w', \alpha) \in \cong_{\mathbf{On}}$ for every $\alpha \in \mathbf{On}$. Two Kripke worlds (M_1, w) and (M_2, w') are* bisimilar *($(M_1, w) \cong (M_2, w')$) iff there is an ordinal bisimulation $\cong_{\mathbf{On}}$ such that $(w, w', \alpha) \cong_{\mathbf{On}}$ for every ordinal $\alpha \in \mathbf{On}$.*

Clearly, the following holds:

Theorem 4.5.10. *Bisimilarity and α-bisimilarity are equivalence relations.*

Proof Reflexivity follows from the fact that for any Kripke model $M = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ the relation defined by $\{(w, w, \alpha) \mid w \in W\}$ is an ordinal bisimulation for every $\alpha \in \mathbf{On}$. Symmetry is a straightforward consequence of the symmetry of (7) and (8) in (4.5.8). For transitivity, we just choose, in both directions, twice a successor. $\square$

We may also note that partial bisimilarity between Kripke worlds is 'monotone' in the following sense:

Theorem 4.5.11. *Let $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ be two Kripke models for the same set of agents $\mathcal{A}$, $w \in W, w' \in W'$ and $\alpha, \beta \in \mathbf{On}$. If $\beta < \alpha$ and (M_1, w) and (M_2, w') are α-bisimilar, then they are β-bisimilar.*

Proof Suppose $(M_1, w) \cong_\alpha (M_2, w')$, i.e. there is an ordinal bisimulation $\cong_{\mathbf{On}}$ such that $(w, w', \alpha) \in \cong_{\mathbf{On}}$. We define another ordinal bisimulation $\cong^*_{\mathbf{On}} := \{(w, w', \gamma) \mid w \in W \wedge w' \in W' \wedge \gamma < \beta \wedge (w, w', \gamma) \in \cong_{\mathbf{On}}\} \cup \{(w, w', \beta)\}$. Clearly $(w, w', \beta) \in \cong^*_{\mathbf{On}}$. So we have to show that $\cong^*_{\mathbf{On}}$ is an ordinal bisimulation. Suppose that $(w, w', \zeta) \in \cong^*_{\mathbf{On}}$. It is clear that the first condition (6) in (4.5.8) is satisfied, because otherwise (M_1, w) and (M_2, w') would not be α-bisimilar in the first place. (7) is also satisfied. For suppose it is violated, i.e. that there is a $\gamma < \zeta$ (hence $\gamma < \beta$), an $i \in \mathcal{A}$, a $v \in W$ such that wR_iv but $\forall v' \in W'(w'R_i'v' \rightarrow (v, v', \gamma) \notin \cong^*_{\mathbf{On}})$. Because of $\gamma < \alpha$ whenever $\zeta < \alpha$, (M_1, w) and (M_2, w') then would not have been α-bisimilar in the first place. (8) is similar. $\square$

Another, and perhaps more perspicuous way to define partial bisimulations would be by way of the following operation on binary relations:

Definition 4.5.12. *Let $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R_i'\}_{i\in\mathcal{A}}\rangle$ be two Kripke models for the same set of agents $\mathcal{A}$. We define the operation $\mathcal{B}$ on binary relations on $W \times W'$ as follows. Let $S \subset W \times W'$ be any binary relation and $w \in W, w' \in W'$ two worlds. $(w, w') \in \mathcal{B}(S)$ iff $(w, w') \in S$ and, for all $i \in \mathcal{A}$, the following two conditions hold:*

$$\begin{array}{llll} \forall v \in W & \textit{if} \;\; wR_iv & \textit{then} & \exists v' \in W' : w'R_i'v' \wedge (v, v') \in S \\ \forall v' \in W' & \textit{if} \;\; w'R_i'v' & \textit{then} & \exists v \in W : wR_iv \wedge (v, v') \in S \end{array}$$

With the help of $\mathcal{B}$, we may now define the hierarchy of interest to us:

Definition 4.5.13 (Another definition of partial bisimilarity). *Let $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R_i'\}_{i\in\mathcal{A}}\rangle$ be two Kripke models for the same set of agents $\mathcal{A}$. For each ordinal $\alpha \in \mathbf{On}$, we define a relation $\cong^+_\alpha \subset W \times W'$ as follows:*

$$\begin{array}{rcl} \cong^+_0 & := & \{(w, w') \in W \times W' \mid \forall p \in \mathbb{P} : w \in \pi(p) \Longleftrightarrow w' \in \pi'(p)\} \\ \cong^+_{\alpha+1} & := & \mathcal{B}(\cong^+_\alpha) \\ \cong^+_\lambda & := & \bigcap_{\alpha<\lambda} \cong^+_\alpha \quad \textit{for a limit ordinal } \lambda \end{array}$$

Two Kripke models $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R_i'\}_{i\in\mathcal{A}}\rangle$ for the same set of agents $\mathcal{A}$ are α-bisimilar$^+$ *($M_1 \cong^+_\alpha M_2$) iff there are worlds $w \in W$, $w' \in W'$ such that $(w, w') \in \cong^+_\alpha$. Two Kripke worlds (M_1, w) and (M_2, w') are* α-bisimilar$^+$ *($(M_1, w) \cong^+_\alpha (M_2, w')$) iff $(w, w') \in \cong^+_\alpha$.*

In contrast to the α-bisimulations defined by (4.5.8), (4.5.13) defines, for any $\alpha \in \mathbf{On}$, an unique relation. The two definitions, however, give us the same relation of α-bisimilarity (both between models and worlds), a fact we will prove later as th. (4.5.16): whenever there is an α-bisimulation or other between two worlds in the sense of (4.5.8), these two worlds stand in the maximal α-bisimulation$^+$ relation defined by (4.5.13); the maximal bisimulation$^+$ $\cong^+_\alpha$ is an α-bisimulation in the sense of (4.5.8). Whenever there is a α-bisimulation in the sense of (4.5.8), on the other hand, we may construct a maximal one by (4.5.13). Let us also define the corresponding notion of bisimilarity$^+$ *tout court*:

Definition 4.5.14 (Another definition of bisimilarity). *Let $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ be two Kripke models for the same set of agents $\mathcal{A}$. We define a relation $\cong^+ \subset W \times W'$ as follows:*

$$\cong^+ \quad := \quad \bigcap_{\alpha\in\mathbf{On}} \cong^+_\alpha$$

Two Kripke models $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ for the same set of agents $\mathcal{A}$ are bisimilar$^+$ *($M_1 \cong^+ M_2$) iff there are worlds $w \in W$, $w' \in W'$ such that $(w, w') \in \cong^+$. Two Kripke worlds (M_1, w) and (M_2, w') are* bisimilar$^+$ *($(M_1, w) \cong^+ (M_2, w')$) iff $(w, w') \in \cong^+$.*

As the relations defined by (4.5.13) is unique, we have the following stronger version of (4.5.11):

Theorem 4.5.15. *Let $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ be two Kripke models for the same set of agents $\mathcal{A}$ and let $\alpha, \beta \in \mathbf{On}$. If $\beta < \alpha$, then $\cong^+_\alpha \subset \cong^+_\beta$.*

Proof Suppose that there are worlds $w, w' \in W$ such that $(w, w') \in \cong^+_\alpha$. We prove that $(w, w') \in \cong^+_\beta$ by induction on α. If $\alpha = 0$, then $\beta = 0$. But it follows from $\alpha = 0$ that $\forall p \in \mathbb{P}$, $w \in \pi(p)$ iff $w' \in \pi'(p)$ so we are done. The induction step follows from the fact that $\mathcal{B}$ defined by (4.5.12) is a downwards monotone operator ($\mathcal{B}(S) \subset S$ for every $S \subset W \times W'$). The case $\alpha = \lambda$ follows directly from the definition of $\cong^*_\lambda$. $\square$

However, the next theorem shows that the two definitions coincide as far as partial bisimilarity of Kripke worlds is concerned:

Theorem 4.5.16. *Let $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ be two Kripke models for the same set of agents $\mathcal{A}$.*

1. *M_1 and M_2 are α-bisimilar iff they are α-bisimilar$^+$.*
2. *M_1 and M_2 are bisimilar iff they are bisimilar$^+$.*

Proof

[1^{st} claim, $\Longrightarrow$:] We prove by induction on α: If $M_1 \cong_\alpha M_2$, i.e. if there are worlds $w \in W$, $w' \in W'$ and an ordinal bisimulation $\cong_{\mathbf{On}}$ such that $(w, w', \alpha) \in \cong_{\mathbf{On}}$, then (M_1, w)

and (M_2, w') are bisimilar$^+$ in the sense of (4.5.13). To show this, we show the more general result that (4.5.13) defines a relation which is maximal among the α-bisimulations defined by (4.5.8). We first define, for any α, a relation $\cong^*_\alpha$ on $W \times W'$ as follows:

$$(1) \quad (w, w') \in \cong^*_\alpha \quad :\Longleftrightarrow \quad \text{there is an ordinal bisimulation } \cong_{\mathbf{On}} \text{ s.t. } (w, w', \alpha) \in \cong_{\mathbf{On}}$$

We now prove, by induction on α, that $\cong^*_\alpha$ is contained in $\cong^+_\alpha$ (defined in the sense of def. 4.5.13), i.e. that, for every $\alpha \in \mathbf{On}$, $\cong^*_\alpha \subset \cong^+_\alpha$.

If $\alpha = 0$, this is guaranteed by the first clause (6) in (4.5.8). So suppose we have shown that $\cong^*_\alpha \subset \cong^+_\alpha$. We have to show that $\cong^*_{\alpha+1} \subset \mathcal{B}(\cong^+_\alpha)$. Take $(w, w') \in W \times W'$ such that $(w, w', \alpha+1) \in \cong_{\mathbf{On}}$ for some ordinal bisimulation $\cong_{\mathbf{On}}$ and suppose wR_iv. We have to show that there is a $v' \in W'$ such that $w'R'_iv'$ and $(v, v', \alpha) \in \cong_{\mathbf{On}}$. Because of $\alpha < \alpha+1$, this follows automatically from the second clause (7) of (4.5.8). The second clause in our definition of $\mathcal{B}$ (4.5.12) follows analogously by (8) of (4.5.8).

If $\alpha = \lambda$ for some limit ordinal λ, we have to show that $\cong^*_\lambda \subset \bigcap_{\kappa<\lambda} \cong^+_\kappa$. By (4.5.11), we have, for every $\kappa < \lambda$: if there is an ordinal bisimulation $\cong_{\mathbf{On}}$ such that $(w, w', \lambda) \in \cong_{\mathbf{On}}$, then there is an ordinal bisimulation $\cong^*_{\mathbf{On}}$ such that $(w, w', \kappa) \in \cong^*_{\mathbf{On}}$, that is $\cong^*_\lambda \subset \cong^*_\kappa$. By applying the induction hypothesis to $\cong^*_\kappa$, we get $\cong^*_\kappa \subset \cong^+_\kappa$ for all $\kappa < \lambda$ and hence $\cong^*_\lambda \subset \bigcap_{\kappa<\lambda} \cong^+_\kappa$.

[1^{st} claim, $\Longleftarrow$:] Suppose $\cong^+_\alpha$ is the unique α-bisimulation$^+$ between M_1 and M_2 in the sense of (4.5.13). We show that $M_1 \cong_\alpha M_2$ by showing, by induction on α, that the following relation $\cong_{\mathbf{On}} \subset W \times W' \times \mathbf{On}$ is an ordinal bisimulation in the sense of (4.5.8):

$$(2) \quad \cong_{\mathbf{On}} \quad := \quad \{(w, w', \alpha) \mid (w, w') \in \cong^+_\alpha\}$$

Suppose that $(w, w', \alpha) \in \cong_{\mathbf{On}}$. We have to verify that the three conditions in (4.5.8) hold and do so by induction on α.

The base case $\alpha = 0$ is clear. So suppose $\alpha = \gamma + 1$ is a successor ordinal. (6) is satisfied because of $\cong^+_\alpha \subset \cong^+_0$. For the second clause (7), suppose $(w, w', \alpha) \in \{(w, w', \alpha) \mid (w, w') \in \mathcal{B}(\cong^+_\gamma)\}$ and that there are $\beta < \alpha$, $i \in \mathcal{A}$ and $v \in W$ such that wR_iv. We have to show that there is a $v' \in W'$ such that $w'R'_iv'$ and $(v, v', \beta) \in \cong_{\mathbf{On}}$. We have by the first clause in the definition of $\mathcal{B}$ (4.5.12) that there is a $v' \in W'$ such that $w'R'_iv'$ and $(v, v') \in \cong^+_\gamma$. Because the sequence of relations defined by (4.5.13) is (at least weakly) decreasing and $\beta \leq \gamma$, we also have $(v, v') \in \cong^+_\beta$, i.e. $(v, v', \beta) \in \cong_{\mathbf{On}}$. (8) is analogous.

If $\alpha = \lambda$ is a limit ordinal, we suppose $w \cong^+_\lambda w'$ and that there are $\beta < \lambda$, $i \in \mathcal{A}$

and $v \in W$ such that wR_iv; we then have to show that there is a $v' \in W'$ such that $w'R_i'v'$ and $(v, v', \beta) \in \cong_{\mathbf{On}}$. By (4.5.15), it follows from $w \cong_\lambda^+ w'$ that $w \cong_\kappa^+ w'$ for every $\kappa < \lambda$ and, because of $\beta < \lambda$, we get the desired result. (6) follows as before and (8) is again analogous.

[2^{nd} claim, $\Longrightarrow$:] Suppose M_1 and M_2 are bisimilar in the sense of (4.5.9), i.e. that there is an ordinal bisimulation $\cong_{\mathbf{On}}$ and worlds $w \in W, w' \in W'$ such that $(w, w', \alpha) \in \cong_{\mathbf{On}}$ for every $\alpha \in \mathbf{On}$. We define a relation $\cong^*$ as follows:

$$(3) \quad (w, w') \in \cong^* \quad :\Longleftrightarrow \quad \forall \alpha \in \mathbf{On} \,:\, (w, w', \alpha) \in \cong_{\mathbf{On}}$$

We now show that $\cong^* \subset \bigcap_{\alpha \in \mathbf{On}} \cong_\alpha^+$, where the relations $\cong_\alpha^+$ are defined as in (4.5.13). Suppose $(w, w') \in \cong^*$ and take any arbitrary $\alpha \in \mathbf{On}$. We have to show that $(w, w') \in \cong_\alpha^+$. This follows from $(w, w', \alpha) \in \cong_{\mathbf{On}}$ and the inclusion we showed in the proof of the left-to-right direction of the first claim.

[2^{nd} claim, $\Longleftarrow$:] Suppose $\cong^+$ is the unique bisimulation$^+$ between M_1 and M_2 in the sense of (4.5.14). We show that $M_1 \cong M_2$ by observing that $\{(w, w', \alpha) \mid (w, w') \in \cong^+ \wedge \alpha \in \mathbf{On}\}$ is an ordinal bisimulation between M_1 and M_2 which is such that $(w, w', \alpha) \in \{(w, w', \alpha) \mid (w, w') \in \cong^+ \wedge \alpha \in \mathbf{On}\}$ for every $\alpha \in \mathbf{On}$. $\square$

Normally, bisimilarity *tout court* is defined in yet another, more direct, way:

Definition 4.5.17 (Yet another definition of bisimilarity). *Let $M_1 = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}} \rangle$ and $M_2 = \langle W', \pi', \{R_i'\}_{i \in \mathcal{A}} \rangle$ be two Kripke models for the same set of agents $\mathcal{A}$. A nonempty relation $\cong^* \subset W \times W'$ is a* bisimulation* *iff, for for all $w \in W, w' \in W'$, the following three condition hold if $(w, w') \in \cong^*$:*

$$
\begin{array}{llll}
(9) & \forall p \in \mathbb{P} : & w \in \pi(p) & \textit{iff} \quad w' \in \pi'(p) \\
(10) & \forall i \in \mathcal{A}\, \forall v \in W : & \textit{if} \quad wR_iv & \textit{then} \quad \exists v' \in W' \,:\, (w'R_i'v' \wedge v \cong^* v') \\
(11) & \forall i \in \mathcal{A}\, \forall v' \in W' : & \textit{if} \quad w'R_i'v' & \textit{then} \quad \exists v \in W \,:\, (wR_iv \wedge v \cong^* v')
\end{array}
$$

Two Kripke models $M_1 = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}} \rangle$ and $M_2 = \langle W', \pi', \{R_i'\}_{i \in \mathcal{A}} \rangle$ for the same set of agents $\mathcal{A}$ are bisimilar* *($M_1 \cong^* M_2$) iff there are worlds $w \in W$, $w' \in W'$ and a bisimulation* $\cong^* \subset W \times W'$ such that $w \cong^* w'$. Two Kripke worlds (M_1, w) and (M_2, w') are* bisimilar* *($(M_1, w) \cong^* (M_2, w)$) iff there is a bisimulation* $\cong^* \subset W \times W'$ such that $w \cong^* w'$.*

We now show that our notion of bisimulation as the limit case of partial bisimulations (4.5.9 and 4.5.14) coincides with this standard notion of bisimilarity.

Theorem 4.5.18. *Let $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ be two Kripke models for the same set of agents $\mathcal{A}$ and $w\in W, w'\in W'$. Then we have:*

w and w' are bisimilar iff w and w' are bisimilar.*

Proof

[$\Longrightarrow$:] Suppose that (M_1, w) and (M_2, w') are bisimilar, that is, that there is an ordinal bisimulation $\cong_{\mathbf{On}}$ such that $(w, w', \alpha) \in\ \cong_{\mathbf{On}}$ for every ordinal $\alpha\in\mathbf{On}$. We show that the following relation $\cong^*$ is a bisimulation* (this is just (3) of the proof of th. 4.5.16 above):

$$(w, w') \in\ \cong^* \quad :\Longleftrightarrow \quad \forall\alpha\in\mathbf{On} : (w, w', \alpha) \in\ \cong_{\mathbf{On}}$$

The first clause (9) is clearly satisfied. For the second, suppose wR_iv. If there were no $v'\in W'$ such that $(v, v') \notin\ \cong$ then there were an ordinal $\beta\in\mathbf{On}$ such that $(v, v', \beta) \notin\ \cong_{\mathbf{On}}$. But then (M_1, w) and (M_2, w') would not be $\beta + 1$-bisimilar. So there is no such $v'\in W'$ and the second condition (10) is satisfied. The third clause (11) is similar.

[$\Longleftarrow$:] Suppose (M_1, w) and (M_2, w') are bisimilar*, i.e. that there is a bisimulation* $\cong^*$ such that $w \cong^* w'$. We have to show that $\cong^* \times\ \mathbf{On} = \{(w, w', \alpha) \mid (w, w') \in\ \cong^* \wedge\ \alpha \in \mathbf{On}\}$ is an ordinal bisimulation and that $(w, w', \alpha) \in\ \cong^* \times\ \mathbf{On}$ for every $\alpha \in \mathbf{On}$ (cf. def. 4.5.9). While the second claim ($(w, w', \alpha) \in\ \{(w, w', \alpha) \mid (w, w') \in\ \cong^*\ \wedge \alpha \in \mathbf{On}\}\ \forall\alpha\in\mathbf{On}$) is true by construction, we show the first claim as follows. By the first clause (9) in (4.5.17), $\cong^* \times\ \mathbf{On}$ satisfies the first condition (6) of (4.5.8). For the second clause (7), suppose $(w, w', \alpha) \in\ \cong^* \times\ \mathbf{On}$ and that there are $\beta < \alpha$, $i\in\mathcal{A}$ and $v\in W$ such that wR_iv. We have to show that there is a $v'\in W'$ such that $w'R'_iv'$ and $(v, v', \beta) \in\ \cong^* \times\ \mathbf{On}$. But we know that there is a $v'\in W'$ such that $w'R'_iv'$ and $v \cong^* v'$. The third clause (8) is similar. □

The approximations by partial bisimulations are proper because we can, for any α, find two Kripke worlds which are α-bisimilar but not bisimilar *tout court.*

Theorem 4.5.19. *For every $\alpha\in\mathbf{On}$, there are Kripke worlds (M_1, w) and (M_2, w') such that $(M_1, w) \cong_\alpha (M_2, w')$ but $(M_1, w) \ncong (M_2, w')$.*

Proof We show how to construct a Kripke model $M_\alpha = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ for a given ordinal $\alpha\in\mathbf{On}$ (we will later see, in def. 4.6.3, that such a Kripke model is a "picture" of the ordinal α): We first choose $\alpha + 1 = \alpha \cup \{\alpha\}$ as our set of worlds and set $\pi = \pi_\alpha : \mathbb{P} \to \mathcal{P}(\alpha)$, for every $p \in \mathbb{P}$, $\pi(p) := \emptyset$. For the relation R between any two $\beta, \gamma \in \{\alpha\}$ we take the converse of membership: $\beta R \gamma :\Leftrightarrow \gamma \in \beta$. So $M_\alpha = \langle \{\alpha\}, \pi_\alpha, \in^{-1} \restriction \{\alpha\} \times \{\alpha\} \rangle$.

We will prove the following two claims:

$$(1) \quad \alpha \leq \zeta \ \rightarrow \ M_\alpha \cong_\alpha M_\zeta$$

$$(2) \quad \alpha > \zeta \ \rightarrow \ M_\alpha \not\cong_\alpha M_\zeta$$

From (1) and (2) it follows that the two Kripke models M_α and $M_{\alpha+1}$ are α-bisimilar, but not $(\alpha+1)$-bisimilar, and hence, by (4.5.9), not bisimilar *tout court*.

We prove (1) by induction on α, i.e. we show, for every α, that the following relation $\cong_{\mathbf{On}} \subset \ W \times W' \times \mathbf{On}$ is an ordinal bisimulation in the sense of (4.5.8) (where $\cong^+_\alpha$ is to be understood in the sense of def. 4.5.13):

$$(3) \quad \cong_{\mathbf{On}} \ := \ \{(w, w', \alpha) \mid (w, w') \in \cong^+_\alpha\}$$

Suppose that $(w, w', \alpha) \in \cong_{\mathbf{On}}$. We have to verify that the three conditions in (4.5.8) hold and do so by induction on α.

The base case $\alpha = 0$ is clear, for 0-bisimilarity just requires sameness of valuation. So suppose we have proven (1) for all $\kappa < \alpha$. We show that $(\alpha, \zeta, \alpha) \in \cong_{\mathbf{On}}$ and hence that $(M_\alpha, \alpha) \cong_\alpha (M_\zeta, \zeta)$. The first bisimulation clause (6) in (4.5.8) is clearly satisfied, given that our definition of π_ζ did not depend on ζ (π_α will just be the restriction of π_ζ to α). The forth condition (7) is immediate because of $\alpha \subset \zeta$. For the back condition (8), suppose $\beta < \alpha$ and $\zeta R' \gamma$ (i.e. $\gamma \in \zeta$) for some $\gamma \in \{\zeta\}$. We have to show that there is a $\delta \in \{\alpha\}$ such that $\alpha R \delta$ (i.e. $\delta \in \alpha$) and $\gamma \cong_\beta \delta$ ($(\gamma, \delta, \beta) \in \cong_{\mathbf{On}}$).

Because of $\gamma \in \{\zeta\}$, we have $\gamma \leq \zeta$. We distinguish two cases:

1. $\gamma < \beta$. Then $\gamma \in \alpha$ and we choose $\delta = \gamma$ to get $\delta \cong_\beta \gamma$ ($(\delta, \gamma, \beta) \in \cong_{\mathbf{On}}$), which is $\gamma \cong_\beta \gamma$ ($(\gamma, \gamma, \beta) \in \cong_{\mathbf{On}}$).
2. $\gamma \geq \beta$. Then we choose $\beta \in \alpha$ as our δ and apply the induction hypothesis to $\beta = \delta < \alpha$ to get $\beta \cong_\beta \gamma$ ($(\beta, \gamma, \beta) \in \cong_{\mathbf{On}}$) which is $\delta \cong_\beta \gamma$ ($(\delta, \gamma, \beta) \in \cong_{\mathbf{On}}$).

We prove (2) by induction on α. In the base case $a = 0$ we have nothing to prove. So let α be given, take any $\zeta \in \mathbf{On}$ such that $\alpha > \zeta$ and assume $(M_\alpha, \beta) \cong_\alpha (M_\zeta, \gamma)$ (i.e. that there is an ordinal bisimulation $\cong_{\mathbf{On}}$ such that $(\beta, \gamma, \alpha) \in \cong_{\mathbf{On}}$) for some $\beta \in \{\alpha\}$ and some $\gamma \in \{\zeta\}$. By $\zeta < \alpha$. We distinguish two cases:

1. $\zeta < \beta$. Then we have $\zeta < \alpha$, $\zeta \in \{\alpha\}$ and $\beta R \zeta$. By the forth clause (7) of (4.5.8), it

follows that there is a $\delta \in \{\zeta\}$ such that $\gamma R' \delta$ ($\delta \in \gamma$) and $\delta \cong_\zeta \zeta$ ($(\delta, \zeta, \zeta) \in \cong_{\mathbf{On}}$). $\delta \in \gamma$ implies $\delta < \zeta$, however, while $\delta \cong_\zeta \zeta$ implies $\delta \geq \zeta$ by the induction hypothesis. So we have a contradiction.

2. $\zeta \geq \beta$. Then we have $\beta < \alpha$, $\beta \in \{\zeta\}$ and $\gamma R' \beta$. By the back clause (8) of (4.5.8), it follows that there is a $\epsilon \in \{\alpha\}$ such that $\beta R \epsilon$ ($\epsilon \in \beta$) and $\epsilon \cong_\beta \beta$ ($(\epsilon, \beta, \beta) \in \cong_{\mathbf{On}}$). $\epsilon \in \beta$ implies $\epsilon < \beta$, however, while $\epsilon \cong_\beta \beta$ implies $\epsilon \geq \beta$ by the induction hypothesis. So we have a contradiction.

$\square$

Using the `depth` function defined above, we may define a notion of partial indistinguishability: Whenever two Kripke worlds cannot be distinguished by sentences of modal depth $\leq \alpha$, we say that they are indistinguishable to `depth` α.

Definition 4.5.20 (Indistinguishability wrt languages). *Two Kripke worlds (M_1, w) and (M_2, w') are* indistinguishable to depth α, *written $(M_1, w) \equiv_\alpha (M_2, w')$, iff, for all $\phi \in \mathcal{L}_\alpha$, $(M_1, w) \models \phi \iff (M_2, w') \models \phi$, i.e. iff (M_1, w) and (M_2, w') satisfy the same formulae of $\mathcal{L}_\alpha$. If (M_1, w) and (M_2, w') satisfy the same formulae of $\mathcal{L}$, they are called* finitely indistinguishable, *written $(M_1, w) \equiv_\mathcal{L} (M_2, w')$. If (M_1, w) and (M_2, w') satisfy the same formulae of $\mathcal{L}_\infty$, they are called* indistinguishable, *written $(M_1, w) \equiv_\infty (M_2, w')$.*

(4.5.20) may be seen to be a special case of a more general notion of indistinguishability with respect to a set of formulae, which we will later need to define filtrations of models. To arrive at this notion, we first have to define a constraint on sets of formulae by which we filter:

Definition 4.5.21 (Subformulae closed sets). *A set of formulae $\Sigma \subset \mathcal{L}_\infty$ is* closed under subformulae *iff the following conditions hold for all subsets $\Phi \subset \Sigma$ and for all formulae $\phi \in \Sigma$:*

- *if $\neg\phi \in \Sigma$, then $\phi \in \Sigma$*
- *if $\phi \wedge \psi \in \Sigma$, then $\phi, \psi \in \Sigma$*
- *if $\bigwedge \Phi \in \Sigma$, then $\psi \in \Sigma$ for all $\psi \in \Phi$*
- *if $\Box\phi \in \Sigma$, then $\phi \in \Sigma$*

It is clear that $\mathcal{L}_\alpha$, for any $\alpha \in \mathbf{On}$, is closed under subformulae, hence that the following definition is a proper generalisation of (4.5.20):

Definition 4.5.22 (Indistinguishability wrt formulae). *If $\Sigma \subset \mathcal{L}_\infty$ is a set of formulae closed under subformulae, two Kripke worlds (M_1, w) and (M_2, w') are* indistinguishable with respect to Σ *($(M_1, w) \equiv_\Sigma (M_2, w')$) iff, for all $\phi \in \Sigma$, $(M_1, w) \models \phi \Leftrightarrow (M_2, w') \models \phi$.*

$\equiv_\Sigma$ is clearly an equivalence relation and we denote its equivalence classes by "$|w|_\Sigma$".

The importance of partial bisimulation derives from the following theorem, which states that the indistinguishability with respect to $\mathcal{L}_\alpha$ formulae corresponds exactly to α-bisimulation:

Theorem 4.5.23. *Let $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ be two Kripke models for the same set of agents $\mathcal{A}$. Then the following holds for every $w\in W, w'\in W'$:*

$$(M_1, w) \cong_\alpha (M_2, w') \quad \Longleftrightarrow \quad (M_1, w) \equiv_\alpha (M_2, w')$$

PROOF

[$\Longrightarrow$:] Let ϕ be a sentence of $\mathcal{L}_\infty$. We prove by induction on ϕ that: if $\mathtt{depth}(\phi) = \alpha$, $w \models \phi$ and $(M_1, w) \cong_\alpha (M_2, w')$ (i.e. there is an ordinal bisimulation $\cong_{\mathbf{On}}$ such that $(w, w', \alpha) \in \cong_{\mathbf{On}}$), then $w' \models \phi$.

- Let $\phi\in\mathbb{P}$. Then $w\models\phi \leftrightarrow w'\models\phi$ by definition of $\cong_{\mathbf{On}}$.
- Let $\phi = \neg\psi$. Then $w\models\psi \leftrightarrow w'\models\psi$ by induction hypothesis and $w\models\phi \leftrightarrow w'\models\phi$ by definition of $\neg$.
- Let $\phi = \bigwedge\Phi$. Then $w\models\psi$ for all $\psi\in\Phi$ and thus $w'\models\psi$ for all $\psi\in\Phi$ by induction hypothesis. This implies $w'\models\phi$ by definition of $\bigwedge$.
- Let $\phi = [a]\psi$ for some $a\in\mathcal{A}$. So $\mathtt{depth}([a]\psi) = \alpha$ is a successor ordinal $\alpha = \beta+1$. Suppose $w'R_a v'$. By the third bisimulation clause (8), there is a $v \in W$ such that $wR_a v$ and $(v, v', \beta) \in \cong_{\mathbf{On}}$. From $w \models [a]\psi$ it follows that for all $v \in W$ such that $wR_a v$ it holds that $v\models\psi$. From this and the induction hypothesis for ϕ we conclude that $v'\models\psi$ and hence, because v' was an arbitrary a-successor, $w'\models\phi$.

[$\Longleftarrow$:] We show that the following defines an ordinal bisimulation:

$$(1) \quad \cong_{\mathbf{On}} \;:=\; \{(w, w', \alpha) \mid (M_1, w) \equiv_\alpha (M_2, w')\}$$

Suppose $(w, w', \alpha) \in\cong_{\mathbf{On}}$. We have to verify that the three conditions in (4.5.8) hold.

Because indistinguishability requires sameness of valuation, the first bisimulation clause (6) is clearly fulfilled. For the second (7), take any $\beta < \alpha$, any $v \in W$ and any $a\in\mathcal{A}$ such that $wR_a v$. Assume that there is no $v'\in W'$ such that $w'R'_a v'$ and $(v, v', \beta) \in \cong_{\mathbf{On}}$. This means that for every $v' \in W'$ with $w'R'_a v'$ the following holds: $(M_1, v) \not\equiv_\beta (M_2, v')$, i.e. there is a $\mathcal{L}_\beta$-sentence $\phi_{v'}$ such that $v \models \phi_{v'}$ but $v' \not\models \phi_{v'}$. But then $w \models \langle a\rangle \bigwedge\{\phi_{v'} \mid w'R'_a v'\}$, but $w' \not\models \langle a\rangle \bigwedge\{\phi_{v'} \mid w'R'_a v'\}$. Because $\beta < \alpha$ "$\langle a\rangle \bigwedge\{\phi_{v'} \mid w'R'_a v'\}$"

is at most a $\mathcal{L}_\alpha$-sentence. This, however, contradicts our assumption that $(M_1, w) \equiv_\alpha (M_2, w')$. The third clause (8) is similar. □

We note, as an important special case, a result due to van Benthem (Benthem 1991: 211): For any $\alpha \in \mathbf{On}$, $\mathcal{L}_\alpha$-formulae do not distinguish between bisimilar models.

Theorem 4.5.24. *Let $M_1 = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i \in \mathcal{A}}\rangle$ be two Kripke models for the same set of agents $\mathcal{A}$. Then the following holds for every $w \in W, w' \in W'$:*

$$(M_1, w) \cong (M_2, w') \qquad \Longrightarrow \qquad (M_1, w) \equiv_\infty (M_2, w')$$

PROOF The proof is the same as that of the [$\Longrightarrow$] direction in (4.5.23). □

The converse of (4.5.24), however, holds only under additional assumptions either on α or on the Kripke worlds: either the language has to be infinitary or else the models have to be image-finite. A Kripke model $M = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}}\rangle$ is *image-finite* if for any for any world $w \in W$ and any relation R_i the set $\{v \mid wR_iv\}$ is finite. A Kripke world is image-finite iff is is based on a image-finite Kripke model.

Theorem 4.5.25. *If (M_1, w) and (M_2, w') are two image-finite Kripke worlds, the following holds:*

$$(M_1, w) \equiv_{\mathcal{L}} (M_2, w') \qquad \Longrightarrow \qquad (M_1, w) \cong (M_2, w')$$

PROOF Image-finiteness just means that the conjunction we construct in the proof of the [$\Longleftarrow$] direction of (4.5.23) is finite and hence that the sentence so constructed belongs to $\mathcal{L}$. □

4.6 Epistemic states

As we have seen in sct. 4.4, the intuitive idea behind using Kripke models in epistemic logic is that every possible world $w \in W$ of a model M corresponds to a possible *epistemic state* of an agent $a \in \mathcal{A}$. This epistemic state is characterised by the truth-values the agent assigns to the propositional variables $p \in \mathbb{P}$ by taking himself to be in w and the *information state* he believes himself to be in, i.e. the set $\{x \in W \mid wR_ax\}$ of those worlds which are epistemic alternatives for a in w, that is worlds that, as far as a can tell, could be what he takes to be his actual world.

We will later see (with th. 4.6.13) that, given our infinitary language, epistemic states are uniquely characterised not only by bisimulation classes of Kripke models but also by single sentences.

Following Barwise and Moss, we will characterise the way a Kripke models represents epistemic states (by the value of the propositional variables and the information states of the agents) as sets which are allowed to contain themselves as elements and therefore do not satisfy the foundation axiom of ordinary ZFC set theory. We therefore have to introduce some machinery, namely *graphs*, *decorations* and *hypersets*.

Definition 4.6.1 (Graphs). *A* graph $\mathcal{G}$ *is a pair* $(G, \rightarrow)$ *consisting of a set of nodes* G *and a binary relation* $\rightarrow$ *over* G. *A* pointed graph $(\mathcal{G}, n)$ *is the pair of a graph* $\mathcal{G} = (G, \rightarrow)$ *and a distinguished node* $n \in G$.

Graphs and pointed graphs are just mono-modal Kripke frames and their worlds. To cover the multi-modal case, we introduce labelled graphs as follows:

Definition 4.6.2 (Labelled Graphs). *Given some set of agents* $\mathcal{A}$, *a* labelled graph $\mathcal{G}$ *is a pair* $(G, \{\xrightarrow{a}\}_{a\in\mathcal{A}})$ *consisting of a set of nodes* G *and a family of binary relations* $\xrightarrow{a}$ *over* G *indexed by* $\mathcal{A}$. *A* pointed labelled graph $(\mathcal{G}, n)$ *is the pair of a labelled graph* $(G, \{\xrightarrow{a}\}_{a\in\mathcal{A}})$ *and a distinguished node* $n \in G$.

Labelled graphs with a finite family of relations and a nonempty set of nodes are just Kripke frames (4.5.1). In order to introduce hypersets, which may contain themselves as members, we introduce the notion of a decoration of a graph (for decorations of labelled graphs, see def. 4.6.9 below).

Definitions 4.6.3 (Decorations, solutions, pictures). *A* decoration *of a graph* $\mathcal{G}$ *is a function* δ *that assigns to each node* n *of* $\mathcal{G}$ *a set* δ_n[14] *such that* $\delta_n = \{\delta_{n'} \mid n \rightarrow_{\mathcal{G}} n'\}$. *If* δ *is a decoration of* $\mathcal{G}$, *any pointed graph* $(\mathcal{G}, n)$ *(for* $n \in G$*) is called a* picture *of the respective set* δ_n, *assigned to* n *by* δ, *and* δ_n *is called a* solution *of* $(\mathcal{G}, n)$.

It can be proved within ordinary Zermelo-Fraenkel set theory ZFC (without making use of the axiom of foundation) that every set has a picture, i.e. that for every set there is a graph which has that set as the decoration of one of its nodes. These graphs are all *well-founded*, i.e. they do not contain an infinite path $n \rightarrow n_1 \rightarrow n_2 \rightarrow \ldots$. The foundation axiom is in ZFC equivalent to the following statement:

[14] "Set" has to be taken with a grain of salt: in the case of non-wellfounded graphs, some δ_n will be hypersets.

4.6.4 (Foundation axiom). *Only well-founded graphs have decorations.*

In order to get hypersets, i.e. sets which are allowed to contain themselves as elements, we drop the foundation axiom from ZFC and replace it by the following:

4.6.5 (Anti-foundation axiom (AFA)). *Every graph has exactly one decoration.*

We will call anything that satisfies ZFC(AFA) (the ZFC axioms minus the axiom of foundation, plus the anti-foundation axiom) and that is built up from our set of urelements $\mathbb{P}$ a *hyperset.*

Hypersets will allow us to encode the relational structure of a Kripke model into its possible worlds. To do this, some preliminary work is needed.

Definition 4.6.6. *Let $\mathcal{A}$ be a set of agents, $\mathbb{P}$ a set of propositional variables and s any hyperset. Call a function f an* epistemic s-valuation *which assigns to each propositional variable $p \in \mathbb{P}$ a truth-value ($\perp$ or $\top$) and to each agent $a \in \mathcal{A}$ some subset $t \in \mathcal{P}(s)$ of s. Consider now the following operator on hypersets:*

(12) $\Phi(s) \;:=\; \{f \mid f \textit{ is an epistemic } s\textit{-valuation}\}$

This operator is clearly monotone (i.e. $s \subset t$ implies $\Phi(s) \subset \Phi(t)$) and hence we know – by Barwise and Moss (1996: 216) – that it has a greatest fixed point. We are now in a position to define *epistemic states*:

Definition 4.6.7 (Epistemic states). *Let $\mathcal{A}$ be a set of agents and $\mathbb{P}$ a set of propositional variables. The class of* epistemic states P *for $\mathcal{A}$ relative to $\mathbb{P}$ is the greatest fixed point of the operator Φ defined by 12.*

We may now characterise epistemic states in somehow more intuitive terms:

Theorem 4.6.8. *The class of* epistemic states P *for $\mathcal{A}$ relative to $\mathbb{P}$ is the largest class such that:*

- *An epistemic state $w \in \mathsf{P}$ is a function that assigns to each $p \in \mathbb{P}$ a truth-value $w(p) \in \{\top, \perp\}$ and to each $a \in \mathcal{A}$ an information state $w(a)$.*
- *An information state σ is a hyperset of epistemic states.*

Proof By Barwise's and Moss's representation theorem for greatest fixed points (1996: 235) and by the fact that P is uniform (cf. Barwise and Moss (1996: 230–233) for a definition), we know that P is the union of all solution-sets of flat P-coalgebras. A flat P coalgebra is a pair $\Xi = \langle b, e\rangle$, where b is a set of urelements new for P[15] and e is a function $e : b \to \mathsf{P}(b)$. We now have to show that each function w assigning to each $p \in \mathbb{P}$ a truth-value $w(p) \in \{\top, \bot\}$ and to each $a \in \mathcal{A}$ an hyperset $w(a)$ of epistemic states is in the solution-set of some flat P-coalgebra. So let some such function w be given. We choose some set of urelements $\mathcal{X}$ new for P and define the following set of equations: for all $p \in \mathbb{P}$, include $w(p) = i$ (for $\in \{\top, \bot\}$) and for each $a \in \mathcal{A}$, include exactly one equation of the form $w(a) = \sigma$ for $\sigma \subset \mathcal{X}$. Such a system of equations can be turned into a flat coalgebra. By the Solution Lemma Lemma (Barwise and Moss 1996: 236) each such flat coalgebra has a unique solution and its solution set is contained in its largest fixed point. $\square$

We are now in a position to decorate Kripke models with epistemic states:

Definition 4.6.9. *If $M = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ is a Kripke model with respect to a set of propositional variables $\mathbb{P}$ and a set of agents $\mathcal{A}$, a function δ that assigns to each world $v \in W$ an epistemic state δ_v is a* decoration *of M iff it holds for each such world v that:*

$$\begin{array}{lccc} \forall p \in \mathbb{P}: & \delta_v(p) = \top & \longleftrightarrow & v \in \pi(p) \\ \forall a \in \mathcal{A}: & \delta_v(a) & = \{\delta_z \mid vR_a z\} & \end{array}$$

If δ is a decoration of M, δ_w is called the solution *of the corresponding Kripke world (M, w) and (M, w) is then a* picture *of δ_w.*

We now have the following correspondence between Kripke worlds and epistemic states:

Theorem 4.6.10. *Every Kripke world has an epistemic state as its unique solution. Each epistemic state is pictured by a Kripke world.*

PROOF

[1^{st} claim:] A Kripke model $M = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ is a labelled graph. Labelled graphs have unique decorations, for they can be uniquely characterised as systems of equations by taking the nodes $w \in W$ as the indeterminates and including the following equations: for all $p \in \mathbb{P}$, include $w(p) = \top$ if $w \in \pi(p)$ and $w(p) = \bot$ otherwise. For all $a \in \mathcal{A}$, include $w(a) = \{v \mid wR_a v\}$. By the solution lemma, provable in ZFC(AFA) (cf. Barwise and Moss 1996: 72), this system of equations has a unique solution, i.e. there is a unique function

[15] Sets of urelements disjoint from both $\mathcal{A}$ and $\mathbb{P}$ are automatically new for P.

δ that assigns to each $w \in W$ a hyperset δ_w such that, for all $p \in \mathbb{P}$, $\delta_w(p) = i$, whenever $w(p) = i$ $(i \in \{\top, \bot\})$, and for all $a \in \mathcal{A}$, $\delta_w(a) = \{d_v \mid v \in s_w\}$ whenever $w(a) = s_w$. Decorations of Kripke models assign unique epistemic states to the worlds in the model, so every Kripke world gets assigned exactly one epistemic state.

[2^{nd} claim:] Given an epistemic state δ_w, we construct the underlying Kripke model $M = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}} \rangle$ as follows: For W, we take all the elements in the R_a chains of δ_w (for all $a \in \mathcal{A}$). To do this, we define an auxiliary function f_δ which takes a hyperset of epistemic states C and gives us the hyperset of their successors:

$$f_\delta(\mathsf{C}) := \bigcup_{w \in \mathsf{C}} \bigcup_{i \in \mathcal{A}} \delta_w(i)$$

f_δ is a monotone operator on epistemic states and W is its least fixpoint. We then have an epistemic state δ_v for any $v \in W$. The relations R_a (for $a \in \mathcal{A}$) are defined by: $(u, v) \in R_a :\Leftrightarrow v \in \delta_u(a)$, whereas the valuation π is given by $u \in \pi(p)$ iff $\delta_u(p) = \top$. □

Whereas every Kripke world pictures a unique epistemic state, the Kripke worlds picturing a given epistemic state are not identical but only bisimilar.[16] We can thus regard epistemic states as representatives of equivalence classes of Kripke worlds under bisimulation:

Theorem 4.6.11. *Two Kripke worlds are pictures of the same epistemic state iff they are bisimilar.*

Proof (M_1, w) and (M_2, w') are pictures of the same epistemic state iff they validate the same sentences: $(M_1, w) \equiv_\infty (M_2, w')$. The claim thus follows from (4.5.23). □

By establishing a correspondence between epistemic states which are non-wellfounded sets and (bisimilarity classes of) Kripke worlds, (4.6.11) allows us to 'import' the following theorem (4.6.12), which is some kind of converse to (4.5.17): Not only is there a 'characteristic' Kripke world for each ordinal but also every Kripke world is uniquely characterised by some ordinal. Barwise and Moss prove that every set in ZFC(AFA) is characterised by a sentence of the infinitary language. We will now extend this result to Kripke models:

Theorem 4.6.12. *For each Kripke world (M_1, w), there is an α which is the smallest ordinal such that for all Kripke worlds (M_2, w'):*

$$(M_1, w) \cong_\alpha (M_2, w') \quad \Longleftrightarrow \quad (M_1, w) \cong (M_2, w')$$

[16]This is easily seen by noticing that the Kripke model we get by this method of 'unfolding' δ_w will be a submodel generated by w (in the sense of def. 4.7.1 below).

PROOF We show, first, that for any Kripke model M_1 there is an ordinal β which characterises that any world in that modal up to bisimilarity *within the underlying model*, i.e. that there is a $\beta \in \mathbf{On}$ such that, for any $v, v' \in W$:

$$(1) \quad (M_1, v) \cong_\beta (M_1, v') \quad \Longleftrightarrow \quad (M_1, v) \cong (M_1, v')$$

Suppose the contrary. This would mean that we find, for every $\beta \in \mathbf{On}$ two worlds $v, v' \in W$ such that $(M_1, v) \cong_\beta (M_1, v')$ but $(M_1, v) \not\cong (M_1, v')$, i.e. such that there is a $\gamma \in \mathbf{On}$ ($\gamma > \beta$ by (4.5.11)) such that $v \not\cong_\gamma v'$. Because the identity relation is a 0-bisimulation and ordinal bisimilarity is transitive, this would mean that we find, for any $\beta \in \mathbf{On}$ at least one *new* world. W, however, was stipulated (in def. 4.5.1) to be a set, while $\mathbf{On}$ is not a set. So we have a contradiction.

Let then β be the ordinal characterising (M_1, w) 'internally' in the sense of (1). Let α be the smallest limit ordinal $> \beta$. We will show that for all Kripke worlds (M_2, w') the following holds:

$$(2) \quad (M_1, w) \cong_\alpha (M_2, w') \quad \Longleftrightarrow \quad (M_1, w) \cong (M_2, w')$$

For the left-to-right direction, suppose $(M_1, w) \cong_\alpha (M_2, w')$, i.e. that there is an ordinal bisimulation $\cong_{\mathbf{On}}$ such that $(w, w', \alpha) \in \cong_{\mathbf{On}}$. Define, for every α, the following relation $\cong_\alpha$ on $W \times W'$:

$$(3) \quad (w, w') \in \cong_\alpha \;:\Longleftrightarrow\; \text{there is an ordinal bisimulation } \cong^*_{\mathbf{On}} \text{ s.t. } (w, w', \alpha) \in \cong^*_{\mathbf{On}}$$

We show that $\cong_\alpha$ is a bisimulation in the sense of def. 4.5.17. (9) is clear.

For (10), choose any $a \in \mathcal{A}, v \in W$ such that $wR_a v$. So there has to be a $v' \in W'$ such that $vR'_a v'$ and $(v, v', \beta) \in \cong^*_{\mathbf{On}}$. Because α is a limit ordinal, we show that $(v, v', \alpha) \in \cong^*_{\mathbf{On}}$ and hence that $(M_1, v) \cong_\alpha (M_2, v')$ by showing that $(v, v', \gamma) \in \cong^*_{\mathbf{On}}$ for all $\gamma < \alpha$. This is clear for all $\gamma \leq \beta$. So choose a $\gamma \in \mathbf{On}$ such that $\beta < \gamma < \alpha$. Since $(w, w', \alpha) \in \cong^*_{\mathbf{On}}$ and $w'R'_a v'$, there is a $u \in W$ such that $(M_1, u) \cong_\gamma (M_2, v')$. By $\beta < \gamma$ and (4.5.11), this implies $(M_1, u) \cong_\beta (M_2, v')$ and by the transitivity of $\cong_\beta$, $(M_1, u) \cong_\beta (M_1, v)$. By our choice of β and (1), this entails $(M_1, u) \cong (M_1, v)$ and hence $(M_1, u) \cong_\gamma (M_1, v)$, which by the transitivity of $\cong_\gamma$ gives us $(M_1, v) \cong_\gamma (M_2, v')$ and hence $v \cong_\gamma v'$.

For (11), choose any $a \in \mathcal{A}, v' \in W'$ such that $w'R'_a v'$. Since $\beta < \alpha$, there has to be a $v \in W$ such that $vR_a v$ and $(v, v', \beta) \in \cong^*_{\mathbf{On}}$. Again, we show that $(v, v', \alpha) \in \cong^*_{\mathbf{On}}$ and hence that $(M_1, v) \cong_\alpha (M_2, v')$ and choose a $\gamma \in \mathbf{On}$ such that $\beta < \gamma < \alpha$. So there is an $u \in W$ such that $(M_1, u) \cong_\gamma (M_2, v')$ and $wR_a u$ and hence $(M_1, u) \cong_\beta (M_1, v)$, that is $(M_1, u) \cong (M_1, v)$. It then follows that $v \cong_\gamma v'$. $\square$

We are now in a position to strengthen our previous results: every Kripke world is characterised by a unique sentence of our infinitary language.

Theorem 4.6.13. *For every Kripke world (M_1, w) there is a sentence ϕ_w of $\mathcal{L}_\infty$ such that for every world (M_2, w'):*

$$(M_2, w') \models \phi_w \quad \Longrightarrow \quad (M_2, w') \cong_\alpha (M_1, w)$$

where $\alpha = \mathtt{depth}(\phi_w)$.

Proof Let (M_1, w) be given. By (4.6.12), there is a unique ordinal α such that for all Kripke worlds (M_2, w'): $(M_1, w) \cong_\alpha (M_2, w') \iff (M_1, w) \cong (M_2, w')$. By (4.5.23), (M_1, w) is characterised up to bisimilarity by its $\mathcal{L}_\alpha$ theory $\{\phi \in \mathcal{L}_\alpha \mid (M, w) \models \phi\}$. So $\phi_w := \bigwedge\{\phi \in \mathcal{L}_\alpha \mid (M, w) \models \phi\}$ and ϕ_w is of depth α (by the minimality of α). $\square$

4.7 Further machinery

In this section, I would like to introduce some further notions which will turn out to be useful in the discussion of what will later be called Kripke classifications.

We first note that we may not only restrict models to ordinals (as we did in the proof of (4.6.12)) but also to generated submodels:

Definition 4.7.1 (Submodels). *Let $M = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}}\rangle$ be a Kripke model and $w \in W$ a world. The* submodel M^w generated by w *is the model $M^w := \langle W^w, \pi \restriction W^w, \{R_i \restriction W^w\}_{i \in \mathcal{A}}\rangle$, where W^w is the intersection of all $\bigcup_{i \in \mathcal{A}} R_i$-closed subsets of W containing w, i.e. $W^w := \bigcap\{V \subset W \mid w \in V \wedge \forall i \in \mathcal{A}\ \forall v \in W (v \in V \rightarrow \{u \in W \mid vR_i u\} \subset V)\}$*

There are two other important ways of getting Kripke models from others: disjoint unions and bounded morphisms. To introduce the latter, we first need the notion of an isomorphism between two Kripke models.

Definition 4.7.2. *Let $M_1 = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R_i'\}_{i \in \mathcal{A}}\rangle$ be Kripke models for the same set of agents $\mathcal{A}$. A function $f : W \rightarrow W'$ is an* isomorphism *iff it satisfies* *the following conditions:*

$$
\begin{array}{lrl}
(13) & & f \textit{ is bijective} \\
(14) & \forall p \in \mathbb{P}: & f(\pi(p)) \;=\; \pi'(p) \\
(15) & \forall i \in \mathcal{A}\, \forall w, v \in W: & wR_i v \;\longleftrightarrow\; f(w)R'_i f(v)
\end{array}
$$

M_1 and M_2 are isomorphic *iff there is an infomorphism $f : W \to W'$.*

Definition 4.7.3 (Disjoint unions). *If $\{M_i\}_{i\in\mathcal{I}} = \{\langle W_i, \pi_i, \{R_{ij}\}_{j\in\mathcal{A}}\rangle\}_{i\in\mathcal{I}}$ is a family of Kripke models for the same set of agents $\mathcal{A}$. Let $\{M'_i\}_{i\in\mathcal{I}}$ denote a family of isomorphic, but pairwise disjoint Kripke models. A* disjoint union *$\biguplus_{i\in\mathcal{I}} M_i$ of $\{M_i\}_{i\in\mathcal{I}}$ is then any Kripke model isomorphic to the following:*

$$\biguplus_{i\in\mathcal{I}} M_i := \langle \bigcup_{i\in\mathcal{I}} W'_i, \bigcup_{i\in\mathcal{I}} \pi'_i, \{(\bigcup_{i\in\mathcal{I}} R'_i)_j\}_{j\in\mathcal{A}}\rangle$$

Definition 4.7.4 (Bounded morphisms). *Let $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ be Kripke models for the same set of agents $\mathcal{A}$. A function $f : W \to W'$ is a* bounded morphism *iff it satisfies the following conditions:*

$$
\begin{array}{lrl}
(16) & \forall p \in \mathbb{P}: & f(\pi(p)) \;=\; \pi'(p) \\
(17) & \forall i \in \mathcal{A}\, \forall w, v \in W: & wR_i v \;\longrightarrow\; f(w)R'_i f(v) \\
(18) & \forall i \in \mathcal{A}\, \forall w \in W\, \forall v' \in W': & f(w)R'_i v' \;\longrightarrow\; \exists v \in W(wRv \wedge f(v) = v')
\end{array}
$$

It is easy to see that these three constructions (4.7.1, 4.7.3 and 4.7.4) give us bisimilar models, i.e. that $(M, w) \cong (M^w, w)$, that (M, w) is bisimilar to any disjoint union containing it and to any image under a bounded morphism. It thus follows from th. 4.5.23 that modal formulae are invariant under these constructions too.

Theorem 4.7.5. *Let $M = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$, $\{M_i\}_{i\in\mathcal{I}} = \{\langle W_i, \pi_i, \{R_{ij}\}_{j\in\mathcal{A}}\rangle\}_{i\in\mathcal{I}}$, $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ be Kripke models based on the same set of agents $\mathcal{A}$. If M^w is the submodel of M generated by $w \in W$, we have for every $v \in W^w$:*

(19) $(M, v) \cong (M^w, v)$

If $\biguplus_{i\in\mathcal{I}} M_i$ *is a disjoint union of* $\{M_i\}_{i\in\mathcal{I}}$*, we have for every* $i\in\mathcal{I}$ *and every* $v\in W_i$*:*

(20) $(M_i, v) \cong (\biguplus\limits_{i\in\mathcal{I}} M_i, \vec{v})$[17]

If $f : M_1 \to M_2$ *is a bounded morphism, then the following holds for every* $w\in W$*:*

(21) $(M_1, w) \cong (M_2, f(w))$

Proof

[(19):] We define the following relation on $W \times W^w$: $\cong^* := \{(v, v) \mid v \in W^w\}$ and show that it satisfies the three conditions of def. 4.5.17. (9) is clearly satisfied. For (10), suppose there is an $i\in\mathcal{A}$ and $u\in W$ such that vR_iu. Because W^w is the intersection of all $\bigcup_{i\in\mathcal{I}}$-closed subsets of W, $u\in W^w$. Thus we have the required element $u\in W^w$ such that vR_iu and $u \cong^* u$. (11) is immediate because of $W^w\subset W$.

[(20):] This case reduces to (19) for M_i is (isomorphic to) a generated submodel of the disjoint union $\biguplus_{i\in\mathcal{I}} M_i$.

[(21):] We define the following relation on $W \times W'$: $\cong^* := \{(v, f(v)) \mid v \in W\}$ and show that it is a bisimulation. (9) is guaranteed by (16). For (10), suppose that there is an $i \in \mathcal{A}$ and $v \in W$ such that wR_iv. By (17), we have $f(w)R'_if(v)$ and thus there is $f(v)\in W'$ such that $v \cong^* f(v)$. For (11), suppose that there is an $i\in\mathcal{A}$ and $u\in W'$ such that $f(w)R'_iu$. By (18), there is a $v\in W$ such that $f(v) = u$ ($v \cong^* u$) and wR_iv, which is just what (11) requires. □

We will discuss two other satisfaction preserving constructions: unraveling and filtration.

Definition 4.7.6 (Tree-like Kripke models). *A Kripke model* $M = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ *is* tree-like *iff the following conditions are satisfied:*

- *There is a unique world* $w\in W$ *such that any other world* v *is* R^**-reachable from it, where* R^* *is the reflexive transitive closure of* $\bigcup\limits_{i\in\mathcal{A}} R_i$*.*
- *Every world* $v\in W$ *different from* w *has a unique* $\bigcup\limits_{i\in\mathcal{A}} R_i$ *predecessor.*
- *There is no* R^+ *cycle, where* R^+ *is the transitive closure of* $\bigcup\limits_{i\in\mathcal{A}} R_i$*.*

Theorem 4.7.7 (Unraveling). *Any Kripke world is bisimilar to the root of a tree-like model.*

[17] "$\vec{v}$" here stands for that element of $\bigcup_{i\in\mathcal{I}} W'_i$ the i-th component of which is v.

PROOF Given a Kripke world $(M_1, w) = (\langle W, \pi, \{R_i\}_{i \in \mathcal{A}}\rangle, w)$ we construct a Kripke model $M_2 = \langle W', \pi', \{R'_i\}_{i \in \mathcal{A}}\rangle$, show that it is tree-like with root w and that M_1 is its image under a bounded morphism.

Construct $M_2 = \langle W', \pi', \{R'_i\}_{i \in \mathcal{A}}\rangle$ as follows: W' is the set of all finite sequences of elements of W $\langle w, v_1, \dots v_n\rangle$ $(n \geq 0)$ such that, for some $i_1, \dots i_n \in \mathcal{A}$, there is a path $wR_{i_1}v_1R_{i_2}\dots R_{i_n}v_n$. The R'_i (for $i \in \mathcal{A}$) are defined as follows:

$$\langle w, v_1, \dots v_n\rangle R'_i \langle w, u_1, \dots u_m\rangle :\Longleftrightarrow m = n+1 \ \wedge \ u_i = v_i \ (\forall i \leq n) \ \wedge \ v_n R_i u_m$$

The valuation π' is given by $\pi'(p) := \{\langle w, v_1, \dots v_n\rangle \mid n \in \mathbb{N} \wedge v_n \in \pi(p)\}$.

We now show that $f : \langle w, v_1, \dots v_n\rangle \mapsto v_n$ is a bounded morphism onto M_1. Surjectivity and the first condition (16) are clear from the construction. For the homomorphism condition (17), suppose that $\langle w, v_1, \dots v_n\rangle R'_i \langle w, u_1, \dots u_m\rangle$. Then $u_m R_i v_n$ by the definition of R'_i. For the back condition (18), suppose $f(\langle w, v_1, \dots v_n\rangle) R_i u$. Then there is a sequence, namely $\langle w, v_1, \dots v_n, u\rangle$, such that $\langle w, v_1, \dots v_n\rangle R'_i \langle w, v_1, \dots v_n, u\rangle$ and for which $f(\langle w, v_1, \dots v_n, u\rangle) = u$. □

Recall from def. 4.5.22 that $|w|_\Sigma$ is the equivalence class of all worlds indistinguishable from w with respect to Σ. We are now in a position to state the conditions under which a Kripke model is a filtration of another Kripke model:

Definition 4.7.8 (Filtrations). *Let $M = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}}\rangle$ be a Kripke model. A* filtration *of M* through Σ, *for any subformulae closed set of formulae Σ, is a Kripke model $M_\Sigma = \langle \{|w|_\Sigma \mid w \in W\}, \pi', \{R'_i\}_{i \in \mathcal{A}}\rangle$, where $\pi'(p) := \{|w|_\Sigma \mid w \in \pi(p)\}$ for any $p \in \mathbb{P}$, and the relations R'_i satisfy the following conditions:*

- *If wR_iv, then $|w|_\Sigma R'_i |v|_\Sigma$, for all $i \in \mathcal{A}$*
- *If $|w|_\Sigma R'_i |v|_\Sigma$, then, for all $\Box\phi \in \Sigma$, if $(M, w) \models \Box\phi$, then $(M, v) \models \phi$*

These conditions ensure that we have neither too few nor too many relations in the filtration. The first condition makes filtration a homomorphism, while the second is needed to have the induction step in the proof of satisfaction preservation go through. The conditions are minimal: there will in general be more than one filtration for a given model.

Theorem 4.7.9. *Let M be a Kripke model. The relations $\{R'_i\}_{i \in \mathcal{A}}$ for the smallest filtration $|M|^{min}_\Sigma$ of M through Σ are, for any $i \in \mathcal{A}$, are given by:*

$$|w|_\Sigma R'_i |v|_\Sigma \quad :\Longleftrightarrow \quad \exists w, v \ (w \in |w|_\Sigma \wedge v \in |v|_\Sigma \wedge wR_iv)$$

The relations for the largest filtration $|M|^{max}_\Sigma$ of M through Σ are given by:

$$|w|\Sigma R_i'|v|_\Sigma \quad :\Longleftrightarrow \quad \forall\Box\phi\in\Sigma\ ((M,w)\models\Box\phi \rightarrow (M,v)\models\phi)$$

PROOF We first show that these two Kripke models are filtrations. For $|M|^{min}_\Sigma$, the first condition is clearly satisfied. For the second, suppose $|w|_\Sigma R_i'|v|_\Sigma$, $(M,w)\models\Box\phi$ for a $\Box\phi\in\Sigma$. Then there are $w', v' \in W$ such that $w'R_iv'$. Because of $w, w' \in |w|_\Sigma$, $(M,w') \models \Box\phi$. So $(M,v')\models\phi$ by the definition of $\Box$. As $v, v'\in|v|_\Sigma$ and $\phi\in\Sigma$ because Σ is subformulae-closed, $(M,v)\models\phi$. For $|M|^{max}_\Sigma$, the second condition is clearly satisfied. For the first, suppose wR_iv and $(M,w)\models\Box\phi$ for $\Box\phi\in\Sigma$. Then $(M,v)\models\phi$ by the definition of $\Box$. As this holds for all $\Box\phi\in\Sigma$, we have $|w|\Sigma R_i'|v|_\Sigma$, as required. $\Box$

There is a long tradition in philosophy taking possible worlds to be complete stories or maximally specific states of affairs, settling truth or falsity of every sentence of the underlying language (cf. sct. 4.8.1). As propositions can be identified with the sets of possible worlds where they are true, possible worlds may in turn be identified with the set of propositions true in them. Theories (sets of propositions) then become sets of sets of worlds. Some such sets have properties that allow us to see them as logically structured. Filters (upwards closed sets closed under finite intersections and including the necessary proposition) over the set of possible worlds, e.g., may be seen as *theories* (Blackburn et al. 2001: 94), closed under finite conjunction and entailment. Proper filters, not containing the empty set (the impossible proposition $\perp$), then correspond to consistent theories, ultrafilters (proper filters such that every set of possible worlds either is included in them or in their complement) to complete theories, settling every issue.

To bring out this connection between propositions and sets of possible worlds more vividly, we may make a harmless notational change: instead of defining valuations only for the propositional variables and then give recursive clauses for truth in a world, we can give these clauses directly for the valuation function (interpreting negation by complement, conjunction by intersection and the modalities, e.g. $[a]$, by the corresponding state transformers, e.g. $|a| : \mathcal{P}(W) \rightarrow \mathcal{P}(W)$, $|a|(A) := \{w \in W \mid \forall v(wR_av \rightarrow v \in A)\}$), and then define $(M,w)\models\phi$ just as $w\in\pi(\phi)$. In the rest of this chapter, I will tacitly take this more algebraic perspective.

In a Kripke frame, not every set of propositions describes a possible world. For a given set of propositions, there need not be a possible world where all and only these propositions are true: in this sense, not every proposition is 'realised' by a maximally specific state of affairs. Those realised in this sense are the principal ultrafilters, i.e. filters which are intersections of all filters having a given possible world as a member. In the following, we denote the principal ultrafilter uniquely determined by some possible world w by "u_w". In a frame, however, not every ultrafilter need to be principal. We thus may want to complete our model by adding all non-principal ultrafilters respecting the modal structure imposed by the accessibility relations.

Definition 4.7.10 (Ultrafilter extension). *The* ultrafilter extension of a Kripke frame $\mathcal{F} = \langle W, \{R_i\}_{i\in\mathcal{A}}\rangle$ *is the Kripke frame* $\mathfrak{ue}\mathcal{F} = \langle \mathtt{Uf}(W), \{R^{\mathfrak{ue}\mathcal{F}}_{i\in\mathcal{A}}\rangle$*, where* $\mathtt{Uf}(W)$ *is the set of ultrafilters over W and the* $R_i^{\mathfrak{ue}\mathcal{F}}$*, for* $i\in\mathcal{A}$*, are defined as follows:*

$$u_1 R_i^{\mathfrak{ue}\mathcal{F}} u_2 \quad :\Longleftrightarrow \quad \forall A\subset W\ (\{w\in W \mid wR_iv \text{ for all } v\in A\}\in u_1 \ \rightarrow\ A\in u_2)$$

The ultrafilter extension of a Kripke model M *is* $\mathfrak{ue}M = \langle \mathfrak{ue}\mathcal{F}, \pi^{\mathfrak{ue}M}\rangle$ *where, for every* $p\in\mathbb{P}$*,* $\pi^{\mathfrak{ue}M}(p) = \{u\in\mathtt{Uf}(W) \mid \pi(p)\in u\}$*.*

By mapping any possible world to the principal ultrafilter generated by it, we may embed Kripke frames into their ultrafilter extensions: any Kripke world in the original model is indistinguishable from the corresponding principal ultrafilter in the ultrafilter extension. We can even generalise this result and show that *any* case of finite indistinguishability between Kripke worlds and models is a case of bisimilarity, namely bisimilarity between their ultrafilter extensions:

Theorem 4.7.11. *Let* $M_1 = \langle W, \pi, \{R_i\}_{i\in\mathcal{A}}\rangle$ *and* $M_2 = \langle W', \pi', \{R'_i\}_{i\in\mathcal{A}}\rangle$ *be Kripke models for the same set of agents* $\mathcal{A}$*. Then the following equivalence holds:*

$$(M_1, w) \equiv_{\mathcal{L}} (M_2, w') \quad \Longleftrightarrow \quad \mathfrak{ue}M_1 \cong \mathfrak{ue}M_2$$

PROOF

[$\Longrightarrow$:] The proof has three steps. We show first, that any Kripke world is finitely indistinguishable from its corresponding principal ultrafilter. Then we show that ultrafilter extensions are what is called "modally saturated", i.e. they are models such that whenever every finite subset of a set of modal formulae Σ is satisfiable in the set of successors of some world w, then that set Σ itself is satisfiable in the set of successors of w. Third, we show that any class of modally saturated models has the Hennessy-Milner property, i.e. that any two finitely indistinguishable worlds are bisimilar.

For the first step, we have to show, for any model M, world w, and formula $\phi\in\mathcal{L}$

that

$$(M,w) \models \phi \iff (\mathfrak{ue}(M), u_w) \models \phi$$

where $\mathfrak{ue}(M)$ is the ultrafilter extension of the model and u_w the principal ultrafilter corresponding to w. We show a stronger result, namely the following:

(1) $\quad \pi(\phi) \in u \iff (\mathfrak{ue}(M), u) \models \phi$

for any ultrafilter u. From this, the claim follows by

$$(M,w) \models \phi \iff w \in \pi(\phi) \iff \pi(\phi) \in u_w$$

We show the stronger claim (1) by induction on ϕ.

- Let $\phi \in \mathbb{P}$. We have $u \in \pi^{\mathfrak{ue}M}(p) \Longrightarrow \pi(p) \in u$ by definition of $\pi^{\mathfrak{ue}M}$.
- Let $\phi = \neg\psi$. Then we have:

$$\pi(\neg\psi) \in u \overset{\text{Def. of }\neg}{\iff} W \setminus \pi(\psi) \overset{u \text{ is an ultrafilter}}{\iff} \pi(\psi) \notin u \overset{\text{induction hypothesis}}{\iff}$$
$$(\mathfrak{ue}(M), u) \not\models \psi \overset{\text{Def. of }\neg}{\iff} (\mathfrak{ue}(M), u) \models \neg\phi$$

- Let $\phi = \psi \wedge \chi$. Then we have:

$$\pi(\psi \wedge \chi) \in u \overset{\text{Def. of }\wedge}{\iff} (\pi(\psi) \cap \pi(\chi)) \in u \overset{u \text{ closed wrt finite intersections}}{\iff} \pi(\psi) \in u \wedge \pi(\chi) \in u$$
$$\overset{\text{induction hypothesis}}{\iff} (\mathfrak{ue}(M), u) \models \psi \wedge (\mathfrak{ue}(M), u) \models \chi \overset{\text{Def. of }\wedge}{\iff} (\mathfrak{ue}(M), u) \models \psi \wedge \chi$$

- Let $\phi = [a]\psi$ for some $a \in \mathcal{A}$. We show the two directions separately:
[$\Longrightarrow$:] Suppose $\pi([a]\psi) = |a|(\pi(\psi)) = \{w \in W \mid \forall v(wR_av \to v \in \pi(\psi))\} \in u$. We have to show that $(\mathfrak{ue}(M), u) \models [a]\psi$, i.e. that $(\mathfrak{ue}(M), u') \models \psi$ for any $uR_a^{\mathfrak{ue}\mathcal{F}}u'$. Suppose the contrary, i.e. that there is an ultrafilter u' such that $uR_a^{\mathfrak{ue}\mathcal{F}}u'$ and $(\mathfrak{ue}(M), u') \not\models \psi$. By the induction hypothesis, we conclude that $\pi(\psi) \notin u'$, i.e. that $\pi(\neg\psi) \in u'$, u' being an ultrafilter. But this means that $\pi(\langle a\rangle\neg\psi) \in u$ and so that $\pi([a]\psi) \notin u$, for by the definition of $R_a^{\mathfrak{ue}\mathcal{F}}$, $\pi(\neg\psi) \in u'$ implies $\{w \in W \mid wR_av \text{ for some } v \in \pi(\neg\psi)\} \in u$ and hence $\pi(\langle a\rangle\neg\psi) \in u$.
[$\Longleftarrow$:] Suppose $(\mathfrak{ue}(M), u) \models [a]\psi$. We have to show that $\pi([a]\psi) \in u$, i.e. that $|a|(\pi(\psi)) = \{w \in W \mid \forall v(wR_av \to v \in \pi(\psi))\} \in u$. Again, we prove the contraposition and suppose that there is a $w \in W$ such that there is a $v \in W$ with wR_av and

$v \notin \pi(\psi)$. We have to show that there is an ultrafilter u' on $\mathcal{P}(W)$ such that $uR_a^{\mathfrak{ue}\mathcal{F}}u'$ and $(\mathfrak{ue}(M), u') \not\models \psi$. To do this, we first show that $\{A \subset W \mid |a|(A) \in u\} \cup \{\pi(\neg\psi)\}$ has the finite intersection property (i.e. that any finite intersection of elements of this set is non-empty) and then apply the ultrafilter theorem to get an ultrafilter u' extending it.[18]
$u_0 := \{A \subset W \mid |a|(A) \in u\}$ has the finite intersection property, for $|a|(A) \in u$ and $|a|(B) \in u$ imply $|a|(A) \cap |a|(B) \in u$ (u being a filter) and $|a|(A \cap B) = |a|(A) \cap |a|(B)$ for any two members A and B of u_0.[19] We show that the addition of $\pi(\neg\psi)$ does not destroy the finite intersection property, i.e. that $A \cap \pi(\neg\psi) \neq \emptyset$ for any $A \in u_0$: $|a|(A) \in u$ and $|a|(A) \cap \pi(\langle a \rangle \neg\psi)$ contains at least one element x (u, being a filter, has the finite intersection property and, being proper, does not contain the empty set). This means that x has a successor y such that $y \in \pi(\neg\psi)$. y is also in A (and hence in the intersection $A \cap \pi(\neg\psi)$) for $x \in |a|(A)$ and xR_ay. So $\{A \subset W \mid |a|(A) \in u\} \cup \{\pi(\neg\psi)\}$ has the finite intersection property and there is an ultrafilter u' extending it. Because u' extends $\{A \subset W \mid |a|(A) \in u\}$, we have $uR_a^{\mathfrak{ue}\mathcal{F}}u'$. Because u' extends $\{\pi(\neg\psi)\}$, we have $(\mathfrak{ue}(M), u') \not\models \psi$. Taken together, this completes the proof of (1) and hence of the first claim, which is that any Kripke world is finitely indistinguishable from its corresponding principal ultrafilter.

In the second step, we show that ultrafilter extensions are modally saturated, i.e. we show that whenever every finite subset of a set of modal formulae Σ is satisfiable in the set of successors of some world w of some ultrafilter extension, then that set Σ itself is satisfiable in the set of successors of w. Let $M = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}} \rangle$ be a Kripke model and $\mathfrak{ue}M = \langle \mathtt{Uf}(W), \{R_{i \in \mathcal{A}}^{\mathfrak{ue}\mathcal{F}}, \pi^{\mathfrak{ue}M} \rangle$ be its ultrafilter extension. Suppose Σ is a set of modal formulae such that every finite subset $\Delta \subset \Sigma$ is true, for some $i \in \mathcal{A}$, in some $R_i^{\mathfrak{ue}}$-successor u' of some ultrafilter u on $\mathcal{P}(W)$. We have to show that there is a $R_i^{\mathfrak{ue}}$-successor u'' of u such that $(\mathfrak{ue}M, u'') \models \sigma$ for all $\sigma \in \Sigma$. To do this, we show that the following set has the

[18]The ultrafilter theorem is a consequence of Zorn's lemma, as applied to the special case of lattices of sets. It follows from Zorn's Lemma (as applied to the set of all proper ideal extending a given ideal) that any proper ideal in a Boolean lattice can be extended to a prime ideal (Davey and Priestley 1990: 186). In Boolean lattices, every prime ideal is maximal. A proper ideal (maximal ideal) in a Boolean lattice B is a proper filter (an ultrafilter) in B^{∂}. So it follows that any proper filter in a Boolean lattice, in particular in a lattice of sets, can be extended to an ultrafilter. Because any subset of W generates a filter (the collection of all filters extending it) and any subset which has the finite intersection property generates a proper filter, any such subset can be extended to an ultrafilter in a lattice of sets.

[19]To see this, note that $|a|(A \cap B) = \{w \in W \mid \forall v(wR_av \to v \in A \cap B)\} = \{w \in W \mid \forall v(wR_av \to (v \in A \wedge v \in B)\} = \{w \in W \mid \forall v(wR_av \to v \in A) \wedge (wR_av \to v \in B)\} = \{w \in W \mid \forall v(wR_av \to v \in A)\} \cap \{w \in W \mid \forall v(wR_av \to v \in B)\} = |a|(A) \cap |a|(B)$.

finite intersection property:

$$\Gamma := \{\pi(\phi) \mid \exists n \in \mathbb{N}\ \exists \psi_1, \dots \psi_n \in \Sigma\ :\ \phi \equiv \bigwedge_{i=1}^{n} \psi_i\} \cup \{A \subset W \mid |a|(A) \in u\}$$

From this, we conclude that there is an ultrafilter u'' extending Γ by the ultrafilter theorem. This ultrafilter u'', extending $\{A \subset W \mid |a|(A) \in u\}$, is a successor of u. Because it extends in particular the set of all unary conjunctions over Σ and hence Σ itself, we have $(\mathfrak{ue}M, u'') \models \sigma$ for all $\sigma \in \Sigma$. So we only have to show that Γ has the finite intersection property. We already know that both conjuncts taken separately have this property. So suppose $\phi \equiv \bigwedge_{i=1}^{n} \psi_i$ is an arbitrary finite conjunction of formulae from Σ and $A \subset W$ is such that $|a|(A) \in u$. By assumption, there is a successor u' of u such that $\pi(\phi) \in u'$. By definition of $R_a^{\mathfrak{ue}}$, $|a|(A) \in u$ implies $A \in u'$. So $\pi(\phi) \cap A$, being an element of the ultrafilter u', cannot be the empty set.

The third step in our proof of th. 4.7.11 is to show that any class of modally saturated models has the Hennessy-Milner property, i.e. that any two finitely indistinguishable worlds are bisimilar. It suffices to show that finite indistinguishability $\equiv_{\mathcal{L}}$ between modally saturated Kripke models $M_1 = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}} \rangle$ and $M_2 = \langle W', \pi', \{R'_i\}_{i \in \mathcal{A}} \rangle$ for the same set of agents is a bisimulation. The first condition (9) of (4.5.17), sameness of valuation for propositional variables, is trivially satisfied. For the forth condition (10), suppose that $w, v \in W$, $wR_a v$, $w' \in W'$ and $(M_1, w) \equiv_{\mathcal{L}} (M_2, w')$. It is clear that every finite subset Δ of $Th(v) := \{\phi \in \mathcal{L} \mid (M_1, v) \models \phi\}$ is true in v and hence that $\Diamond \bigwedge \Delta$ is true in w. By the finite indistinguishability of w and w' it follows that $(M_2, w') \models \Diamond \bigwedge \Delta$ so that w' has for any such finite conjunction a successor in which it is true. But this means that $Th(v)$ is finitely satisfiable in the successors of w' and thus, by the modal saturation of M_2, there has to be a successor of w' where $Th(v)$ is true. The proof of the back condition (11) is entirely parallel, making use of the modal saturation of M_1 instead of that of M_2.

So we have shown that ultrafilter extensions are modally saturated and that for modally saturated Kripke models, finite indistinguishability implies bisimilarity. Given that any Kripke model is finitely indistinguishable from its ultrafilter extension, finitely indistinguishable models have finitely indistinguishable ultrafilter extensions which, being modally saturated, are automatically bisimilar.

[$\Longleftarrow$:] Given that any model is a submodel of its ultrafilter extension and that bisimulation of ultrafilter extensions implies indistinguishability by (4.5.23), this direction is immediate. $\square$

A last notion we will need in the following is that of a general frame, which is a Kripke frame with a distinguished set of subsets of possible worlds indicating which propositions are of special interest to us:

Definition 4.7.12 (General frames). *A (multi-modal)* general frame $\mathcal{F}_A$ *is a pair* $(\mathcal{F}, A) = (\langle W, \{R_i\}_{i\in\mathcal{A}}\rangle, A)$ *of a (multi-modal) Kripke frame* $\mathcal{F}$ *together with a non-empty set* $A \subset \mathcal{P}(W)$ *of* admissible *subsets of* W *which is closed under the following operations:*

- *If* $X, Y \in A$*, then* $X \cup Y \in A$
- *If* $X \in A$*, then* $W \setminus X \in A$
- *If* $X_1, \ldots X_n \in A$ *and* $i \in \mathcal{A}$*, then* $\{w \in W \mid wR_iv \text{ for some } v \in W\} \in A$

Models on general frames are constrained by the condition that their valuation maps propositional variables only onto admissible subsets; such models are called *admissible*. Ordinary frames can be taken to be general frames where any subset is admissible. Validity at a state in a general frame is defined as truth at that state in every admissible model; validity *tout court* in a general frame is truth at every state in every admissible model. Validity in a general frame is thus a stronger condition on propositions than just validity and properly includes the latter.[20]

4.8 Outlook

4.8.1 The problem with possible worlds

In this section, I would like to stake stock and identify the principal desiderata for a further development of the theory of which we have seen the main philosophical motivations and the dialectical starting point set by its main competitor. I will identify five principal problems of the epistemic interpretation of modal logic and will show in the following chapters how the theory of information flow devised by Barwise and Seligman tackles – or tries to tackle – them. Given the impressive machinery developed by modal logicians over the years and given the possibility to interpret ordinary modal logic in epistemic terms, one might well question the need to develop another and independent "logic of information". Before presenting such a theory, it may therefore be useful to state again and develop in more detail some of the reasons to be sceptical about the prospects of such a transposition.

Epistemic modal logic, as modal logic, in general, is based on an ontology of possible world, a notion built right into the definition of a Kripke frame (cf. def. 4.5.1).

[20] To see this, consider the McKinsey formula $\Box\Diamond p \rightarrow \Diamond\Box p$, invalid e.g. on the frame $(\mathbb{N}, <)$, but valid on a general frame based on $(\mathbb{N}, <)$ in which only finite and cofinite subsets are admissible.

The first and foremost worry with possible worlds is ontological. Situation theory, according to its proponents, is firmly committed to the claim that there is only one possible world, namely the actual one:

> "A basic idea of situation theory is that there is a concrete reality, which has concrete parts but not concrete alternatives. This reality can be thought about, perceived, studied, and analysed in a variety of different ways, from a variety of different perspectives, for a variety of different purposes." (Israel and Perry 1990: 7)

Seligman, in his work on perspectives and taking up the original motivation of situation theory of describing *scenes*, i.e. what is seen by epistemic agents and what makes their perceptual reports true, has shown that things are more complicated than Israel and Perry make them appear. "What is seen" is ambiguous: in one sense of the locution, you and me, watching Sam eating a sandwich, are seeing the same thing; in another sense, what we see is different, for we see Sam from different angels. As we noted in sct. 4.4, situation theory but not epistemic modal logic has the resources to allow for a necessary degree of perspectivism.

The ontological worry about the status of possible world is not confined to epistemic modal logic but is much more general. Various anti-realistic proposal have been made, but I lack the space to discuss them here. Suffice it to note that classifications, the ontological basis of the theory exposed in later chapters, offer an attractive alternative.

Even given an ontologically innocent reading of "possible worlds", however, there are other problems with this notion.[21] We already noted that situation semantics was partially motivated by a focus on situations as opposed to models and hence by a bias towards partial modelling.[22] There are both positive and negative reasons for partiality. One positive reason is that it just seems implausible that ordinary, limited situation like my reading a book or the two of us having a discussion really are sets of all total situations (models of the world) that contain it. The negative reasons against taking them for that are, first, that we cannot be sure that the collection of all these total situations forms a set (cf. sct. 4.8.5). Second, even if it is a set, there are much more issues resolved in the set of all total situations containing a given situation than in the situation itself. In every total

[21] The criticism of the possible-world approach chosen by Hintikka (1962) is almost as old as the approach itself. Cf.: "In possible-world semantics, "$\exists x \mathsf{K}_a Fx$" is read as "There is something which exists in all worlds compatible with a's knowledge and which is F in all such worlds." This, to my mind, in no way illumines the meaning of "There is something known to a which is known by him to be F."" (Stine 1974: 129–130)

[22] I do not want to imply by this that anyone working in the possible-worlds tradition is committed to all these features of 'possible worlds' as understood here. Hintikka, e.g., has acted the innocent: "[...] situation semantics complements rather than contradicts rightly understood possible world semantics ..." (Hintikka 1983: 207)

situation that contains my writing this section, e.g., Bin Laden either is alive or dead. My tipping on my computer, however, has nothing to do with Bin Laden at all. No terrorist is in any way part of the situation(s) I am now in (or so I hope).[23] Another negative reason resides in the fact that taking possible worlds to be entities of their own, taking them to be *represented* and not *representing* items, means to loose our grip on relevant (epistemic or informational) equivalences. Possible worlds are supposed to model real situations and there just seems no room for differences between them which are not differences with respect to how they model the world. Differences between representationally equivalent possible worlds seem a gratuitous artifact of the theoretical machinery in use.

Bisimulation is a good example of such a unnecessarily non-trivial relation. The richer our language, the more expressive it is a a representational system, the more issues it can describe and distinguish. In addition to the expressiveness brought in at the ground level, by the choice of the relevant propositional variables $\mathbb{P}$, there is also the logical complexity of the formulae we allow for which gives us resources to distinguish possibilities. So it seems that the infinitary language $\mathcal{L}_\infty$ is maximally expressive among languages based on some given choice of logically simple propositional variables. So it does not seem to be justifiable to allow for $\mathbb{P}$-models which cannot be distinguished even by this maximally expressive language. Given (4.5.17), there therefore seems no justification to distinguish bisimilar models.

Two Kripke worlds represent the same situation iff they give the same interpretation to the propositional variables and for every world a-accessible from one of them there is a world a-accessible from the other that represents the same situation. So two Kripke worlds represent the same situation iff they are bisimilar. Bisimilarity is informational equivalence, i.e. carrying and making available the same information – there is thus no reason to individuate worlds finer than by bisimilarity. From a representational viewpoint, taking possible worlds to be representing devices, not what is represented, bisimulation as representational equivalence is indeed the right point to stop distinguishing between possible

[23]The fact that the possible worlds machinery – at least if the worlds are taken to be the metaphysically possible worlds – imports factors into epistemic logic which are extraneous to it, may be even more clearly brought out with respect to first-order epistemic logic. If we agree that identity is (metaphysically) necessary and that (at least some) individual constants are rigid designators (designate the same individual in all possible world), then it is impossible not to know any true identity statement, the Babylonians' ignorance about whether Hesperus is Phosphorus notwithstanding. Even without this assumption (which goes back to Kripke (1972)), however, we have contra-intuitive results (for bound variables are treated as rigid designators in the ordinary semantics for first-order logic: it follows even then that, for all x and y, if x and y are identical, it is not compatible with anyone's knowledge that they are not identical; and that everything is such that it is not compatible with anyone's knowledge that it does not exist (cf. Stine 1974: 137).

worlds.[24] Changing possible worlds for epistemic states, as we did in (4.6), remedies this situation.

Barwise, calling informational alternatives (as described in sct. 4.4) *relevant issues*, has indeed suggested the following substitute for the notion of "possible world":

> "Given the relevant issues [...], whatever they are, one can consider all the various ways of resolving these issues. This is what we shall mean by a state: *a way of resolving all the relevant issues.*" (Barwise 1997: 494)

If we conceive of issues as states of affairs (objects, locations, relations and polarities) made factual or supported by situations, this gives us a somehow anti-realistic picture of situations, identifying situations which resolve issues in the same way. There are thus two crucial differences between possible worlds and situations: the former are all-embracing, the latter partial; the former are abstract ways of specifying a total state of the world, the latter are real limited parts of reality.[25] Partiality is not to be confused with relativity: even if possible worlds are "way[s] of representing information about possibility *relative to* some set I of issues" (Barwise 1986d: 431), it is a further question whether any such possible world provides *total* information about every issue $i \in I$.

Apart from their all-embracing globality, there is another quite general worry with possible worlds. Whether they are conceived of as alternative total states of the world, as abstract entities or as ways of resolving all relevant issues – they are something which has to be assumed to be given at the outside, on the modelling level, not the level modelled, something the nature and dynamics of which is not capable of being modelled inside the theory. Possible worlds seem to be given from the theoretician's viewpoint, as formal tools to represent the cognitive agents' epistemic behaviour and reasoning, not as something itself within the cognitive reach of the agent.

This perspective misrepresents both the nature of the relevant issues themselves, coming into existence as an *answer* to, not a precondition of, the queries of an agent, and the intended field of applications of constraints, which are not supposed to be relative to some complete state of the world, but supposed to apply relative to some epistemic situation or other. We will later see how the normal situations used to define information contexts (cf. def. 7.1.1) bring us much closer to an adequate solution of these two issues.

[24] Seligman and Moss (1997: 252) take bisimulation to be the appropriate principle of identity for infons.

[25] Cf.: "Situations are with worlds; a world determines the answer to every issue, the truth-value of every statement. Situation[s] correspond[...] to the limited parts of reality we in fact perceive, reason about, and live in. What goes on in these situations will determine answers to some issues and the truth[-]values of some statements, but not all." (Barwise and Perry 1999: xxv)

Possible worlds, as used in the semantics of modal logics, are not things that are given to the theorists and on which he may freely draw to model whatever is of interest to him: they are themselves modelling devices, of the same order than infons and propositions and of a different order than situations, agents and information states.[26]

Both Dretske's information-theoretic account[27] and Barwise's and Perry's situation-theoretic account of knowledge in terms of epistemic alternatives, however, construe the range of alternatives as subject to change as a function of the agents' total epistemic state. Of course epistemic modal logic – at least for systems other than **S5** – allows for that by offering different accessibility or alternativeness relations for different agents. But the choice is very limited: it concerns only the selection and not the very nature of the alternatives. If not only the question which issues are relevant and how they are solved, but also the question, e.g., how the relevant issues are specified, receives different answers for different agents in different cognitive states, something very different is needed. We will go some steps in this direction in our final chapter on information frames (ch. 8).

4.8.2 The problem of omniscience

A general problem for any possible-worlds approach to information is that – according to the "Inverse Relationship Principle" principle (1.4.1) – necessary truths, excluding no possibilities, carry no information. And yet it seems plain that true mathematical, logical and metaphysical statements *tell us* something, that we learn something by becoming convinced of their truth by a reliable source. So any theory of information should account for the information carried by necessary truths and be able to distinguish between different necessary and impossible truths in terms of the information they carry.

The same holds for epistemic logic. Standard epistemic logics impute logical omniscience on the agents the cognitive behaviour they claim to model. The problem is connected to their use of possible worlds: what is true in all possible worlds is, a fortiori,

[26]This corresponds roughly to the way Barwise interprets Stalnaker: "By a possible world, Stalnaker does not mean a world at all. Rather, what he has in mind is information about *possible ways* the world might be, or might have been." (Barwise 1986d: 431) I think this is a charitable, though probably exegetically false, interpretation of Stalnaker.

[27]As we saw in sct. 2.3, the perspectival relativity built in Dretske's notion of informational content has been overlooked by many of his readers. There are, though, some clear indications that he regards both $I(s)$ and $I_s(r)$ (the information generated at s and the part of this which is received at r) as relative to an agreed upon way of dividing up the possibilities at the source: "...whether or not a signal is equivocal depends on how we carve up the possibilities at the source" (Dretske 1981: 61). Though he seems to think of this relativity, as we noted on p. 50, as some kind of dependence on notation, he nevertheless acknowledges its existence.

true in all the possible worlds which are epistemic alternatives for a given agent at a given time. In classical modal epistemic logic (as exposed above in sct. 1.3.2 and 4.5), this is mirrored by the necessitation rule (Nec on p. 13):

$$\frac{\vdash p}{\vdash \Box p}\ \text{Nec}$$

In the case of epistemic logic, however, logical omniscience is just a special case of a more general problem: epistemic alternatives, being worlds the actuality of which a given agent is in no position to exclude, do not have to be – and in practical cases never are – transparent to the agent in question. This means that they are given to him under descriptions, in typical cases by some sort of indirect description of the form "a world where all I think I know is true". But even where they are given in some more explicit way, e.g. by "a world where some proposition p I do not know is true", they are not (and perhaps never) completely described. This means that many things may be true of them – and may even be true of *all* of them – without the agent having any idea about them.

Suppose there is some unheard-of general feature of the world, which happens to be had by all the worlds the actuality of which I cannot exclude on the basis of what I know. I will then, quite implausibly, be said to *know* that the actual world has that feature. In the normal treatment of the possible-world framework, this danger is normally evaded by some strong assumption of *plenitude*, assuring us that there is enough variety among possible worlds. This is questionable, however, for insofar as the epistemic logic is taken to be based on some sort of "metaphysical" possibility, there might be not only unknown metaphysical necessities (as there certainly are), but all kinds of metaphysical connections among the possible worlds invoked to model my knowledge. This problem then reduces to the case of logical omniscience and the closure of knowledge under known entailment, which makes the omniscience problem especially pressing in our context (cf. p. 13):

(K) $\vdash \Box p \wedge \Box(p \rightarrow q) \ \rightarrow\ \Box q$

Unknown, but necessary connections between propositions become modalised and license inferences from known propositions (p) to other, perhaps unheard-of, propositions (q).

The most popular approach to the problem of logical omniscience, the one invoking "impossible worlds" of different kinds, consists in denying (1.4.1). Not only the exclusion of possible, but also that of impossible situations is said to carry information. While allowing for worlds where, say, $2 + 2 = 5$ is true solves the problem of mathematical omniscience, it does not solve, e.g. the problem of logical omniscience, except if, say, all the axioms of Peano Arithmetic are true in that world. In that case, however, and given classical logic, all the worlds where $2+2=5$ are worlds in which contradictions are true and vice versa. But it is perfectly possible to believe $2+2=5$ without believing all contradictions whatsoever (it is not even clear that the latter is possible at all). So we do not only need impossible worlds, but different (kinds of) impossible worlds to be able to discriminate the information carried by different impossible statements. On the other hand, we should not individuate them too fine-grainedly neither: even if the information carried by "$2+3=4$" is different from that carried by "$2+2=5$", it is the same than that carried by "$3+2=4$". So our impossible worlds better had not to be too impossible after all.[28]

Another, to my view more promising route, is to sacrify closure of knowledge or belief under conjunction, i.e. the factorisation principle $(\Box p \wedge \Box q) \rightarrow \Box(p \wedge q)$.[29] This would allow for the following

> "... the very possibility to model the context-sensitivity of the inferences rests on the (partial) removing of the structural rules: in order to prevent dissonant cognizers from admitting blatant inconsistencies, we have to restrict the scope of these rules." (Dubucs (1991: 53), also cited in Gochet and Gribomont (2002: 17))

As we saw (in fn. 20 in sct. (1.3.3), this was Ruth Barcan Marcus' way of distinguishing having contradictory beliefs from believing in contradictions.

The case against closure under conjunction is particularly strong if we accept (9) discussed on p. 16 in sct. 1.3.2, i.e. closure under valid equivalences. Suppose there are only three Fs and all of them are G: a, b and c. I may well know of every F I know that it is G (I am, e.g., a good G-detector among Fs). Suppose I happen to know of a, b and c that they are F and G. Knowing that I am a good G-detector among Fs, but not a good F-detector in general, I know that $Fa \rightarrow Ga$, that $Fb \rightarrow Gb$ and that $Fc \rightarrow Gc$. It is clear, however, that I do not know that all Fs are G – just because there may well be, for all I know, other Fs which are not G (I am, after all, a bad F-detector). I do not know that all Fs are G even though $\forall x\ (Fx \rightarrow Gx)$ and $Fa \rightarrow Ga\ \wedge\ Fb \rightarrow Gb\ \wedge Fc \rightarrow Gc$ are logically equivalent. So, given (9), I fail to know the conjunction of what I know.

[28] There are, of course, other problems with impossible worlds: if impossible worlds are worlds where contradictions are true, then we need something different from the ordinary definition (4.5.6) of truth-in-a-world. For else we get a contradiction *in our world* and "there is no subject matter, however marvellous, about which you can tell the truth by contradiction yourself" (Lewis 1986: 7, n. 3). But even if this is rejected, the move from $w \models \neg\phi$ to $w \not\models \phi$ gives us a reductio, for it makes the actual world (and, presumably every other world), impossible (cf. Stalnaker 1996: 196).

[29] This is the principle (12) – closure of belief or knowledge under conjunction – we met in sct. 1.3.2.

We will later, in sct. 7.1.1 and 7.2.1, see how Barwise and Seligman (1997) deal with the problem of omniscience and solve it by invoking impossible worlds of a very special kind.

4.8.3 The problem with constraints

Another problem already mentioned (cf. sct. 3.3) is that constraints, while important already in the original 1983 theory (cf. 3), are not yet adequately modelled – neither in situation semantics nor in epistemic modal logic.[30] The crucial *desideratum* is to make room for error, i.e. constraints attunement to which may lead us astray in certain special circumstances.[31]

In Barwise's and Perry's *Situation and Attitudes* theory *conditional constraints* account for error (Barwise and Perry 1983: 94). Conditional constraints are constraints which hold only under certain circumstances, and hence subjects may falsely believe that these circumstances obtain and falsely rely on the constraints requiring them. The problem then is that conditional constraints are conditional on some background types, i.e. that the conditions under which they hold have to be specified at the outset.[32] A much more flexible way to model such 'exceptions' is desirable, if only because it is in many practical cases virtually impossible to state the intended background types.[33]

The problem, then, is how to model the conditions on which most, if not all, constraints depend – and this is tantamount to the problem of specifying under what conditions information flows.[34] It is also connected to the problem of accounting for the non-monotonicity of much of ordinary-life reasoning. A piece of reasoning is non-monotonic,

[30] Cf. Seligman and Moss in 1997: "Despite its importance, and considerable research effort, the theory of constraints has remained elusive." (1997: 299)

[31] This is reiterated in *Information Flow*, cf. "...a theory of representation must be compatible with the fact that representation is not always veridical" (Barwise and Seligman 1997: 235)

[32] This is also how conditional constraints are described in Seligman and Moss (1997: 393).

[33] Barwise became soon aware of this lacuna in the 1983 account: "When I speak of certain conditions obtaining, I am not referring to the conditions described by the antecedent of these conditionals, but to other, more pervasive, background conditions that generally obtain. [...] There is simply no reason to suppose that there is any way to flesh out a conditional statement to incorporate a description of the exact conditions under which the conditionals holds, or even the conditions under which the speaker believes it to hold." (Barwise (1986a: 107, 109), cf. also Barwise (1986b: 160) and Devlin (1991: 109)) Jerry Seligman (1990) subsequently proposed to make constraints conditional on what he calls perspectives, precursors of what are here called information contexts (to be discussed in ch. 7.1).

[34] The connection between the two issues is explicitly made by Barwise as early as in 1986: "In terms of Perry's and my theory, informational relations are usually *conditional* constraints, constraints that only apply under certain conditions, but there is often nothing there in the head that corresponds to these conditions." (Barwise 1986b: 151)

the idea is, iff it exploits conditional constraints the non-obtaining of which under certain additional conditions the agent is disposed to recognise.

In Barwise's and Seligman's account, non-monotonicity is accounted for by a relative picture of information channels (we will come back to this issue in sct. 5.2.2): what information flows is relative to some channels; channels exist at different levels of 'refinement', that is, granularity. A piece of reasoning is non-monotonic iff it relies on a channel that actually is more refined than the agent takes care to state: as soon as the exception arises, the agent thus switches back to the more refined channel, on which he was relying all along, according to which the anomalous situation does no longer count as an exception.

This feature of the channel-theoretic account of error and non-monotonicity has been criticised by Lemon:

> "However, this [the suggestion that non-monotonicity shows that "people are rather good at changing channels when exceptions arise" (Barwise and Seligman 1997: 45)] seems to miss the point of non-monotonicity. The model presented here claims that, once we know the background conditions for the inference, and so change to the 'right channel', there is no need for non-monotonicity – but the whole point of such logics is that we have to reason under uncertainty (precisely when we do *not* know all the relevant background conditions), and the channel-theoretic account says little about these cases." (Lemon 1998: 398–399)

Lemon here misses the point noted above (sct. 1.4.7) that people do not have to be aware nor to somehow explicitly represent the regularities on which information flow they receive or produce depends. "Changing the channel", then, just means becoming aware of an exception and recognising it as an exception – exactly what happens in non-monotonic reasoning. Recognising a situation as exceptional means to classify it as abnormal relative to a former, less inclusive classification.

Attunement to a constraint is not being aware of it, but relying on it in information-exploiting behaviour. Attunement does not require mental representation. It is therefore an error to identify, as Devlin does (1991: 91), attunement and awareness. Whether or not an agent may be said to be attuned to a constraint depends on what he does, in particular on how we react to putative counterexamples to inferences he made, not on what he believes or even on what he is explicitly aware of.[35]

Apart from the worry about conditional constraints, another general problem with constraints as modelled in the 1983 theory is that their ontological status is unclear. What

[35]Devlin is not particularly clear about this point. He characterises attunement to the constraint "smoke means fire" e.g. as "a form of familiarity with, or behavioral adaption to, the way the world *vis à vis* smoke and fires" (Devlin 1991: 91) and as "behavior-guiding awareness of knowledge" (1991: 100).

exactly is described by a true conditional statement $p \to q$? Does it have the form $\sigma \models p \Rightarrow \sigma \models q$ or rather $\sigma \models p \Rightarrow \exists \sigma' (\sigma' \models q)$? If the latter, how are σ and σ' related? The answer, in short, is: by a constraint. What, then, is a speaker classifying by uttering $p \to q$?

The key move, advocated by Barwise (1993), to answer this question is to make use of the crucial difference between situation meaning and situation-type meaning discussed in sct. 3.3. There are two very different relations involved here: a *channel*, i.e. an informational dependency between situations, and a *constraint*, i.e. an abstract regularities between types of situations that correctly describes the channel in question:

> "When one expresses a constraint $\phi \to \psi$, one makes a claim about one of these relationships [between particular situations], say $\overset{c}{\mapsto}$, that it supports the constraint. What it means for $\overset{\mapsto}{c}$ to support one of these constraints, say $\phi \to \psi$, is roughly this: under normal circumstances, if $s_1 \models \phi, s_1 \overset{c}{\mapsto} s_2$ and $\phi \to \psi$, then $s_2 \models \psi$." (Barwise 1993: 5)

We will see much latter, in sct. 6.1.1, how this crucial distinction finds its place in Barwise's and Seligman's theory of information flow.

Another motivation for the explicit incorporation of constraints is given by our two observations about Grice and Dretske (on p. 10 and p. 52 respectively): information flow is some sort of *specific* dependence, where the dependent fact allows for a road back to (makes information available about) what it depends on. My utterance of p, if meaningful, not only allows you to infer that I intend to make you believe that p but it allows you to infer that my intending you to believe p is a *reason* for my intention to be understood, i.e. for my intention to make you believe that I intend you to believe that p. So you not only ascribe to me an intention of some general type, but an intention that is supposed to play a particular rôle in my cognitive architecture. Similarly, if I see some smoke, this does not carry just the information that there is some fire or other. It carries information only about one specific fire, namely the one that produced the smoke (this is the upshot of Dretske's 'aboutness' condition in 2.2.2).

To account for information flow, it is therefore not enough to construe a system which satisfies certain constraints. The constraints themselves, in virtue of their rôle of connecting two different facts or propositions, have to enter the picture, because only the particular fact that a constraint holds in a certain situation gives me a road back from an indicating fact to the particular fact it indicates.

Though they did not acknowledge this explicitly, this was a major drawback of Barwise's and Perry's early theory. Consider e.g. the following passage:

> "In a world knitted together by constraints – whether these be constant conjunctions

> or some more metaphysically potent connections – situations carry information. The fact that there is a situation of one type, carries the information that there are situations of the types that one involves. If it is a constraint that objects left unsupported near the surface of the earth fall, then the fact that a certain apple near the surface of the earth is left unsupported, carries the information that it will fall." (Israel and Perry 1990: 3–4)

I do not disagree with this passage, though I think it important to recognise that the claim made by the third sentence is in no way a specialisation of the one made by the second. The information *that there are situations of some types or others* does not allow me to infer anything about a particular apple, *unless* I already know that the situations in questions involve the apple. This, however, is a new piece of information I might lack.

In the theory to be developed below, this crucial distinction between informational connections as abstract regularities of a system and actualised information channels between real objects is nicely mirrored by the distinction between information channels, which connect particular classifications involving concrete objects, and regular theories, which describe classifications at the level of their types. Every informational channel, by th. 5.2.7 guaranteeing that it has a minimal cover, can be taken to be an entity in its own right, itself capable of being classified and itself entering into infomorphisms (and hence channels) of all kind. Regular theories, on the other hand, are not objects, but families of claims about objects.

I take this explanation to be superior to the one provided by Israel and Perry (1990: 5), which is that a connecting fact turns pure (general) information into incremental information (about specific objects). The problem with this, as with the handling of such problems in situation semantics in general, is that it inflates situations with 'aboutness' facts, i.e. facts of the type that something (say an x-ray) is about something (say a particular dog). As Dretske remarked (cf. p. 2.2.1), however, this is not necessary: a situation carries information about another situation not in virtue of the fact that we have to mention (parts of) the latter to completely specify the former, but just by the fact that they are appropriately connected to each other. Whether or not this connection holds, is a fact external to the situations at hand. This is just another manifestation of the relativity of information to constraints and gives us another reason to model such constraints explicitly in our theory.

The problem with incremental information is that it can only be handled by *relative constraints* which are constraints connecting pairs of situations with some connecting situation. The problem (as remarked above) is that there is, at least in normal cases, no one situation type which holds of all the connecting situations. Even if the involvement between two situation types is perfectly general, situations of one of these types may in various ways be *about* some other situations (Israel and Perry (1990: 13) use e.g. the type *is-x-ray-of*). It seems much preferable, then, to be able to directly represent constraints: the Xerox principle (2.2.1), e.g., then becomes not a dubious claim about the existence of general enough situation types, but a simple existence statement to the effect that channels may be compounded.[36]

4.8.4 The problem of epistemic dynamics

There is another facet to the 'room for error' problem not adequately handled either by epistemic modal logic nor by Dretske's or Barwise's and Perry's more philosophical theories. Information is essentially something the learning of which leads us to (and justifies us in) correcting our belief sets.

The problem is related in the following way to that of the modelling of constraints. We saw that both Dretske's theory and situation semantics insists on the veracity of information (sct. 1.4.5) and allows for a straightforward account of knowledge, while the picture of belief is both much more complicated and draws on resources external to the main theoretical setting. In both cases, the "knowledge implies belief" principle appears to be super-imposed, motivated more by a desire to 'save the phenomena' (or not to offend common sense) than by a sound inner-theoretical foundation (cf. sct. 1.3.7).

What makes beliefs difficult to model in both theories is the fact that beliefs may be (and too often are) false. Dretske tackles the problem of misrepresentation by bestowing on symbols a standard informational rôle in learning, thereby reifying the use they are put to in successful cases of representation.[37] Barwise's and Perry's attempt at "representing the mental" may be seen to be a similar, though more sophisticated attempt.

The crucial point now is that, as in the case of information carrying signals and token meaning, the very possibility of misrepresentation depends on regularities, though they are not causal and veridical but conventional, and situated not at the level of signals and situation tokens, but at the level of signal coding and situation types. The difference between information and misinformation, then, is not, as Dretske and Barwise/Perry seem to have believed in the early eighties, to be seen as that between reliance on actual or merely apparent informational connections, but as the distinction between two levels on

[36] Israel and Perry (1991) explicitly formulated the xerox principle as saying that there are some specific constraints and connecting facts (Israel and Perry 1991: 157).

[37] In his 1988 book *Explaining Behavior* Dretske (1988) he tries to improve this explanation by invoking natural functions of representations. The general gist of the explanation, however, remains the same.

which actual informational connections operate.

Dynamics forces us to make room for different states of an agent at different instants in time. There is an independent motivation for this: information flow, like transmission of knowledge, typically takes place in a multi-agent environment. There is no reason to assume that these agents differ only in what situation they are in (i.e. with respect to which elementary propositions are true in their situation) and in which worlds are compatible with what they know. They might differ, e.g., also in the standards they have for knowledge, or in the formal structure of their epistemic alternativeness relation. So a pluralistic picture seems to be called for. Barwise realised this as early as in 1988: he suggested to conceive of the theory of information content not on the model of set theory, where people take their task to be modelling one (supposedly) coherent and shared set of intuitions, but on the model of topology, which provides the general framework in which many different topologies (and, associated with them, many different conceptions of 'nearness') may be studied:

> "The final analogy I want to draw from topology comes from the proven importance of studying not just topological spaces themselves, but *relationships between* different spaces. Indeed, it has turned out that one of the most important notions of topology has been that of a homeomorphism, a function from one space to another that respects nearness relations in an appropriate way. I think that something similar is going to happen here. I suspect that we are going to find ourselves studying "infomorphisms", maps from one information space to another that preserve information. But before we can define these maps, we need to understand the basic structure they need to preserve." (Barwise 1989a: 256–257)

And indeed this was what happened: adopting the most liberal stance possible, Barwise and Seligman set out, in 1997, to study the interrelation of different "information contexts" associated to different classifications (cf. ch. 7.1 below). They thus took into account the pluralistic nature, not only of what is true of different agents (or time-slices of an agent) and of what they know, but also of *how* they know things, i.e. how they individuate the world they know things about. This need for pluralism gives us another, independent, reason for explicit modelling of the constraints in virtue of which information flow takes place:

> "The situation theoretic model of an organism's understanding of its environment is therefore composed of two parts: a scheme of individuation and a collection of constraints. One of the potential advantages of this model is that it allows a flexible approach to describing the ways in which different organisms conceive of their world." (Seligman 1990: 148)

The problem of dynamics is correlated with the modelling of the efficiency of language, which was one of the guiding ideas behind situation semantics (cf. p. 70). Sentences are essentially re-usable and their uses may in different situations have different causes and effects: they may be used to say different things. One way of modelling this re-usability (which most prominently occurs with indexicals) in standard possible world semantics is to add context sets, determining a speaker, a time and place of utterance etc., and make the evaluation of uttered sentences depend on the context set associated with the utterance. This procedure, however, does not extend to demonstratives, nor does it provide us with a clear picture on what the meaning of such sentences depends: context sets seem a theoretician's artifact, without independent plausibility of their own. Going partial, on the other hand, allows a unified treatment:

> "...by admitting situations into the theory, you no longer need *ad hoc* devices, but rather, the same devices that serve to be what statements are about can also serve as the context of statements, as well as the embedding circumstances of other meaningful items. It is in this way that one gets a relational theory of meaning where the things being related are the same kinds of things." (Barwise 1987: 88)

The need for explicit modelling of the background conditions on which information flow depends brings us back to one of the main motivations behind situation theory: everything which is relevant from the modeller's perspective may be *made* relevant from the agents' perspective.[38] This means that we cannot, indeed must not, draw a categorial divide between the inferences the epistemic agents are supposed to draw and the ones the modeller is entitled to: while they have to be kept apart, a modeller's perspective should in principle be open to the agents as well.[39]

4.8.5 The problem with set theory

One of the reasons Barwise and Perry give for their dissatisfaction with conventional possible-world semantics is the use it makes of an underlying set-theoretical machinery. Their reasons are both specific and general.

The first specific problem is that at least some formal developments of possible-worlds semantics takes the space of possibilities to be a set:

[38] Cf. Barwise's *credo*: "...in situation theory, anything we use can be objectified and talked about. This applies to situations themselves, to relations, operations, conditions, parameters – whatever; and it gives situation theory a rather different flavor from more traditional, "closed" logical theories." (Barwise 1986e: 179–180)

[39] I do not, therefore, agree with one of the main ideas of Devlin's, which is to base situation theory's claim on empirical adequacy on a distinction between the agents' and the theorists' schemes of individuation (Devlin 1991: 27) and even logics (1991: 128), which buys into a dubious relativism (cf. e.g.: "...we may have no way of knowing how things *really* appear to the agent" (Devlin 1991: 128)).

> "Not only does the formal theory take other [merely possible] realities as primitive, it assumes that there is a *set* of all such. It must, in order to have arbitrary subcollections (propositions) count as legitimate objects. But the assumption is at odds with the view of sets built into the set-theoretic metatheory of possible worlds semantics." (1985: xlvii)

Possible world semanticists may retort, as Stalnaker (1986) did, that possible world semantics is not committed to the claim that there is a space of all possibilities (or the "space of all issues", as Barwise (1987: 83) prefers to call it in):

> "If one thinks of possible worlds simply as a formal device for modeling alternative total ways the real world might have been, not as alternative concrete realities, then there is nothing to preclude setting the whole apparatus up *relative to* some particular limited space $\mathcal{L}$ of issues, so that the "total" is relative to this space of issues." (1987: 83–84)

The problem with such an instrumentalist stance, however, is that it forsakes the explanatory potential of theories such as situation semantics and epistemic logic, which are not just formal tools to uncover the logical structure of, say, naked infinitive perception reports and knowledge ascriptions, but are supposed – by their protagonists, at least – to state what such sentences are *about*. This is evidenced by the fact that such frameworks are regularly used to introduce theoretical entities – propositions and epistemic states for example –, which should play an explanatory rôle. This ontological worry is not even addressed by a Stalnakerian instrumentalism.

Another problem in this area, not noted by Barwise but equally relevant to his concerns, is the one stated by David Kaplan (1994) in *A Problem in Possible-Worlds Semantics*: If propositions are subsets of the set of possible worlds, then, by Cantor's Theorem, there are more of them than there are worlds. So there are not enough possible worlds to differentiate contingent attitudes towards propositions. What is more, there is only a set of contingent propositions and only a set of objects which are such that some contingent proposition is about them (if we assume that for a proposition to be about an object is, at least, for it to be true in at least one world where that object exists). As there is no set of all objects, though, there are many objects of which no contingent proposition is true, hence many objects which only have essential properties.

To bring the problem to the fore, consider the following proposition:[40]

(22) $\forall p \diamond \forall q \; (Qp \leftrightarrow p = q)$

[40] In the **S5** case, where accessibility is universal, we may replace (22) by $\forall p \diamond \forall q (Qp \leftrightarrow \Box(p \leftrightarrow q))$.

The problem now is that (22), though intuitively contingent,[41] is necessarily false.[42]

After many years of discussion, there is still no consensus what to do about this. Kaplan draws the conclusion that intensional logic should not decide metaphysical questions like the truth of (22) and thus turns his paradox into an argument against possible-world semantics:

> "... if [possible-worlds semantics] is to serve for intensional *logic*, we should not build such metaphysical prejudices into it. We logicians strive to *serve* ideologies not to constrain them." (Kaplan 1994: 42)
> "If propositions are entities [as they are assumed to be in possible-worlds semantics], then our modal logic, in its premetaphysical purity, should accommodate a nonlogical sentential operator [Q] satisfying (22)." (Kaplan 1994: 44)

I think this is very much in the spirit of situation theory: logic should be maximally neutral. If logic decides substantial questions, this in itself (whatever the answer these questions receive) is an argument *against* the logical system in question. As David Lewis aptly remarked, however, "... the way of the neutralist is hard" (Lewis 1986: 105, n. 2). One way to deal with (22), it seems, is to bite the bullet and argue that not every proposition defines a specific possibility. This was the way chosen by Lewis (1986: 105):

> "Not just any set of worlds is a set that might possibly give the content of someone's thought. Most set of worlds, in fact all but an infinitesimal minority of them, are not eligible contents of thought. It is absolutely impossible that anybody should think a thought with content given by one of these ineligible sets of worlds".

Against this and in favour of his own favourite way out – a ramification of models, i.e. possible worlds – David Kaplan made the observation that, even without the assumption that there is a possibility for every proposition, (22) gives us a contradiction when combined

[41] To see this, interpret Q as meaning "being queried". Is it not then possible that for every proposition, it and only it is queried? But even if it is not, say because there are propositions that are not expressible in any possible language used by any possible being (Kaplan 1994: 49, n. 11), this should not be excluded by *logic*.

[42] That is, it is not true in any model. For if were true, that model would give the following conditions to sentences containing the intensional operator Q (where w is a world, f an assignment of values to variables and I an assignment Carnapian intensions (functions from possible worlds to sets of propositions) to non-logical constants):

$$f \text{ satisfies } Q\phi \iff \{w' \in W \mid f \text{ satisfies } \phi \text{ in } w'\} \in I(Q)(w)$$

The problem now is that we may deduce from this truth-clause the existence of a bijective correspondence between the sets of elements in the individual domain and some of the elements of that domain, which goes against Cantor's Theorem.

with the following sentence:[43]

(23) $\forall p\ (Qp\ \rightarrow\ \neg p)$

C. Anthony Anderson (2001) has shown that the following sentence, which he calls "Epimenides' Theorem", is a theorem of any formulation of the propositional calculus where the individual variables range over the natural numbers and the function-variables range over functions from natural numbers to natural numbers:

(24) $\forall p\ (((p = \forall s(Qs \rightarrow \neg s))\ \rightarrow\ \neg\forall s(Qs \leftrightarrow p = s))$

(24) saying that every proposition p is such that if p is the proposition that all Q's are untrue (as is the proposition expressed by (23) above), then it is not the unique Q proposition. This shows that the second, Liar-like Kaplan paradox is not due to the contentious modelling of propositions as sets of possible worlds, but to the underlying conception of words as points, entities which are quantified over in the second-order version of propositional modal logic. (24), then, gives us another reason to be suspicious about our modelling tools.[44]

The specific problems with the modelling of situation theory in ZFC soon became to be recognised as symptoms of a more general mismatch of the latter and classical set theory. Commenting, in 1986, on the problem (mentioned in fn. 17 in ch. 3, p. 84) that a proper class of events is incompatible with a given act of seeing, Barwise considered ZFC as a trap he and Perry had fallen into:

> "If you start with the cumulative picture of sets as collections, and then try to model various real world situations by imposing structure on sets, you may well get into trouble." (Barwise 1986e: 178)

The possibility of doing without the axiom of foundation was already considered in *Situation and Attitudes* (Barwise and Perry 1983: 95), but then eschewed in favour of Kripke-Platek set theory with urelements which does not allow for infinite collections of the primitives. This was done in an exploratory spirit, without any claim to empirical adequacy.[45] It was clear from the beginning, however, that some situations are circular, at least prima

[43] Instantiate (22) with (23), i.e. assume that (23) is the only sentence/proposition of which Q holds. Then if (23) is true and we call the sentence expressed by it "s", it is false (for we have $Qs \rightarrow \neg s$), and if it is false, there is a true sentence of which Q holds, i.e. (23) is true.

[44] Another, much simpler example is $\exists! P\ \forall x\ Px$ which translates into $\exists! p\ \Box p$, i.e. that there is only one necessary proposition.

[45] Cf.: "...it interests us to see just where in language we have to abandon this constraint metatheory [KPU] and assume the existence of infinite objects." (Barwise and Perry 1983: 52)

facie. Even without commitment to the well-foundedness of the constituent-of relation on situations, however, some restrictions on the construction of abstract situations are needed, as Lindström (1991: 752) pointed out: for if we allow, e.g., for the property of a situation being factual, it follows, assuming AFA, that there is a situation:

$$e := \quad \mathsf{at}\ l_u : \quad \mathsf{factual}, \dot{e}; \mathsf{no}$$

which is factual iff it is not.

Soon after 1983, Barwise reinforced the case for allowing for circular situations (cf. e.g. Barwise 1986e: 194f.).[46] Not only is non-wellfounded set theory needed to model self-referential situations, but the very existence of (and hence the theoretical need for) non-wellfounded sets may be seen to follow from the existence of such self-referential situations. Barwise, in Barwise (1986e), opposed to the standard cumulative picture of sets as collections of elements a picture of sets as *forgetful situations.* This gives us a rather new and – in my view – exciting perspective on set theory in general: sets are no longer abstract tools on the modelling level, implicitly thought of in instrumentalistic or structuralistic terms, but enter right into the intended domain of application of the theory to be developed.[47] Even if the existence of non-wellfounded sets may appear ill-motivated and the usefulness of circular graphs dubious, it is clear that there are self-referential pieces of informations, called 'hyperinfons' in situation theory. Their study became soon one of the focal points of situation theory and the stimulation of interest in 'circular phenomena' may be seen to be one of its lasting effects.[48]

When Barwise got interested in non-wellfounded sets to model real circular situations and considered ZFC/AFA as background theory for situation semantics, he not just

[46] He later argued at length that circular situations are needed for his favourite account of common knowledge, according to which two agents a and b have common knowledge that p iff there is a 'shared situation' s such that $s \models p$, $s \models a$ knows that p, $s \models b$ knows that p (Barwise 1988: 203). If we were to buy into Devlin's idiosyncratic self-referential account of seeing, we would have many other examples of circular situations. I do not, however, see any good reason for doing so, for we are clearly not – pace Devlin – seeing our own visual experiences. Cf.: "The visual experience, v, figures in its own external content: it is part of the external content of a visual experience, *that* it is caused by what is seen." (Devlin 1991: 196) Apart from its inherent implausibility, this proposal has the drawback to make any instance of seeing (at least partly) an instance of seeing that, a type of experience we are rather reluctant to ascribe to language-incompetent and conceptually unsophisticated creatures.

[47] Barwise goes even further than that and claims that set theory is but a special case of the general theory of situations: "By using the same symbol for the membership relation and the constituent relation we are reflecting our prejudice that the former is just a special case of the latter." (Barwise 1986e: 189)

[48] Cf.: "The elucidation of the structure of hyperinfons and other 'circular phenomena' is one of Situation Theory's main achievements." (Seligman and Moss 1997: 252).

changed one "empirically plausible" model of reality for another. He instead adopted a pluralistic stance, allowing different set theories to be useful modelling tools for different purposes.[49] The mathematically and philosophically interesting question then becomes: what is invariant among different modelling tools? The 1997 theory, covering the classification of elements by sets of which they are members as a special case, may also be seen as a (partial) answer to this question.

Summing up: one part of the general problem with set theory is that the choice of a particular set theory, be it Kripke-Platek in *Situations and Attitudes* (Barwise and Perry 1983: 52) or non-wellfounded set theory in later works, inevitably brings with it some extra constraints, ill-suited and extraneous to the modelling purposes at hand. The other part of the problem is that *any* set theory seems bound to conceive of the domain it models in *unsituated*, *non-local* terms: the danger then is the same as with possible worlds, i.e. to succumb to a temptation to take features of the modelling instruments for features of what is modelled, thereby misconstruing the nature of situated and essentially local human activities as reasoning, meaning and conveying information.[50]

There is a third, even more general, worry pertaining to the very project of modelling situation theory within set theory. Situation theory is, or at least has the ambition to become, a theory about what there is. The theory of information flow, e.g., should eventually provide us with a theory about information flow between real situations, objects in (and making up) the world. It is certainly awkward to take such situations to be *sets*: information flow between sets, if it existed, would not be of any particular interest. Taking set-theory to be the principal modelling device, then, brings with it a constant danger of confusing model and reality. Take, e.g., situations. While Barwise and Perry are careful to distinguish between real, factual and possible situations (the first are entities in the world, the latter two are special kinds of sets), Devlin (1991: 33) takes situations to be real entities and then introduces abstract situations as sets of infons, thereby introducing a basic ontological schism between the two. This not only forces him to identify indistinguishable abstract situations,[51] but allows for the weird question how, when an abstract situation becomes actual (as when Sam hopes that he wins the lottery and then actually wins it), a set of infons is somehow transformed into an entity located at some point in space-time.

[49] He did not consider such a pluralism just a useful methodological position, but instead claimed empirical adequacy for such a "situated set theory" itself (cf. Barwise 1989b: 291).

[50] Barwise was quite explicit about this from 1989 on: "...the set-theoretical modeling of properties and relations by sets of tuples is inadequate in a mathematical model of meaning. Hence the emphasis on properties and relations, not just their extensions, in situation theory." (Barwise 1989e: 294)

[51] Devlin acknowledges this much and says that "to all intents and purposes we have an 'extensionality principle'." (Devlin 1991: 72)

4.8.6 What to do next

Given the different points elaborated above, it at least seems worthwhile (if not necessary) to investigate whether a different and more direct approach might prove fruitful. This is what Barwise and Seligman (1997) tried in *Information Flow – The Logic of Distributed Systems*. Unfortunately, this theory is very abstract and it is difficult to see how exactly it was meant to apply to all the issues brought up.[52] I will ow present this theory, try to stress its general and distinguishing features and make some comparisons with epistemic modal logic. In particular, I will investigate in what sense it may be seen as a *generalisation* of the latter. A word of caution before we start: my interest in the theory will be primarily conceptual, the focus on the elucidation it may provide us with of the nature of information and on the mathematical features of the theory. It is clearly to early to judge the merits of potential applications.[53]

---

[52]Oliver Lemon shares this view: "Evaluating the theoretical merit of such novel work [as done in *Information Flow*, the book he reviews] presents serious problems, for scientific progress has most often been the result of tackling some recognised problem in an established discipline, rather than of theory-building (which is the nature of this work). It is not clear, for example, whether there is any central problem that the theory is supposed to tackle. (What are the requirements on a satisfactory theory of information flow? – How do we know when we've got one?) This lack of philosophical clarity complements the definition-rich flavour of the technical work – many formal structures and operations are introduced to model information flow, but the reader is left pondering the necessity of the particular formalism presented." (1998: 400)

[53]Such applications, in design theory and refinements of the Standard Upper Ontology have been made (cf. e.g. the unpublished papers by Robert Kent, "A KIF Formalization for the IFF Category Theory Ontology" Kent (2001), or by Yuzuru Kakuda / Makoto Kikuchi, "Topology on Classifications in Abstract Design Theory" Kakuda and Kikuchi (2001). Even if such applications may some day prove useful, it strikes me as a grotesque exaggeration to claim, as Devlin does, that "to survive [sic!] in the information society, anyone in a position of responsibility or authority – at any level – will need to understand what information is." (Devlin 2001: 12)

Chapter 5

Representing meaning by classifications

5.1 Classifications

5.1.1 Token meaning

Information involves both bearers and contents. It is the bearers that carry the information; the contents are the pieces of information carried by the bearers. Bearers are particulars, individual spatio-temporal objects, whereas contents are repeatable and abstract objects used to classify the bearers. The crucial relation connecting the two, then, is one of *classification*: information bearers are classified according to the informational content they carry. This relation may in special cases taken to be one of *meaning*, the bearers' meaning the pieces of information they are classified with. In general, however, the relation is one of *relative indication*: relative to a given classification, a bearer a indicates a piece of information α. Following Jon Barwise and Jerry Seligman (1997), let us start with the following definition, where we call bearers "tokens" and contents "types".

Definition 5.1.1 (Classifications). *A* classification *is a triple* $A = \langle \mathsf{tok}(A), \mathsf{typ}(A), \models_A \rangle$ *of a class of tokens* $\mathsf{tok}(A)$*, a class of types* $\mathsf{typ}(A)$ *and a binary modelling relation* $\models_A \subset \mathsf{tok}(A) \times \mathsf{typ}(A)$*.*

If $a \models_A \alpha$ we say that *a indicates* α with respect to A, or that *a is of* type α (more precisely: that a is modelled as being of type α in classification A). This reflects the fact that the same physical objects, e.g. the same marks on a piece of paper, may mean different things to different people or in different contexts. There just is no such thing as *absolute* informational content: one and the same token, within different classifications, may be of different and mutually exclusive types.

Within one classification, tokens differ in what they indicate. Some pieces of information may however be indicated by all the tokens. They are, so to say, already given to us at the level of the classification and do not have to be 'anchored' in specific tokens. Such universal types may be called *valid* within or with respect to a given classification:

Definition 5.1.2 (Validity). *If all the tokens of a classification A are of some type $\alpha \in \mathsf{typ}(A)$, α is called* valid *in A.*

Information is not tied to particular classifications. We can, e.g., embed a classification into another, more expressive, one. The basic means to do this is delivered by the following definition:

Definition 5.1.3 (Infomorphisms). *An* infomorphism *$f : A \rightleftarrows B$ from A to B is a contravariant pair of functions $f = (f^\wedge, f^\vee)$ satisfying the fundamental property of infomorphisms:*

$$\text{(fund)}\quad f^\vee(b) \models_A \alpha \iff b \models_B f^\wedge(\alpha) \qquad \textit{for all } b \in \mathsf{tok}(B) \textit{ and for all } \alpha \in \mathsf{typ}(A)$$

Whenever it is clear whether f means $f^\wedge$ or $f^\vee$, we will drop the superscripts.

Infomorphisms are the morphisms of the category **Class** of all classifications and the basic means of information flow (see sct. 5.2.2 below). Infomorphism allow us to talk of systems of connected classifications with a common "core", i.e. channels:

Definition 5.1.4 (Channels). *A* channel *is an indexed family of infomorphisms $\mathsf{C} = \{f_i : A_i \rightleftarrows C\}_{i \in I}$.*

We may also define subclassifications in the following way:

Definition 5.1.5 (Subclassifications). *A classification B is a* subclassification *of another classification A (denoted by "$B \sqsubseteq A$") iff the identity function $(\mathbf{1}_{\mathsf{typ}(A)}, \mathbf{1}_{\mathsf{tok}(B)})$ is an infomorphism $A \rightleftarrows B$.*

Subclassifications inherit all the valid types of their parents:

Theorem 5.1.6. *If $B \sqsubseteq A$, then all the types valid in A are valid in B.*

PROOF Suppose $B \sqsubseteq A$. It follows that $\mathsf{typ}(A) \subset \mathsf{typ}(B)$, $\mathsf{tok}(B) \subset \mathsf{tok}(A)$ and the modelling relation agrees on types and tokens common to both. □

Within a classification, a token is characterised by the types classifying it, whereas types may be taken extensionally, as nothing but the grouping together of their tokens. So we adopt the following definitions:

Definitions 5.1.7 (Coextensionality and indistinguishability). *The* description *of a token a is the set* $\mathsf{typ}(a) = \{\alpha \in \mathsf{typ}(A) \mid a \models_A \alpha\}$. *The* extension *of a type α is the set* $\mathsf{tok}(\alpha) = \{a \in \mathsf{tok}(A) \mid a \models_A \alpha\}$. *Two types α_1 and α_2 are* coextensive *($\alpha_1 \sim_A \alpha_2$) if* $\mathsf{tok}(\alpha_1) = \mathsf{tok}(\alpha_2)$. *Two tokens a_1 and a_2 are* indistinguishable *($a \sim_A a$) if* $\mathsf{typ}(a_1) = \mathsf{typ}(a_2)$.

As before, we have to bear in mind that coextensiveness and indistinguishability are notions relative to some given classification: tokens indistinguishable in one classification may well be distinguishable by another one.

By $\mathsf{typ}(G) := \bigcap\{\mathsf{typ}(s) \mid s \in G\}$ and $\mathsf{typ}(\Gamma) := \bigcap\{\mathsf{tok}(\sigma) \mid \sigma \in \Gamma\}$, we extend these definitions to sets of types and tokens respectively. Whenever distinctions between tokens or types are not relevant, we may take invariants and construct the corresponding quotients:

Definitions 5.1.8 (Invariants and dual invariants). *An* invariant *I to a classification A is a pair $I = \langle \Sigma, R \rangle$ of a set $\Sigma \subset \mathsf{typ}(A)$ and a binary relation $R \subset \mathsf{tok}(A) \times \mathsf{tok}(A)$ satisfying:*

$$aRb \qquad \Longrightarrow \qquad \forall \alpha \in \Sigma : a \models_A \alpha \leftrightarrow b \models_A \alpha$$

A dual invariant *J to a classification A is a pair $J = \langle A, R \rangle$ of a set $A \subset \mathsf{tok}(A)$ and a binary relation $R \subset \mathsf{typ}(A) \times \mathsf{typ}(A)$ satisfying:*

$$\alpha R \beta \qquad \Longrightarrow \qquad \forall a \in A : a \models_A \alpha \leftrightarrow a \models_A \beta$$

Invariants are thus given by relations on tokens or types which imply indistinguishability or coextensionality respectively.

Suppose we are given an arbitrary binary relation R on $\mathsf{tok}(A)$. We then denote by "$\equiv_R$" the smallest equivalence relation on $\mathsf{tok}(A)$ containing R and by "$[a]_R$" the $\equiv_R$-equivalence class containing $a \in \mathsf{tok}(A)$ and adopt the following definition:

Definition 5.1.9 (Quotients and dual quotients). *The* quotient *A/I of (a classification) A by (an invariant) $I = \langle \Sigma, R \rangle$ is the classification with:*

- $\mathsf{tok}(A/I) = \{[a]_R \mid a \in \mathsf{tok}(A)\}$
- $\mathsf{typ}(A/I) = \Sigma$
- $[a]_R \models_{A/I} \alpha \quad :\Longleftrightarrow \quad a \models_A \alpha$

The dual quotient A/J *of (a classification)* A *by (a dual invariant)* $J = \langle A, R\rangle$ *is the classification with:*

- $\mathsf{tok}(A/J) = A$
- $\mathsf{typ}(A/J) = \{[\alpha]_R \mid \alpha \in \mathsf{typ}(A)\}$[1]
- $a \models_{A/J} [\alpha]_R \quad :\Longleftrightarrow \quad a \models_A \alpha$

As coextensionality and indistinguishability are equivalence relations, we may form the corresponding quotients to get separate or extensional classifications. In separate classifications, tokens which indicate the same (and thus have the same 'meaning' within the classification) are identified. Types in extensional classifications are uniquely characterised by their tokens:

Definition 5.1.10 (Separateness and extensionality). *The classification A is* separated *iff $a_1 \sim_A a_2$ implies $a_1 = a_2$. A is* extensional *iff $\alpha_1 \sim_A \alpha_2$ implies $\alpha_1 = \alpha_2$.*

By taking indistinguishability of tokens as the equivalence relation of our invariant, we get the separated quotient $\mathsf{Sep}(A)$ of A. Taking coextensionality of types as the equivalence relation of our dual invariant, we get the (dual) extensional quotient $\mathsf{Ext}(A)$ of A.

Theorem 5.1.11. *Any classification isomorphic to $\mathsf{Sep}(A)$ for some classification A is separated.*[2]

Proof Suppose $\mathsf{Sep}(A) \rightleftarrows B$ is an infomorphism. By (fund), we have:

$$f^{\vee}(b) \models_{\mathsf{Sep}A} \alpha \quad \Longleftrightarrow \quad b \models_B f^{\wedge}(\alpha)$$

By the one-one correspondence of types, it follows from $\mathsf{typ}(b_1) = \mathsf{typ}(b_2)$ that $\mathsf{typ}(f^{\vee}(b_1)) = \mathsf{typ}(f^{\vee}(b_2))$. Because $\mathsf{Sep}(A)$ is separated, this implies $f^{\vee}(b_1) = f^{\vee}(b_2)$, and, by the injectivity of $f^{\vee}$, $b_1 = b_2$. $\square$

[1] "$[\alpha]_R$" denotes the $\equiv_R$-equivalence class containing $\alpha \in \mathsf{typ}(A)$ and "$\equiv_R$" the smallest equivalence relation on $\mathsf{typ}(A)$ containing R

[2] When Barwise first introduced the term "infomorphism", he discussed the issue of whether situation theory should identify situations supporting the same infons. In our terminology, his "choice 5" (Barwise 1989a: 264) corresponds to the question whether situation theory has any use for non-separated classifications. It seems clear that the answer is in the affirmative.

Infomorphisms may be said to respect invariants, the biggest respected invariants and dual invariants being kernel and cokernel respectively:

Definition 5.1.12. $f : A \rightleftarrows B$ respects *an invariant* $I = \langle \Sigma, R \rangle$ *on* B *if:*

$$\begin{aligned} \forall \alpha \in \mathsf{typ}(A) &\quad : \quad f(\alpha) \in \Sigma \\ \forall b_1, b_2 \in \mathsf{tok}(B) &\quad : \quad b_1 R b_2 \rightarrow f(b_1) = f(b_2) \end{aligned}$$

Such an infomorphism $f : A \rightleftarrows B$ respects *a dual invariant* $J = \langle A, R \rangle$ *on* A *if:*

$$\begin{aligned} \forall b \in \mathsf{tok}(B) &\quad : \quad f(b) \in A \\ \forall \alpha_1, \alpha_2 \in \mathsf{typ}(A) &\quad : \quad \alpha_1 R \alpha_2 \rightarrow f(\alpha_1) = f(\alpha_2) \end{aligned}$$

Quotients and dual quotients allow us to factorise infomorphisms uniquely:

Theorem 5.1.13. *If* $I = \langle \Sigma, R \rangle$ *is an invariant and* $f : B \rightleftarrows A$ *an infomorphism which respects* I*, there is a unique infomorphism* $g : B \rightleftarrows A/I$ *such that the following diagram commutes, where* $\tau_I : A/I \rightleftarrows A$ *is the canonical quotient infomorphism, which is the inclusion on types and maps each tokens on its* R *equivalence class:*

$$\begin{tikzcd} A/I \arrow[rr, "\tau_I"'] & & A \\ & \circlearrowleft & \\ B \arrow[uu, "\exists!\ g"] \arrow[uurr, "f"'] & & \end{tikzcd}$$

If $J = \langle A, R \rangle$ *is a dual invariant and* $f : A \rightleftarrows B$ *an infomorphism which respects* J*, there is a unique infomorphism* $g : A/J \rightleftarrows B$ *such that the following diagram commutes, where* $\tau_J : A \rightleftarrows A/J$ *is the canonical quotient infomorphism, which is the inclusion on tokens and maps each type on its* R *equivalence class:*

$$\begin{tikzcd} A \arrow[rr, "\tau_J"'] \arrow[ddrr, "f"'] & & A/J \arrow[dd, "\exists!\ g"] \\ & \circlearrowleft & \\ & & B \end{tikzcd}$$

Proof

[$1^s t$ claim:] Define g as the identity function on types and on tokens by $g([a]_R) := f(a)$. g is well-defined because f respects I, i.e. because $[a]_R = [b]_R$ implies aRb and hence $f(a) = f(b)$.

[$2^n d$ claim:] Define g as the identity function on tokens and on types by $g([\alpha]_R) := f(\alpha)$. g is well-defined because f respects J, i.e. because $[\alpha]_R = [\beta]_R$ implies $\alpha R \beta$ and hence $f(\alpha) = f(\beta)$. $\square$

We will often use (e.g. in the proof of theorem (5.2.7)) the fact that any infomorphism respecting some invariant may be lifted to the corresponding quotient.

Infomorphisms are embeddings in the sense that they do not allow us to draw distinctions we could not draw before:

Theorem 5.1.14. *If $f : A \rightleftarrows B$ is an infomorphism, $f^\vee(b) \sim_A f^\vee(b')$ whenever $b \sim_B b'$.*

Proof Suppose $b \sim_B b'$. We have to show $\mathsf{typ}(f^\vee(b)) = \mathsf{typ}(f^\vee(b'))$. Whenever $\alpha \in \mathsf{typ}(f^\vee(b))$, $f^\vee(b) \models \alpha$, and, by (fund), $b \models f^\wedge(\alpha)$, thus $f^\wedge(\alpha)$ is in $\mathsf{typ}(b)$, which is $\mathsf{typ}(b')$ and so $b' \models f^\wedge(\alpha)$, which, again by (fund), means $\alpha \in \mathsf{typ}(f^\vee(b'))$. The converse is similar. $\square$

Only type surjective infomorphisms, however, preserve indistinguishability of tokens:

Theorem 5.1.15. *If $f : A \rightleftarrows B$ is a type surjective infomorphism, $a \sim_A a'$ implies $b \sim_B b'$ for any $f^\vee(b) = a$ and $f^\vee(b') = a'$.*

Proof Suppose $a \sim_A a'$. Suppose we have $b, b' \in \mathsf{tok}(B)$ with $f^\vee(b) = a$ and $f^\vee(b') = a'$. By (fund), b and b' are of the same types in $f^\wedge(\mathsf{typ}(A))$. Because $f^\wedge$ is surjective, they are of the same types in $\mathsf{typ}(B)$ and thus indistinguishable in B. $\square$

We will now study some examples of classifications.

5.1.2 Truth classifications

Our first example are truth classifications, where we take the sentences of a given language as our types and the corresponding structures as our tokens:

Definition 5.1.16 (Truth classifications). *A* truth classification *A_L of a first-order language L is a triple $A_L = \langle \mathsf{Struct}(L), \mathsf{Form}(L), \models \rangle$ of the formulae of L, the L-structures and $\models$ defined by $M \models \phi :\Leftrightarrow \phi$ is true in M.*

In truth classifications, the description of a structure M is more commonly called the theory $\mathsf{Th}(M)$ of M, whereas the extension of a sentence σ is called the model set $\mathsf{Mod}(\sigma)$ of σ. Coextensive sentences have the same model set, indistinguishable structures the same theory. Truth classifications are never separated (as isomorphic structures do not need to be identical) nor extensional (as $\phi \wedge \phi \neq \phi$).

In the case of truth classifications, we can view an infomorphism $f : A \rightleftarrows B$ as an embedding of A into B, consisting of a translation of the types of A into types of B and a corresponding function mapping tokens of B into tokens of A such that (fund) specifies the obvious adequacy condition for the translation. If we take, e.g., the models of Peano arithmetic as our tokens, a suitable language L_{PA} and normal validity and as the tokens of B models of ZFC-set theory, with language L_{ZFC} and normal validity, we get e.g. the following embedding of elementary number theory into ZFC:

Theorem 5.1.17. *The following function f is an infomorphism $PA \rightleftarrows ZFC$: $f^\wedge(\alpha)$ is, for any term t of $\mathcal{L}_{PA}$ defined inductively as follows:*

- *If $t = 0$, then $f^\wedge(t) = \emptyset$.*
- *If $t = S(s)$, then $f^\wedge(t) = f^\wedge(s) \cup \{f^\wedge(s)\}$*
- *If $t = x_i$ for any $i \in \mathcal{I}$, then $f^\wedge(t) = v_i$*
- *If $t = s + 0$, then $f^\wedge(t) = f^\wedge(s)$*
- *If $t = s + S(u)$, then $f^\wedge(t) = f^\wedge(s+u) \cup \{f^\wedge(s+u)\}$*
- *If $t = s \cdot 0$, then $f^\wedge(t) = \emptyset$*
- *If $t = s \cdot S(u)$, then $f^\wedge(t) = f^\wedge(s \cdot u) + f^\wedge(s)$*

We now define the translation of a formula α:

- *If $\alpha \equiv (s = t)$ for two terms s and t, then $f^\wedge(\alpha) = (f^\wedge(t) = f^\wedge(s))$*
- *If $\alpha \equiv (s \neq t)$ for two terms s and t, then $f^\wedge(\alpha) = (f^\wedge(t) \neq f^\wedge(s))$*
- *If $\alpha \equiv (s < t)$ for two terms s and t, then $f^\wedge(\alpha) = (f^\wedge(t) \in f^\wedge(s))$*
- *If $\alpha \equiv (s \not< t)$ for two terms s and t, then $f^\wedge(\alpha) = (f^\wedge(t) \notin f^\wedge(s))$*
- *If $\alpha \equiv \beta \wedge \gamma$ for two formulae β and γ, then $f^\wedge(\alpha) = f^\wedge(\beta) \wedge f^\wedge(\gamma)$*
- *If $\alpha \equiv (\exists x_i < t)\ \beta(x)$ for some term t and $i \in \mathcal{I}$, then $f^\wedge(\alpha) = (\exists v_i \in f^\wedge(t))\ f^\wedge(\beta(x))$*
- *If $\alpha \equiv (\forall x_i < t)\ \beta(x)$ for some term t and $i \in \mathcal{I}$, then $f^\wedge(\alpha) = (\forall v_i \in f^\wedge(t))\ f^\wedge(\beta(x))$*

$f^\vee(M)$ is defined for any model M of ZFC as follows:

- $|f^\vee(M)| := \{\alpha \mid \mathbf{On}(\alpha) \wedge \forall\beta(\mathbf{On}(\beta) \wedge \beta \in^M \alpha \wedge \beta \neq^M \emptyset \to \exists\gamma(\beta = \gamma \cup^M \{\gamma\}))\}$[3]
- $0^{f^\vee(M)} := \emptyset^M$
- $S^{f^\vee(M)}(w) := w \cup^M \{w\}$
- $w +^{f^\vee(M)} \emptyset^M := \emptyset^M$
- $w +^{f^\vee(M)} (v \cup^M \{v\}) := (w +^{f^\vee(M)} v) \cup^M \{w +^{f^\vee(M)} v\}$

- $w \cdot^{f^\vee(M)} \emptyset^M := \emptyset^M$
- $w \cdot^{f^\vee(M)} (v \cup^M \{v\}) := (w \cdot^{f^\vee(M)} v) +^{f^\vee(M)} v$
- $=^{f^\vee(M)} := =^M$
- $<^{f^\vee(M)} := \in^M$

PROOF We have to check that f, so defined, is an infomorphism. So we have to verify (fund):

$$f^\vee(M) \models_{PA} \alpha \quad \Longleftrightarrow \quad M \models_{ZFC} f^\wedge(\alpha)$$

This is done by an induction on the structure of the formula α. □

Every sentence α of number theory is translated into one of set theory (replacing, e.g., numerals by names for the finite von Neumann ordinals). To every structure M of set theory corresponds a structure of number theory, satisfying the PA axioms. It is instructive to see that the "translations" of formulae we chose are tailor-made in order to get the proof of (fund) go through. Any problem in that proof would lead us to change our definition of either $f^\vee$ or $f^\wedge$. This seems to indicate that we in fact have chosen the right adequacy condition (fund).

5.1.3 Lewis classifications

If we just start with some tokens (which we may call "possible worlds"), and take our types to be arbitrary subsets of tokens (which may be called "propositions"), we have what we will call a Lewis classification:

Definition 5.1.18 (Lewis classifications). *A* Lewis classification A_W *of a set* W *of possible worlds is a triple* $A_W = \langle W, \mathcal{P}(W), \models\rangle$ *of a set of possible worlds* W*, a set of propositions* $\mathcal{P}(W)$ *and a binary relation* $\models \subset W \times \mathcal{P}(W)$ *with* $w \models \alpha :\Leftrightarrow w \in \alpha$*.*

Any Lewis classification is separated and extensional: possible worlds are characterised by the propositions which are true in them, while necessary equivalent propositions, i.e. propositions sharing all their members, are identified. Infomorphisms are just functions paired with their inverses, considered as operating on the respective power sets:

[3] $\mathbf{On}(\alpha)$ here abbreviates a long formula in the language of set theory with quantifiers ranging over M (cf. Krivine 1971: 14).

Theorem 5.1.19. *$f : A_W \rightleftarrows A_V$ is an infomorphism between the two Lewis classifications A_W and A_V iff $f^\vee : V \to W$ is any function and $f^\wedge(\alpha) = f^{-1}[\alpha] := \{v \in V \mid f(v) \in \alpha\}$ for any $\alpha \subset W$.*

PROOF This is immediate from (fund) which becomes $f^\vee(v) \in \alpha \iff v \in f^\wedge(\alpha)$. $\square$

Lewis classifications represent epistemic situations in which we not only have complete information, but also access to a complete and unique description of any set of tokens. Moreover, the set of types has a rich internal structure, given by intersection, union and complement. We will later see that Lewis classifications can be seen as a special family of Boolean classifications and then generalise them to frame classifications.

While Lewis classifications have an expressively rich class of types, they do not allow for interesting "theorems". The only type valid in them is the universal proposition W.

5.1.4 Topological classifications

A special case of Lewis classifications are topological spaces:

Definition 5.1.20 (Topological classifications). *A* topological classification *$X_\mathcal{O}$ of a topological space $(X, \mathcal{O})$ (a set X with the family of its open subsets $\mathcal{O} \subset \mathcal{P}(X)$) is a triple $X_\mathcal{O} = \langle X, \mathcal{O}, \in \rangle$ of a space X, its open sets $\mathcal{O}$ and membership as the classification relation.*

Any topological classification is extensional. If $(X, \mathcal{O})$ is T_0 (or even Hausdorff), the topological classification $X_\mathcal{O}$ is separated. With topological classifications, infomorphisms are just continuous functions.

Theorem 5.1.21. *$f : X_{\mathcal{O}_1} \rightleftarrows Y_{\mathcal{O}_2}$ is an infomorphism between the two topological classifications $X_{\mathcal{O}_1}$ and $Y_{\mathcal{O}_2}$ iff $f^\vee : Y \to X$ is a continuous function and $f^\wedge(A) = f^{-1}[A]$ for any $A \subset X$.*

PROOF This holds because inverse images of open sets under continuous functions are open. $\square$

We may not only build topological classifications out of topological spaces, but conversely "topologise" any given classification we start with:

Definition 5.1.22. *Let $A = \langle \mathsf{tok}(A), \mathsf{typ}(A), \models_A \rangle$ be a classification. Then $(\mathsf{tok}(A), \mathcal{O}(\{\mathsf{tok}(\alpha) \mid \alpha \in \mathsf{typ}(A)\}))$ is a topological space, where $\mathcal{O}(Y)$ is the weakest topology containing Y. If A is Boolean, $\{\mathsf{tok}(\alpha) \mid \alpha \in \mathsf{typ}(A)\}$ is a base of this topology and the topological space is zero-dimensional.*

PROOF For the first claim, we have to check for any two subsets U and V of $\{\mathsf{tok}(\alpha) \mid \alpha \in \mathsf{typ}(A)\}$, if $a \in U \cap V$, then there is a $W \subset \{\mathsf{tok}(\alpha) \mid \alpha \in \mathsf{typ}(A)\}$ such that $a \in W$ and $W \subset U \cap V$. So let $a \in \mathsf{typ}(\Phi) \cap \mathsf{typ}(\Psi)$ for $\Phi, \Psi \subset \mathsf{typ}(A)$. But then $a \models \bigcap \Phi \cup \Psi$ and $\bigcap \Phi \cup \Psi \in \mathsf{typ}(A)$.

For the second claim, we have to show that any $U \subset \{\mathsf{tok}(\alpha) \mid \alpha \in \mathsf{typ}(A)\}$ is closed and thus that the clopen sets form a basis. But this just follows from the presence of a negation. □

We will later see how to construct filters on the set of all sequents of the regular theory of some classification, which will give us another way of "topologising" classifications.

5.1.5 Kripke classifications and modal infomorphisms

Given the encoding of Kripke models as epistemic states in section (4.5), we may also define Kripke classifications:

Definition 5.1.23 (Kripke classifications). *A* Kripke classification A_C *of a set* C *of epistemic states is a triple* $A_C = \langle \mathsf{C}, Sent(\mathcal{L}), \models \rangle$ *of a class of epistemic states* C*, the formulae of a modal language* $\mathcal{L}$ *and a binary relation* $\models \subset \mathsf{C} \times \mathsf{Form}(\mathcal{L})$ *defined as follows:*

$$\begin{aligned} w \models p &\;:\Longleftrightarrow\; w(p) = \top \\ w \models \neg\phi &\;:\Longleftrightarrow\; w \not\models \phi \\ w \models \bigwedge \Phi &\;:\Longleftrightarrow\; w \models \phi \quad \forall \phi \in \Phi \\ w \models [a]\phi &\;:\Longleftrightarrow\; w' \models \phi \quad \forall w' \in w(a) \end{aligned}$$

By (4.5.23), Kripke classifications with respect to $\mathcal{L}_\infty$ are separated. Kripke classifications with respect to $\mathcal{L}_{<\omega}$ are not separated when taken to include the class of all epistemic states. They are, however, separated with respect to the class $HF^1[\mathbb{P}]$, the largest class of hereditary finite sets with $\mathbb{P}$ as urelements (Barwise and Moss 1996: 137). Kripke classifications are never extensional.

Given a Kripke model $M = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}} \rangle$, we may turn it into a classification A_M by taking A_M to be the Kripke classification of the set of those (uniquely determined) epistemic states which are solutions of all its worlds (M, w) $(\forall w \in W)$. Validity in the original Kripke model M then corresponds with validity in the corresponding Kripke classification A_M:

Theorem 5.1.24. *If* M *is a Kripke model and* A_M *the corresponding Kripke classification,* *we have for any modal formula ϕ:*

$$M \models \phi \quad \Longleftrightarrow \quad A_M \models \phi$$

Proof We prove this by induction on ϕ.

- If $\phi \equiv p$ for some $p \in \mathbb{P}$, then $M \models \phi$ iff $\pi(p) = W$. So we have, for all $w \in W$, $\delta_w(p) = \top$ for the unique solution δ_w of (M, w), which means that all the epistemic states in our classification are of the type p. If, conversely, all the epistemic states are of type p, then $\delta_w(p) = \top$ for all solutions δ_w, which means $w \in \pi(p)$ for all $w \in W$.
- If $\phi \equiv \neg\psi$, then $M \not\models \psi$, which holds, by the induction hypothesis, iff $A_M \not\models \psi$.
- If $\phi \equiv \bigwedge \Phi$, then $M \models \psi$ for all $\psi \in \Phi$. By the induction hypothesis, this holds iff $A_M \models \psi$ for all $\psi \in \Phi$, which is, by definition of $\wedge$, $A_M \models \phi$.
- If $\phi \equiv [a]\psi$, we have, by the induction hypothesis, $v \models \psi$ $\forall v \in W$ such that $wR_a v$ iff $\delta_v \models \phi$ $\forall \delta_v \in \delta_w(a)$. The result follows from the respective definitions of $M \models [a]\psi$ and $A_M \models [a]\psi$.

□

Whenever we have a Kripke classification A_C, we may construct a corresponding Kripke model M_C by taking W to be our set of epistemic states C, defining π by $\pi(p) := \{v \in \mathsf{C} \mid v(p) = \top\}$ and defining the relations R_i by $wR_i v :\Leftrightarrow v \in w(a)$. Given (4.6.10), M_C is unique up to bisimulation. Validity in the Kripke classification coincides with validity in the corresponding Kripke model:

Theorem 5.1.25. *If A_C is a Kripke classification and M_C the corresponding Kripke model, we have for any modal formula ϕ:*

$$A_C \models \phi \quad \Longleftrightarrow \quad M_C \models \phi$$

Proof We show this by induction on ϕ.

- If $\phi \equiv p$ for some $p \in \mathbb{P}$. Then $v \models p$ for all $v \in \mathsf{C}$ iff $\pi(p) = \mathsf{C}$.
- If $\phi \equiv \neg\psi$, then $M_C \models \phi$ iff $M_C \not\models \psi$, which holds, by the induction hypothesis, iff $A_C \not\models \psi$, i.e. iff $A_C \models \phi$.
- If $\phi \equiv \bigwedge \Phi$, then $M_C \models \psi$ for all $\psi \in \Phi$. By the induction hypothesis, this holds iff $A_C \models \psi$ for all $\psi \in \Phi$, which is, by definition of $\wedge$, $A_C \models \phi$.
- If $\phi \equiv [a]\psi$, we have, by the induction hypothesis, $v \models \psi$ $\forall v$ such that $wR_a v$ iff $\delta_v \models \phi$ $\forall \delta_v \in \delta_w(a)$. The result follows from the respective definitions of $M_C \models [a]\psi$ and $A_C \models [a]\psi$.

$\square$

To get a handle on different interesting families of Kripke classifications, we define some closed classes of epistemic states (where such a class C is closed iff $v \in \mathsf{C}$ whenever $w \in \mathsf{C}$ and $v \in w(a)$ for some $a \in \mathcal{A}$) as follows:

Definitions 5.1.26. *We define the following closed classes of epistemic states:*

- *The class of all* positively introspective *epistemic states* PI *is the largest closed class such that, for each* $w \in \mathsf{PI}$ *and* $a \in \mathcal{A}$, $v \in w(a)$ *implies* $v(a) \subset w(a)$.
- *The class of all* negatively introspective *epistemic states* NI *is the largest closed class such that, for each* $w \in \mathsf{NI}$ *and* $a \in \mathcal{A}$, $v \in w(a)$ *implies* $w(a) \subset v(a)$.
- *The class of all* introspective *epistemic states* I *is* $\mathsf{PI} \cap \mathsf{NI}$.
- *The class of all* reflective *epistemic states* T *is the largest closed class such that, for any* $a \in \mathcal{A}$, $w \in \mathsf{T}$ *implies* $w \in w(a)$.

By the familiar soundness and completeness proofs of multi-modal logics and the correspondence between Kripke classifications and (bisimulation-equivalence classes of) Kripke models, we have:

Theorem 5.1.27. *The following correspondances between Kripke classifications and modal logics hold for any modal language* $\mathcal{L}$*:*

- **(K)** *The* $\mathcal{L}$*-formulae valid in the Kripke classification* A_M *with respect to all epistemic states* C *are exactly the theorems of the modal logic* **K**.
- **(T)** *The* $\mathcal{L}$*-formulae valid in the Kripke classification* A_T *with respect to all reflective epistemic states* T *are exactly the theorems of the modal logic* **T**.
- **(S4)** *The* $\mathcal{L}$*-formulae valid in the Kripke classification* A_{TPI} *with respect to all reflective and positively introspective epistemic states* TPI *are exactly the theorems of the modal logic* **S4**.
- **(S5)** *The* $\mathcal{L}$*-formulae valid in the Kripke classification* A_{TI} *with respect to all reflective and introspective epistemic states* TI *are exactly the theorems of the modal logic* **S5**.
- **(K45)** *The* $\mathcal{L}$*-formulae valid in the Kripke classification* A_I *with respect to all introspective epistemic states* I *are exactly the theorems of the modal logic* **K45**.

PROOF

K: **K** is sound and complete with respect to validity in the class of all Kripke models. Validity in this class is validity in the Kripke classification A_C, because every Kripke world corresponds to an epistemic state which, within the classification, is of all the types which are true in the Kripke world and vice versa. We have already shown that the solution of a Kripke world is an epistemic state and that it is of all the formulae true in the Kripke world (4.6.10). Given an epistemic state we can construct its picture, i.e. a Kripke world validating the same sentences and which is unique up to bisimulation in the way described in the proof of (4.6.10).

T: Given that **T** is sound and complete with respect to validity in the class of all Kripke models with reflexive accessibility relations, we have to show that every Kripke world with reflexive accessibility relations has a reflective epistemic state as its solution (establishing soundness) and that every reflective epistemic state has a picture where the accessibility relations are reflexive (establishing completeness). For soundness, suppose $(\langle W, \pi, \{R_a\}_{a\in\mathcal{A}}\rangle, w)$ is a Kripke world with reflexive $\{R_a\}_{a\in\mathcal{A}}$ and δ_w is its unique solution. For any $a\in\mathcal{A}$ we have wR_aw and thus $\delta_w\in\delta_w(a)$ by the definition of a decoration (4.6.9). So δ_w is reflective. For completeness, suppose that w is a reflective epistemic state and (M, w) one of its pictures ((M, w) is then unique up to bisimulation). For any $a\in\mathcal{A}$, we have $w\in w(a)$ and thus $(w, w)\in R_a$ by the way we construct the picture of an epistemic state.

S4: Given that **S4** is sound and complete with respect to validity in the class of all Kripke models with reflexive and transitive accessibility relations, we have to show that every Kripke world with reflexive and transitive accessibility relations has a reflective and positively introspective epistemic state as its solution (establishing soundness) and that every reflective and positively introspective epistemic state has a picture where the accessibility relations are reflexive and transitive (establishing completeness). For soundness, this follows from the fact that all R_a successors u of any R_a successor v of w (of any v such that $\delta_v\in\delta_w(a)$, where δ is the unique decoration of the Kripke model) are R_a successors of w (i.e. it holds for all u such that $\delta_u\in\delta_v(a)$ that $\delta_u\in\delta_w(a)$). For completeness, transitivity of the accessibility relations follows from the way they are defined on the basis of the epistemic states.

S5: As above, we make use of the soundness and completeness of **S5** with respect to Kripke models with equivalence relations. The solutions are introspective because every R_a successor of a world w has the same R_a successors than w, i.e. for any epistemic state v in $\delta_w(a)$ it holds that $v(a) = w(a)$. Symmetry of the accessibility relations is guaranteed by the fact that whenever $v \in w(a)$, $w \in w(a)$ by reflexivity and thus $w\in v(a)$ by negative introspection.

K45: **K45** is sound and complete with respect to Kripke models with transitive and Euclidean accessibility relations. Euclidity can be seen to correspond to introspection by noting that whenever $v \in u(a)$ $(uR_a v)$ and $w \in u(a)$ (uRw), then $w(a) = u(a)$ by introspection and consequently $v \in w(a)$ (which is wRv); and conversely, given euclidity, $v \in u(a)$ (uRv) entails $u(a) \subset v(a)$ because any $w \in u(a)$ (uRw) is also in $v(a)$ (vRw).

□

What is the relation between the taking of quotients of Kripke classifications and filtration of Kripke models encountered earlier (cf. def. 4.7.8)? Suppose a set of subformula-closed set of formulae of a modal language Σ is given. Any invariant $\langle \Sigma, R \rangle$ will contain an R which is a subset of indistinguishability with respect to Σ. Hence $\equiv_R$, the smallest equivalence relation on the tokens of the Kripke classification containing R, will be such a subset too, indistinguishability with respect to Σ being itself an equivalence relation. So the second condition on filtrations in (4.7.8) is guaranteed. To guarantee the first, we have to assume that Σ is sufficiently expressive, i.e. that it contains, for every $a \in \mathcal{A}$, an $\langle a \rangle \phi$ formula (for some ϕ true in the R_a-successor) true of the first member of any R_a-related pair. Such filtrations are just those quotients where R is not only a subset, but identical to indistinguishability with respect to Σ.

We now see that the quotient-construction on Kripke classifications is a generalisation of filtration in two respects: R may be a subset of Σ-indistinguishability and Σ is not forced to preserve all relational dependencies. The other crucial difference is that quotients, but not filtrations, keep track of the gain in expressive economy: they retain only the types needed to distinguish between the equivalence classes we keep, while filtration does not affect the language in use at all.

Given th. 5.1.27, we can mimic the familiar embeddings of these modal logics into each other by infomorphisms:

Theorem 5.1.28. $(\mathbf{1}_{\mathsf{typ}(A)}, \mathbf{1}_{\mathsf{tok}(B)})$ *is an infomorphism between Kripke classifications A and B in all the following cases:*

- *from $A = A_C$ to A_T and to A_I $(B \in \{A_T, A_I\})$;*
- *from $A = A_T$ to A_{TPI} and to A_{TI} $(B \in \{A_{TPI}, A_{TI}\})$;*
- *from $A = A_I$ to A_{TPI} and to A_{TI} $(B \in \{A_{TPI}, A_{TI}\})$; and*
- *from $A = A_{TPI}$ to $B = A_{TI}$*

Proof This follows from the fact that we can derive the axioms of any weaker in all of the stronger logics and from (5.1.27). □

In terms of subclassifications (cf. definition (5.1.5)), we thus have that A_C (**K**) is a subclassification of A_T (**T**) and A_I (**K45**) (and, by the transitivity of the subclassification ordering, of A_{TPI} / **S4** and A_{TI} / **S5**).

What are modal infomorphisms, i.e. infomorphisms between Kripke classifications, in terms of Kripke models? As a first step at an answer, observe that whenever an infomorphism is the identity function on the types, (fund) reduced to indistinguishability between Kripke worlds:

$$f^{\vee}(w) \models \phi \iff w \models f^{\wedge}(\phi)$$

So, by (4.5.23) the contravariant function $(\mathbf{1}_{\mathsf{typ}(C)}, f^{\vee}) : A_C \rightleftarrows A_D$ is an infomorphism iff $f^{\vee} : M_D = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}} \rangle \to M_C = \langle W', \pi', \{R'_i\}_{i \in \mathcal{A}} \rangle$ between the corresponding Kripke models is a bisimulation. For the "only if" direction, however, we do not need something as strong as bisimulation: a bounded morphism (4.7.4), or any other satisfaction preserving construction:

Theorem 5.1.29. *Let A_C and A_D be two Kripke classifications based on the same set of agents and $f^{\wedge} : \mathsf{typ}(A_C) \to \mathsf{typ}(A_D)$ be the identity function between their types. If $f^{\vee} : M_D = \langle W, \pi, \{R_i\}_{i \in \mathcal{A}} \rangle \to M_C = \langle W', \pi', \{R'_i\}_{i \in \mathcal{A}} \rangle$ is a bounded morphism between the corresponding Kripke models, then $(f^{\wedge}, f^{\vee}) : A_C \rightleftarrows A_D$ is an infomorphism.*

PROOF Suppose $f^{\vee} : M_D \to M_C$ is a bounded morphism. We have to check (fund), that is:

$$f^{\vee}(w) \models \phi \iff w \models \phi$$

We do so by formula induction on ϕ:

- Let $\phi \in \mathbb{P}$. Then $f^{\vee}(w) \models \phi \iff w \models \phi \iff w \models f^{\wedge}(\phi)$ by the first clause in (4.7.4) and the fact that $f^{\wedge}$ is identity.
- Let $\phi = \neg\psi$. Then $f^{\vee}(w) \models \psi \iff w \models f^{\wedge}(\psi)$ by the induction hypothesis. $f^{\vee}(w) \models \phi \iff w \models f^{\wedge}(\phi)$ follows from the definition of $\neg$.
- Let $\phi = \bigwedge \Phi$. Then $f^{\vee}(w) \models \psi$ for all $\psi \in \Phi$ iff $w \models \psi$ for all $\psi \in \Phi$ by the induction hypothesis. $f^{\vee}(w) \models \phi \iff w \models f^{\wedge}(\phi)$ follows from the definition of $\bigwedge$.
- Let $\phi = [a]\psi$ for some $a \in \mathcal{A}$. For the left-to-right direction, suppose $f^{\vee}(w) \models [a]\psi$ and $wR_a v$. We have to show that $v \models f^{\wedge}(\psi)$. From the second clause in (4.7.4), we know that $f^{\vee}(w) R'_a f^{\vee}(v)$ and hence that $f^{\vee}(v) \models \psi$. By the induction hypothesis, we conclude that $v \models f^{\wedge}(\psi)$. Because v was an arbitrary R_a-successor of w, we have $w \models \Box f^{\wedge}(\psi)$, which is all we have to show. For the right-to-left direction, suppose

$w \models \Box f^\wedge(\psi)$ (which is equivalent to $w \models f^\wedge(\Box\psi)$) and $f^\vee w R'_a v'$. By the third clause in (4.7.4), there has to be a $v \in W'$ such that $f^\vee(v) = v'$. So $v \models f^\wedge(\psi)$. By the induction hypothesis, $f^\vee(v) \models \psi$ as well, so $f^\vee w \models \Box\psi$ for v' was an arbitrary R_a-successor.

□

5.1.6 Boolean classifications

We have defined classifications for arbitrary sets of tokens and types, which a priori have no further structure. We will now see how we can introduce (infinitary) Boolean compounds of some types we already have.

Definition 5.1.30. *Given a classification A, the* disjunctive power *of A is the classification $\vee A$ with:*

- $\mathsf{tok}(\vee A) = \mathsf{tok}(A)$
- $\mathsf{typ}(\vee A) = \mathcal{P}(\mathsf{typ}(A))$
- $a \models_{\vee A} \Phi \iff a \models_A \sigma$ *for some* $\sigma \in \Phi$

η^d_A, the natural embedding $\eta^d_A : A \rightleftarrows \vee A$, is the token identical isomorphism which maps each type to its unit set.

Definition 5.1.31. *Given a classification A, the* conjunctive power *of A is the classification $\wedge A$ with:*

- $\mathsf{tok}(\wedge A) = \mathsf{tok}(A)$
- $\mathsf{typ}(\wedge A) = \mathcal{P}(\mathsf{typ}(A))$
- $a \models_{\wedge A} \Phi \iff a \models_A \sigma$ *for every* $\sigma \in \Phi$

η^c_A, the natural embedding $\eta^c_A : A \rightleftarrows \wedge A$, is the token identical isomorphism which maps each type to its unit set.

Definition 5.1.32. *Given a classification A, the* negation *of A is the classification $\neg A$ with:*

- $\mathsf{tok}(\neg A) = \mathsf{tok}(A)$, $\mathsf{typ}(\neg A) = \mathsf{typ}(A)$
- $a \models_{\neg A} \alpha \iff a \not\models_A \alpha$

Definition 5.1.33. *Let A be a classification.*

- *A* disjunction infomorphism *on A is a token identical infomorphism $d : \vee A \rightleftarrows A$.*
- *A* conjunction infomorphism *on A is a token identical infomorphism $c : \wedge A \rightleftarrows A$.*
- *A* negation infomorphism *on A is a token identical infomorphism $n : \neg A \rightleftarrows A$.*

We write "$\vee\Phi$" for $d^\wedge(\Phi)$, "$\wedge\Phi$" for $c^\wedge(\Phi)$ and "$\neg\alpha$" for $n^\wedge(\alpha)$.
A classification A is called Boolean *if it has a disjunction, a conjunction and a negation.*

Truth classifications are Boolean iff their sentences are built up recursively from a given set of propositional variables using the familiar clauses for the Boolean operators. Lewis classifications are Boolean too, with negation, conjunction and disjunction given by complement, union and intersection respectively.

As the conjunctions and disjunctions we introduced are infinitary, only Kripke classifications based on $\mathcal{L}_\infty$ are Boolean. Weaker classifications may be Boolean iff we restrict ourselves to some subclass of possibilities, e.g. the hereditarily finite sets in the case of classifications based on the finitary language $\mathcal{L}$.

The next theorem illustrates how the introduction of Boolean compounds increases the expressive power of some set of types:

Theorem 5.1.34. *A classification A is Boolean iff for every set X of tokens closed under indistinguishability there is a type $\alpha \in \mathsf{typ}(A)$ such that $X = \mathsf{tok}(\alpha)$.*

Proof [$\Longrightarrow$:] Let A be a Boolean classification and define $\alpha := \bigvee_{a\in X} \bigwedge \mathsf{typ}(a)$. We show that $\alpha \in \mathsf{typ}(A)$ and $X = \mathsf{tok}(\alpha)$.

- $\mathsf{typ}(a) \subset \mathsf{typ}(A)$ $\forall a \in X$, hence $\mathsf{typ}(a) \in \mathsf{typ}(\wedge A)$. So we have, $\forall a \in X$, $c^\wedge(\mathsf{typ}(a)) = \bigwedge \mathsf{typ}(a) \in \mathsf{typ}(A)$. So $\{\bigwedge \mathsf{typ}(a) \mid a \in X\} \subset \mathsf{typ}(A)$, hence $\{\bigwedge \mathsf{typ}(a) \mid a \in X\} \in \vee A$. Thus $d^\wedge(\{\bigwedge \mathsf{typ}(a) \mid a \in X\}) = \bigvee_{a\in X} \bigwedge \mathsf{typ}(a) \in \mathsf{typ}(A)$.
- $X \subset \mathsf{tok}(\alpha)$, because, for any $a \in X$, $a \models_A \bigwedge \mathsf{typ}(a)$, and hence $a \models_A \bigvee_{a\in X} \bigwedge \mathsf{typ}(a)$. For $X \supset \mathsf{tok}(\alpha)$, suppose $b \models_A \alpha$, which means that $\exists a \in X$ such that $b \models_A \bigwedge \mathsf{typ}(a)$. We have to show that $a \sim_A b$. If $\beta \in \mathsf{typ}(a)$, then $b \models_A \beta$ from the definition of $\wedge$. If, conversely, $\beta \in \mathsf{typ}(b)$, then $\neg\beta \notin \mathsf{typ}(b)$ and hence $\neg\beta \notin \mathsf{typ}(a)$ (for we already have shown $\mathsf{typ}(a) \subset \mathsf{typ}(b)$). By the definition of $\neg$ in Boolean classifications, $\beta \in \mathsf{typ}(a)$.

[$\Longleftarrow$:] We have to show that the extensions of Boolean types are closed under indistinguishability:

- $\mathsf{tok}(A) \setminus \mathsf{tok}(\alpha)$ for any $\alpha \in \mathsf{typ}(A)$: If $\mathsf{tok}(\alpha) = \mathsf{tok}(A)$, the claim holds. Let $a \in \mathsf{tok}(A) \setminus \mathsf{tok}(\alpha)$ and $a \sim_A b$. From $a \not\models \alpha$ it follows that $b \not\models \alpha$ and hence $b \notin \mathsf{tok}(\alpha)$.

- $\bigcup_{\gamma\in\Gamma}\mathsf{tok}(\gamma)$ for any $\Gamma\subset\mathsf{typ}(A)$: For any two tokens $a\sim_A b$: $a\in\bigcup_{\gamma\in\Gamma}\mathsf{tok}(\gamma)$ iff $\exists\gamma\in\Gamma : a\models\gamma$ iff $\exists\gamma\in\Gamma : b\models\gamma$ iff $b\in\bigcup\limits_{\gamma\in\Gamma}\mathsf{tok}(\gamma)$.
- $\bigcap\limits_{\gamma\in\Gamma}\mathsf{tok}(\gamma)$ for any $\Gamma\subset\mathsf{typ}(A)$: replace "$\bigcup$" by "$\bigcap$" and "$\exists$" by "$\forall$".

$\square$

For an arbitrary classification, we can build its "Boolification" in an canonical way. To do this, we define $\mathbf{Part}(A)$, the set of partitions of (the types of) a given classification A in the following way:

Definition 5.1.35. *Given a classification A, its partitions are given by:*

$$\mathbf{Part}(A) := \{\langle\Gamma,\Delta\rangle \mid \Gamma\cup\Delta = \mathsf{typ}(A) \;\wedge\; \Gamma\cap\Delta=\emptyset\})$$

For any token $a\in\mathsf{tok}(A)$ we define its state description *as:* $\mathsf{state}(a) := \langle\mathsf{typ}(a),\mathsf{typ}(A)\setminus\mathsf{typ}(a)\rangle$.

Definition 5.1.36. *Given a classification A, the* Boolean closure *of A is the classification* $\mathsf{Boole}(A)$ *with:*

- $\mathsf{tok}(\mathsf{Boole}(A)) = \mathsf{tok}(A)$
- $\mathsf{typ}(\mathsf{Boole}(A)) = \mathcal{P}(\mathbf{Part}(A))$
- $a\models_{\mathsf{Boole}(A)}\alpha \iff \mathsf{state}_A(a)\in\alpha$

η_A, the natural embedding $\eta_A : A \rightleftarrows \mathsf{Boole}(A)$, is the token identical isomorphism which maps each type α to $\{\langle\Sigma,R\rangle \mid \alpha\in\Sigma \;\wedge \langle\Sigma,R\rangle$ is a partition of $\mathsf{typ}(A)\}$.

5.1.7 Modal classifications

A more general way to "modalise" classification than the one we used in (4.5) is given by the following definition:

Definition 5.1.37. *Given a classification A, the* modalisation of A with respect to a transition relation R_a with $a\in\mathcal{A}$ *is the classification $[a]A$ with:*

- $\mathsf{tok}([a]A) = \mathsf{tok}(A)$
- $\mathsf{typ}([a]A) = \mathsf{typ}(A)$
- $b\models_{[a]A}\phi \iff c\models_A\phi$ *for every c such that bR_ac*

Definition 5.1.38. *Let A be a classification. A* $[a]$-infomorphism *is a token identical infomorphism $a : [a]A \rightleftarrows A$. We write "$[a]\phi$" for $a^\wedge(\phi)$. A classification is called* $[a]$-modal *if there is such an infomorphism a with respect to a transition relation a on the tokens. A classification is* modal *if it is $[a]$-modal with respect to all $a \in \mathcal{A}$.*

Kripke classifications are modal, with A_T corresponding to a reflexive, A_{PI} to an transitive, A_I to an Euclidean and A_{TI} to an equivalence transition relation on the set of worlds of the corresponding Kripke models.

5.1.8 State spaces

State spaces are another and important class of separated classifications:

Definition 5.1.39. *A* state space *is a classification S for which each token a is of exactly one type $\mathsf{state}_S(a)$. The state space S is* complete *if state_S is onto. S is* ideal *if $\mathsf{typ}(S) = \mathsf{tok}(S)$ and $\mathsf{state}_S(s) = s$ for every token s.*

A state space is a classification where each token is uniquely described (and the state descriptions $\mathsf{state}(s)$ are not just partitions, but (partitions defined by) single types). State spaces therefore represents a situation where we have complete information about the items of interest. In a complete state space, every type is realised by some token. Our information in this case is not only complete but also non-redundant.

A state space can be straightforwardly embedded in another state space:

Definition 5.1.40. *A* (state space) projection *$f : S_1 \rightrightarrows S_2$ is a covariant pair of functions such that for every token $a \in \mathsf{tok}(S_1)$:*

$$f(\mathsf{state}_{S_1}(a)) = \mathsf{state}_{S_2}(f(a))$$

In analogy to (5.1.5), we define:

Definition 5.1.41. *A state space S_0 is a* subspace *of a state space S_1 if the identity on both types and tokens is a projection $S_0 \rightrightarrows S_1$.*

We can build classifications from state spaces if we take complete theories of particular tokens as our types:

Definition 5.1.42. *Given a state space S the* event classification $\mathsf{Evt}(S)$ *associated with S is defined as follows:*

$$\begin{array}{rcl} \mathsf{tok}(\mathsf{Evt}(S)) & := & \mathsf{tok}(S) \\ \mathsf{typ}(\mathsf{Evt}(S)) & := & \mathcal{P}(\mathsf{typ}(S)) \\ a \models_{\mathsf{Evt}(S)} \alpha & :\Longleftrightarrow & \mathsf{state}_S(a) \in \alpha \end{array}$$

Given a state space projection $f : S_1 \rightrightarrows S_2$ we define a contravariant pair of functions $\mathsf{Evt}(f) : \mathsf{Evt}(S_2) \rightleftarrows \mathsf{Evt}(S_1)$ which is f on tokens of S_1 and maps every type α of $\mathsf{Evt}(S_2)$ to the set $f^{-1}(\alpha)$.

The step from ordinary types to complete theories of tokens increases the expressive power of our type set in the same way it is increased by the introduction of arbitrary Boolean compounds:

Theorem 5.1.43. *For any state space S, the classification $\mathsf{Evt}(S)$ is Boolean.*

Proof Negation, conjunction and disjunction are given by complement, union and intersection on the power set of the original types. □

Given a classification, we can build a corresponding state space by taking arbitrary partitions as types:

Definition 5.1.44. *Given a classification A the* free state space $\mathsf{Ssp}(A)$ *of A is defined as follows:*

$$\begin{array}{rcl} \mathsf{tok}(\mathsf{Ssp}(A)) & := & \mathsf{tok}(A) \\ \mathsf{typ}(\mathsf{Ssp}(A)) & := & \mathbf{Part}(A) \\ \mathsf{state}_{\mathsf{Ssp}(A)}(a) & := & \mathsf{state}_A(a) \end{array}$$

$\mathsf{state}_A(a)$ is the state description $\langle \mathsf{typ}(a), \mathsf{typ}(A) \setminus \mathsf{typ}(a) \rangle$ of a in A (defined in def. 5.1.35). Given a contravariant pair of functions $f : A \rightleftarrows B$ we define a covariant pair of functions $\mathsf{Ssp}(f) : \mathsf{Ssp}(B) \rightrightarrows \mathsf{Ssp}(A)$ which is f on tokens of S_1 and maps every partition $\langle \Gamma, \Delta \rangle$ of $\mathsf{Ssp}(B)$ to the partition $\langle f^{-1}(\Gamma), f^{-1}(\Delta) \rangle$.

Barwise and Seligman (1997: 109–110) have shown that projections on state spaces correspond to infomorphisms on their event classifications and that infomorphisms on classifications correspond to projections on their free state spaces:

Theorem 5.1.45. *Given two state spaces S_1 and S_2, we have:*

$$f : S_1 \rightrightarrows S_2 \textit{ is a projection} \iff \mathsf{Evt}(f) : \mathsf{Evt}(S_2) \rightleftarrows \mathsf{Evt}(S_1) \textit{ is an infomorphism}$$

Theorem 5.1.46. *Given two classifications A and B and a contravariant pair of functions $f : A \rightleftarrows B$ we have:*

$$f : A \rightleftarrows B \textit{ is an infomorphism} \iff \mathsf{Ssp}(f) : \mathsf{Ssp}(B) \rightrightarrows \mathsf{Ssp}(A) \textit{ is a projection}$$

The event classification of the free state space of a classification gives us back the Boolean closure of our original classification:

Theorem 5.1.47. *For any classification A,* $\mathsf{Boole}(A) = \mathsf{Evt}(\mathsf{Ssp}(A))$.

PROOF All three functors are the identity function on tokens. For types, $\mathsf{typ}(\mathsf{Boole}(A))$ is the powerset of all partitions of A. $\mathsf{typ}(\mathsf{Ssp}(A))$ is the set of all partitions of $\mathsf{typ}(A)$ and $\mathsf{typ}(\mathsf{Evt}(\mathsf{Ssp}(A)))$ is the powerset of this set. For the modelling relation, we have the following:

$$a \models_{\mathsf{Boole}(A)} \alpha \iff \mathsf{state}_A(A) \in \alpha \iff \mathsf{state}_{\mathsf{Ssp}(A)}(a) \in \alpha \iff a \models_{\mathsf{Evt}(\mathsf{Ssp}(A))} \alpha$$

□

5.1.9 Frame classifications and ultrafilter state spaces

The idea underlying the modalisation of classifications in (5.1.7) can also be implemented as follows: As the "Boolification" of a classification gives us a Boolean algebra structure on the types, modalisation gives us the structure of a modal algebra. We thus arrive at a generalisation of the Lewis classifications defined in (5.1.3) which can now be seen to correspond to ordinary frames in modal logic.

Whenever we choose some subset of $\mathcal{P}(W)$ as admissible valuations and identify our propositional variables $\mathbb{P}$ with these admissible subsets, we get frame classifications, which correspond to what has been called general frames in modal logic (cf. def. 4.7.12):

Definition 5.1.48. *A* frame classification $A_{\mathbb{P}}$ *of a set A of tokens or possible worlds with respect to a class $\mathbb{P} \subset \mathcal{P}(A)$ of admissible propositions and a set of relations $\{R_i\}_{i \in \mathcal{A}}$ is a* *triple* $A_{\mathbb{P}} = \langle W, \mathbb{P}, \in \rangle$ *such that* $\mathbb{P}$ *is nonempty and closed under the following operations:*

$$\begin{aligned}
X, Y \in \mathbb{P} &\rightarrow X \cup Y \in \mathbb{P} \\
X \in \mathbb{P} &\rightarrow A \setminus X \in \mathbb{P} \\
X \in \mathbb{P} &\rightarrow \{w \in A \mid vR_a w \textit{ for some } v \in X\} \in \mathbb{P} \textit{ for all } a \in \mathcal{A}
\end{aligned}$$

A frame classification thus is a Boolean classification with a modal algebra structure on the types. General frames in modal logic based on some set A of admissible valuations are called *differentiated* if the following holds for all states $s, t \in W$:

$$s = t \iff \forall a \in A (s \in a \iff t \in a)$$

It is clear that this property mirrors the separateness and extensionality of the frame classification corresponding to this frame.

5.2 Distributed systems and information flow

5.2.1 Distributed systems and their limits

We already saw how systems of classifications connected with infomorphisms which have a common codomain, called the "core", can be combined into channels. Arbitrary systems of classifications are called distributed systems:

Definition 5.2.1. *A* distributed system $\mathfrak{A}$ *is an indexed family* $\mathsf{cla}(\mathfrak{A}) = \{A_i\}$ *of classifications together with a set* $\mathsf{inf}(\mathfrak{A})$ *of infomorphisms with domain and codomain in* $\mathsf{cla}(\mathfrak{A})$.

To be able to combine classifications, we first need a categorical product within **Class** which we construct as follows:

Definition 5.2.2. *The* sum $\sum_{i \in \mathcal{I}} A_i$ *of the classifications* $\{A_i\}_{i \in \mathcal{I}}$ *is defined as the Cartesian product of the tokens, the disjoint union of the types and a corresponding modelling relation as follows:*

$$\begin{aligned}
\mathsf{tok}(\sum_{i \in \mathcal{I}} A_i) &:= \prod_{i \in \mathcal{I}} \mathsf{tok}(A_i) \\
\mathsf{typ}(\sum_{i \in \mathcal{I}} A_i) &:= \biguplus_{i \in \mathcal{I}} \mathsf{typ}(A_i) = \bigcup_{i \in \mathcal{I}} \{(i, \alpha_i) \mid \alpha_i \in \mathsf{typ}(A_i)\} \\
\langle a_1, a_2, \ldots \rangle \models_{\sum_i A_i} (i, \alpha) &:\iff a_i \models_{A_i} \alpha
\end{aligned}$$

We adopt this definition because of it gives us the following theorem:

Theorem 5.2.3. *$\sum_{i\in\mathcal{I}} A_i$ is a categorical product of $\{A_i\}_{i\in\mathcal{I}}$, which thus has the universal mapping property of products.*

PROOF We will consider the case of two summands, as the general case is only notationally more complex. So we have to show that, given two infomorphisms $f : A \rightleftarrows C$ and $g : B \rightleftarrows C$, there is a unique infomorphism $f + g : A + B \rightleftarrows C$ such that the following diagram commutes:

$$\begin{tikzcd}
A \arrow[rr, "\sigma_A"'] \arrow[drr, "f"'] & & A+B \arrow[rr, "\sigma_B"'] \arrow[d, "f+g"'] \arrow[dl, phantom, "\circlearrowleft"] \arrow[dr, phantom, "\circlearrowleft"] & & B \\
& & C \arrow[urr, "g"'] & &
\end{tikzcd}$$

$\sigma_A : A \rightleftarrows A + B$ is the canonical embedding which maps each type $\alpha \in \mathsf{typ}(A)$ to $\langle 0, \alpha\rangle$ and each pair in $\mathsf{tok}(A + B)$ to its first component. We have to show that the following contravariant pair of functions $f + g$ is an infomorphism:

- $(f + g)^\wedge(\langle 0, \alpha\rangle) := f^\wedge(\alpha)$ and $(f + g)^\wedge(\langle 1, \beta\rangle) := g^\wedge(\beta)$
- $(f + g)^\vee(c) := \langle f^\vee(c), g^\vee(c)\rangle$

We show that it has the fundamental property of infomorphisms (fund) for the left-hand side of the diagram:

$$\begin{aligned}
(f+g)^\vee(c) \models_{A+B} \langle 0, \sigma\rangle \overset{\text{def. of } (f+g)^\vee}{\iff} \langle f^\vee(c), g^\vee(c)\rangle \models_{A+B} \langle 0, \sigma\rangle \overset{\text{def. of} \models_{A+B}}{\iff} f(c) \models_A \sigma \\
\overset{f \text{ is infomorphism}}{\iff} c \models_C f(\sigma) \overset{\text{def. of } (f+g)^\wedge}{\iff} c \models_C (f+g)(\langle 0, \sigma\rangle)
\end{aligned}$$

□

Sums of classification allow us to define classifications which cover some given distributed system, turning it into a channel:

Definition 5.2.4. *A channel $\mathfrak{C} = \{h_i : A_i \rightleftarrows C\}_{i\in I}$* covers *a distributed system $\mathfrak{A}$ if $\mathsf{cla}(\mathfrak{A}) = \{A_i\}_{i\in I}$ and for each $i, j \in I$ and each infomorphism $f : A_i \rightleftarrows A_j$ in $\mathsf{inf}(\mathcal{A})$ $h_i = h_j \circ f$, that is the following diagram commutes:*

$$\begin{tikzcd}
& C \arrow[d, phantom, "\circlearrowleft"] & \\
A_i \arrow[ur, "h_i"] \arrow[rr, "f"'] & {} & A_j \arrow[ul, "h_j"']
\end{tikzcd}$$

$\mathfrak{C}$ *is a* minimal cover *of* $\mathfrak{A}$ *if it covers* $\mathfrak{A}$ *and for each other channel* $\mathfrak{D}$ *covering* $\mathfrak{A}$ *there is a unique infomorphism from* $\mathfrak{C}$ *to* $\mathfrak{D}$.

A channel covering a distributed system preserves the informational dependencies within it, given by the system's infomorphisms.

Theorem 5.2.5. *Every distributed system has a minimal cover, which is unique up to isomorphism.*

PROOF We will see later that the limit $\mathsf{lim}(\mathfrak{A})$ of a distributed system $\mathfrak{A}$ is a minimal cover. Uniqueness up to isomorphism follows from minimality in the following way: Suppose $\mathfrak{C}$ and $\mathfrak{D}$ are minimal covers of $\mathcal{A}$. Both being minimal, there are unique infomorphisms $r : C \rightleftarrows D$ and $r' : D \rightleftarrows C$. $r' \circ r$, an infomorphism $\mathfrak{C} \rightleftarrows \mathfrak{C}$, is the identity, so $r' = r^{-1}$ and C and D are isomorphic, for there is a one-one correspondence between the types and tokens such that: $r(d) \models_C \gamma \Leftrightarrow d \models_D r(\gamma)$. □

We are now ready to define what we will show to be the minimal cover of a given distributed system:

Definition 5.2.6. *The* limit $\mathsf{lim}(\mathfrak{A})$ *of a distributed system* $\mathfrak{A}$ *with classifications* $\{A_i\}_{i\in I}$ *is the channel* $\{g_i : A_i \rightleftarrows C\}_{i\in I}$ *constructed as follows:*

1. *Define a dual invariant* $J = \langle C, R\rangle$ *on* $A = \sum_{i\in I} A_i$ *as follows.*
   - *The set* C *of tokens consists of those tokens* c *of* A *such that* $f(c_j) = c_i$ *for each infomorphism* $f : A_i \rightleftarrows A_j$ *in* $\mathsf{inf}(\mathfrak{A})$ *(where "*c_i*" denotes the i-th coordinate of* $c \in \mathsf{tok}(A)$*).*
   - *The relation* R *on the types of* A *is defined by:*

     $$\alpha R \beta \quad \Longleftrightarrow \quad \exists f : A_i \rightleftarrows A_j \; \exists \gamma \in \mathsf{typ}(A_i) \; : \; \alpha = \sigma_i(\gamma) \wedge \beta = \sigma_j(f(\gamma))$$

     Let the core of the channel $\mathsf{lim}(\mathfrak{A})$ *be the dual quotient* $C := (\sum_{i\in I} A_i)/J$.
2. *Define infomorphisms* $g_j : A_j \rightleftarrows C$*, for any* $j \in I$*, by:*
   - $g_j(\alpha) = [\sigma_j(\alpha)]_{\equiv_R}$ *for each* $\alpha \in \mathsf{typ}(A_j)$
   - $g_j(c) = \sigma_j(c)$ *for each* $c \in \mathsf{tok}(C)$

Theorem 5.2.7. *Any distributed system* $\mathfrak{A}$ *has its limit* $\mathsf{lim}(\mathfrak{A})$ *as a minimal cover.*

Proof By drawing the commutative diagram, it is easy to see that $\mathsf{lim}(\mathfrak{A})$ is a cover. Let us show that it is minimal. Suppose $\mathfrak{D} = \{h_i : A_i \rightleftarrows D\}$ is another channel covering $\mathfrak{A}$ and that the dual invariant used for the construction of $\mathsf{lim}(\mathfrak{A})$ is $\langle C, R \rangle$. We have to show that there is a unique infomorphism $r : A/J \rightleftarrows D$. To show this, consider the following diagram, where $A = \sum A_i$ and $\tau_J : A \rightleftarrows A/J$ is the canonical embedding into the quotient, inclusion on tokens and mapping each type to its R-equivalence class.

$$\begin{tikzcd}
& D & \\
& A/J \arrow[u, "\exists! r"'] & \\
& A \arrow[u, "\tau_J"] & \\
A_i \arrow[uuur, "h_i"] \arrow[uur, "g_i"] \arrow[ur, "\sigma_i"'] \arrow[rr, "f"'] & & A_j \arrow[uuul, "h_j"'] \arrow[uul, "g_j"'] \arrow[ul, "\sigma_j"]
\end{tikzcd}$$

We first show that $h := \sum h_i : A \rightleftarrows D$ respects the dual invariant J:

- We first have to show that, for all $d \in \mathsf{tok}(D)$, $h(d) \in C$. Because $\{\sigma_i : A_i \rightleftarrows \sum A_i\}_{i \in \mathcal{I}}$ is a cover and thus $f(\sigma_j(\vec{a})) = \sigma_i(\vec{a})\ \forall \vec{a} \in \mathsf{tok}(A)$, we have $f(\sigma_j(h(d))) = \sigma_i(h(d))$. But by the definition of h this means $f(h_j(d)) = h_i(d)$ which is the defining condition for C.
- Second, we show that $\alpha_1 R \alpha_2$ implies $h(\alpha_1) = h(\alpha_2)$. So assume the antecedent. Because f is an infomorphism and we have for the canonical embeddings σ_i of A_i and σ_j of A_j into A: $\alpha_1 = \sigma_i(\alpha_0)$ and $\alpha_2 = \sigma_j(f(\alpha_0))$ for some $\alpha_0 \in \mathsf{typ}(A_i)$. $h(\alpha_1) = h(\sigma_i(\alpha_0)) = h_i(\alpha_0)$ and $h(\alpha_2) = h(\sigma_j(f(\alpha_0))) = h_j(f(\alpha_0))$ by the definition of h. Because $\mathfrak{D}$ is a cover, we have $h_i(\alpha_0) = h_j(f(\alpha_0))$ and thus $h(\alpha_1) = h(\alpha_2)$.

Because h respects the dual invariant J, there is an unique infomorphism $r : A/J \rightleftarrows D$ such that $r \circ \tau_J = h$. $\square$

Having the limit construction at our disposal, we may define quotient channels:

Definition 5.2.8. *Given an invariant $I = \langle \Sigma, R \rangle$ on a classification A, the* quotient channel *of A by I is the limit of the distributed system $\mathfrak{A}$ with* $\mathsf{cla}(\mathfrak{A}) = \{A, A/I, A\}$ *and* $\mathsf{inf}(\mathcal{A}) = \{\tau_I : A/I \rightleftarrows A, \tau_I : A/I \rightleftarrows A\}$.

5.2.2 Information Flow

It is now time to take stock. We have introduced the basic information patterns, namely classifications. Classifications consist of some items being classified, some classificatory scheme and a binary relation telling us what is classified how. This notion of classification is completely general. It is important to bear in mind that the basic elements of the theory are neither models (possible worlds) nor sentences of some given formal language. The application in mind probably was situations and infons, but even this is just one special case among others.

Set theory, then, is treated just as a special kind of classifications (of sets by their elements or of elements by the sets they are elements of) and does not play the rôle of an all-embracing meta-theory.

This general and flexible way of handling the primitives of the theory has several advantages: one of them is that it allows us to avoid a stance towards the identification of indistinguishable tokens, another is that we may handle perspectivism, i.e. the phenomenon that the same situation may be given to different agents (or to the same agent at different times) in different, mutually incompatible ways. To handle such phenomena, we just have to find suitable infomorphisms.

A third advantage is that the information-carrying items are conceived of as types, properties of situations, instead of propositions; the theory is therefore much freer in the compounding of such items. We have seen that classifications may or may not be Boolean; their types may have any internal structure you like.

The most important advantage of the new theory, however, comes with the notion of a channel and with the possibility to cover channels with limits. The theoretical importance of core classifications of channels has been aptly pointed out by van Benthem and Israel:

> "The core of the channel is a classification that in a sense represents the information system, the distributed system, as a whole. The type mappings from the constituent classification are all into the types of the core, and "interpret" the properties of the components into properties of the whole; the token mappings from the core into the constituents represent the connections between the whole and its parts along which properties of the parts can be translated into properties of the whole." (Benthem/Israel: ADD REFERENCE)

Apart from minor quarrels, I fully agree with this statement.

What picture now emerges of information flow? How is the fact that a's being F carries the information that b is G represented by our theory at its present state of development? We saw how the singular states of affairs, a's being F and b's being G can be represented within classificatory systems subject to different constraints. Relative to these constraints, the corresponding propositions, that a is F and that b is G, have greater or lesser inferential power, i.e. they allow for the deduction of more or less consequences. We may, for example, choose to adopt a less expressive set of types, identifying tokens which are indistinguishable relative to this subset of our types, or, conversely, restrict our attention to only those tokens which make for a classificatory difference, identifying types which are coextensional with respect to them.

We have seen how we can impose additional structure both on our set of types, by introducing Boolean operators for example, or by modalising them, and on our set of tokens, e.g. by construing them as non-wellfounded sets, encoding into them a structure which captures the accessibility relations between worlds in ordinary epistemic modal logic.

We have defined infomorphisms, the basic tools to embed classifications into another or, to use a metaphor, to make sense of a classification from the point of view of another classification. Infomorphisms, then, are the channels of the flow of information. Whenever a states of affairs, say a's being F relative to some classification, carries – according to another, possibly different, classification – the information that b is G, the latter classification somehow indirectly classifies also what happens in the former. So an infomorphism may be seen as a partial translation of one classificatory terminology into another, together with the simultaneous restriction on the latter that the translation should be conservative, i.e. not telling us new things framed in the translated terminology about the old items.

Recall that an infomorphism $f : A \rightleftarrows B$ from A to B is defined in (5.1.3) as a contravariant pair of functions $f = (f^\wedge, f^\vee)$ satisfying:

$$(1) \quad f^\vee(b) \models_A \alpha \quad \Longleftrightarrow \quad b \models_B f^\wedge(\alpha) \qquad \text{for all } b \in \mathsf{tok}(B) \text{ and for all } \alpha \in \mathsf{typ}(A)$$

Let us take $f^\vee$ to be $\mathbf{1}_{\mathsf{tok}(B)}$. We then have a straightforward change of terminology, a notational change between two equally expressive representational systems. If, on the other hand, we choose $f^\wedge$ to be $\mathbf{1}_{\mathsf{typ}(A)}$, we have an ontological reinterpretation, some sort of isomorphism between two model structures interpreting our language. In both cases, however, we do not, in an intuitive and loose sense of this term, get anything new.

Things get interesting when we combine both moves, adopting a terminological change necessitating new models to interpret the new language, or an ontological swap where the new models call for a description in a new terminology. We are then not just changing language or ontology: we are doing both at once. (fund), in my view, then states the conditions under which we are justified in believing that we did not change the subject, that we, although we are now talking about different things in a different language, are still dealing with the same problems, questions or issues we had before.

The notion of a conservative extension of some classificatory system alluded to above is a very powerful and useful one. It is well captured by (fund), for the latter simultaneously excludes cases where we adopt too weak a terminology to keep what could be said in the old language (failure of the left-to-right direction) as well as cases where we accept entities which fail to support old properties the translations of which we can ascribe to what they were supposed to replace (failure of the right-to-left direction).

Chapter 6

Representing information by regular theories

6.1 Regular theories and information

6.1.1 Sequents and regular theories

We have seen how certain tokens may carry some information and how the information they carry can be related to other tokens carrying other items of information. We have also seen how different classificatory systems may be combined and the information contained in them fused. There is, however, an important element still missing in the picture.

Information flows, we have said, in virtue of constraints, regularities both within a given classificatory system and between different classifications. In the present state of the theory, these regularities may be taken to be represented *within* a given system (by taking the limit of some distributed system), but we cannot representing them as holding *throughout* a system, as giving its structure – to us and to the epistemic agents in question. We cannot but represent all informational links we wish to investigate as further items in the system, thereby drawing on further informational links to use the links so represented to connect different situations.

In this chapter, we will develop, following again Barwise and Seligman (1997), the materials to draw the crucial difference, noted in sct. 3.3 and discussed in sct. 4.8.3, between situation meaning and situation-type meaning and, more generally, between channels, i.e. informational dependencies between situations, and constraints, i.e. abstract regularities between types of situations that correctly describe some active channel. Jon

Barwise pointed out that neglect of this distinction was one of the crucial shortcomings of his 1983 theory:

> "In retrospect, one of the mistakes in S&A [*Situations and Attitudes*] was our decision to model situations with sets of states of affairs, rather than to take them to be primitive objects [...] This made it very difficult to be sure when we were talking about the world (situations) and when we were talking about information (states of affairs). Keeping clear about this is crucial when you are trying to understand constraints, objects at the level of information, and asking what it is about the world that supports them." (Barwise 1993: 11)

We are thus forced into what might be called a *modelling* perspective, an external viewpoint on the system under study, drawing on external resources and background knowledge for investigating it. It would be attractive, then, to have a way of representing globally the informational links a given classificatory system supports. This is done by regular theories, which we will know investigate.

To get a handle on the structure of the information provided by some tokens, we introduce two-sided Gentzen-style sequents which represent information about some token in our classification:

Definition 6.1.1. *A sequent* $\langle\Gamma, \Delta\rangle$ *of sets of types is* information about a token a *or* holds in a *provided that if* a *is of every type of* Γ*, then* a *is of some type in* Δ*. We write* $\Gamma \vdash_a \Delta$ *if* $\langle\Gamma, \Delta\rangle$ *is information about* a *and say that* a satisfies $\langle\Gamma, \Delta\rangle$*. Similarly, a sequent* $\langle\Gamma, \Delta\rangle$ *is information about a set of tokens* X *iff it is information about every token* $a \in X$*.*

Note that $\Gamma \vdash_a \Delta$ and $a \models \sigma$ provide us (and ultimately the epistemic agents of interest to us) with very different kinds of information. $a \models \sigma$ is a simple fact about some situation a that it indicates σ, relative to some classification. $\Gamma \vdash_a \Delta$, however, is conditional information and says of a that it satisfies some (possibly very local) regularity. $\Gamma \vdash_a \Delta$, though conditional, is not bound to any particular classification: it is classificatory, not assertoric information and expresses a view about a classification (though it may be couched in that classification itself). This classificatory information is what models constraints, as opposed to channels, which are the underlying objective phenomena in virtue of which constraints hold.

We will call any such binary relation (given by a set of sequents), together with some set of types, a theory:

Definition 6.1.2. *A* theory *is a pair* $T = \langle\Sigma, \vdash\rangle$ *of a set of types* Σ *together with a consequence relation* $\vdash$*, i.e. a binary relation between subsets of* Σ*. A* constraint *of* T *is a sequent* $\langle\Gamma, \Delta\rangle \subset \vdash$*, i.e. a pair of subsets of* Σ*, for which* $\Gamma \vdash \Delta$*.*

The binary relations of interest to us are particularly well-behaved:

Definition 6.1.3. *A binary relation* $\vdash$ *is* regular *iff it satisfies the following:*

(**Identity**) $\alpha \vdash \alpha$

(**Weakening**) $\Gamma \vdash \Delta \implies \Gamma, \Gamma' \vdash \Delta, \Delta'$

(**Global Cut**) $\forall \langle \Sigma_0, \Sigma_1 \rangle \in \mathbf{Part}(\Sigma')\ (\Gamma, \Sigma_0 \vdash \Delta, \Sigma_1)$[1] $\implies \Gamma \vdash \Delta$

A theory is regular iff its binary relation is.

We will later see how we can further characterise regular relations and theories.

Classifications, by the information they give us about their tokens, give rise to theories in a natural way:

Definition 6.1.4. *The* theory generated by a classification A *is* $\mathsf{Th}(A) = \langle \mathsf{typ}(A), \vdash_A \rangle$, *whose constraints are the set of sequents which are information about every token of* A*.*

We now have the following theorem:

Theorem 6.1.5. $\mathsf{Th}(A)$ *is regular.*

PROOF Any binary relation which is information about or holds in a set of tokens S is regular. (**Identity**) and (**Weakening**) are obvious. For (**Global Cut**), suppose that $\Gamma, \Sigma_0 \vdash \Delta, \Sigma_1$ for each partition $\langle \Sigma_0, \Sigma_1 \rangle$ of a set Σ' and some $s \in S$ is of all the types in Γ but of no type in Δ. Let $\Sigma_0 := \mathsf{typ}(s) \cup \Sigma'$ and $\Sigma_1 := \Sigma' \setminus \Sigma_0$. But then $\Gamma, \Sigma_0 \nvdash_s \Delta, \Sigma_1$, which implies $\Gamma, \Sigma_0 \nvdash_S \Delta, \Sigma_1$. □

In the case of truth-classifications A_L, $\mathsf{Th}(A_L)$ is just the set of sentences true in all structures. The theory of a Lewis classification, is the set of all possible worlds, i.e. the necessary proposition. This illustrates a special case of the Brentano principle: the richer and more comprehensive our classification, the less uniformities hold across all the classified items.

By (5.1.27), the theory of the Kripke classification K_C is $\mathbf{K}$, $\mathsf{Th}(K_T)$ is $\mathbf{T}$, $\mathsf{Th}(K_{TPI})$ is $\mathbf{S4}$ and $\mathsf{Th}(K_{TI})$ is $\mathbf{S4}$. We might want to choose some handier way to axiomatise these theories, e.g. by Gentzen sequents, but this is not an issue I will pursue here any further.

Consistency for regular relations, regular theories and information contexts (cf. sct. 7.1.1) is defined in the following sense as non-triviality:

[1] "$\mathbf{Part}(\Sigma)$" denotes the set of partitions of Σ. We here extend the notation introduced in def. 5.1.35 in a harmless way.

Definition 6.1.6. *A regular relation is* consistent *iff it is not universal. A theory* $T = \langle \Sigma, \vdash \rangle$ *is consistent if it has a sequent* $\langle \Gamma, \Delta \rangle$ *with* $\Gamma \nvdash \Delta$.

Consistency means that the empty constraint does not hold:

Theorem 6.1.7. *Let* $\vdash$ *be any regular relation:*

$$\vdash \text{ is consistent} \quad \Longleftrightarrow \quad \emptyset \nvdash \emptyset$$

PROOF Suppose $\Gamma \nvdash \Delta$. Then, by (**Weakening**), $\emptyset \nvdash \emptyset$. If $\emptyset \vdash \emptyset$, on the other hand, we get by (**Weakening**) $\Gamma \nvdash \Delta$ for any sequent $\langle \Gamma, \Delta \rangle$. So $\vdash$ is inconsistent. □

The information some sequent carries about the token of a classification is preserved by infomorphisms:

Theorem 6.1.8. *Let* A *and* B *be two classifications,* $f : A \rightleftarrows B$ *an infomorphism,* $\Gamma, \Delta \in \mathsf{typ}(A)$ *and* $b \in \mathsf{tok}(B)$*. Then we have the following:*

(1) $\langle \Gamma, \Delta \rangle$ *holds of* $f(b)$ *in* $A \quad \Longleftrightarrow \quad \langle f(\Gamma), f(\Delta) \rangle$ *holds of* b *in* B

PROOF

[$\Longrightarrow$:] Suppose $\langle \Gamma, \Delta \rangle$ holds of $f(b)$ in A and $b \models_B \beta$ for every $\beta \in f(\Gamma)$. It follows, by (fund), that $f(b) \models_A \alpha$ for every $\alpha \in \Gamma$. But then $f(b) \models_A \alpha'$ for some $\alpha' \in \Delta$. By (fund), this gives us $b \models_B f(\alpha')$ for some $f(\alpha') \in f(\Delta)$ as desired.

[$\Longleftarrow$:] Suppose $\langle f(\Gamma), f(\Delta) \rangle$ holds of b in B and $f(b) \models_A \alpha$ for every $\alpha \in \Gamma$. It follows, by (fund), that $b \models_B f(\alpha)$ for every $f(\alpha) \in f(\Gamma)$. But then $b \models_B f(\alpha')$ for some $f(\alpha') \in f(\Delta)$. By (fund), this gives us $f(b) \models_A \alpha'$ for some $\alpha' \in \Delta$ as desired. □

Theories may be embedded in one another in the following way:

Definition 6.1.9. *A* regular theory interpretation $f : T_1 \to T_2$ *is a function from* $\mathsf{typ}(T_1)$ *to* $\mathsf{typ}(T_2)$ *such that for each* $\Gamma, \Delta \subset \mathsf{typ}(T_1)$*:*

(2) $\Gamma \vdash_{T_1} \Delta \quad \Longrightarrow \quad f(\Gamma) \vdash_{T_2} f(\Delta)$

For an infomorphism $f : A \rightleftarrows B$ *we define an interpretation* $\mathsf{Th}(f) : \mathsf{Th}(A) \to \mathsf{Th}(B)$ *by* $\mathsf{Th}(f)(\alpha) = f^{\wedge}(\alpha)$.
For a state-space projection $f : S_1 \rightrightarrows S_2$ *we define an interpretation* $\mathsf{Th}(f) : \mathsf{Th}(S_2) \to \mathsf{Th}(S_1)$ *by* $\mathsf{Th}(f)(X) = f^{-1}(X)$

That $f(\Gamma) \vdash_{T_2} f(\Delta)$ follows from $\Gamma \vdash_{T_1} \Delta$ means that f preserves the constraints valid in T_1 by translating them into constraints valid in T_2. It thus shows that T_1 is in some sense "weaker" than T_2, or perhaps based on a more expressive set of types. Theory interpretations preserve inconsistency in the following way:

Theorem 6.1.10. *Let T_1 and T_2 be regular theories, $f : \mathsf{typ}(T_1) \to \mathsf{typ}(T_2)$ be a function and $\langle \Gamma, \Delta \rangle$ be any partition of $\mathsf{typ}(T_2)$:*

$$f \text{ is an interpretation} \quad \iff \quad \Gamma \nvdash_{T_2} \Delta \Rightarrow f^{-1}(\Gamma) \nvdash_{T_1} f^{-1}(\Delta)$$

PROOF

[$\Longrightarrow$:] If $f^{-1}(\Gamma) \vdash_{T_1} f^{-1}(\Delta)$, then $ff^{-1}(\Gamma) \vdash_{T_2} ff^{-1}(\Delta)$ because f is an interpretation, which is just $\Gamma \vdash_{T_2} \Delta$.

[$\Longleftarrow$:] If $f(\Gamma') \nvdash_{T_2} f(\Delta')$, then $f^{-1}f(\Gamma') \nvdash_{T_1} f^{-1}f(\Delta')$ which is $\Gamma' \vdash_{T_1} \Delta'$. $\square$

6.1.2 Some theory interpretations

If we take as our example the truth classification A_L embedding Peano Arithmetic into ZFC set theory, the theory interpretation given by the infomorphism translates statements about numbers into statements about, e.g., finite von Neumann ordinals.

For Kripke classifications, the trivial embeddings from (5.1.27) also give us (trivial) theory interpretations of the weaker logics in the stronger ones. For the converse, however, we have to work a little more. There is, however, a particularly interesting example:

Theorem 6.1.11. *Let both A_{TI} and A_{TPI} be Kripke classifications defined with respect to a mono-modal language. Then*

$$f(\alpha) := \Box\Diamond\Box\alpha \quad \forall \alpha \in \mathsf{typ}(A_{TI})$$

defines a theory interpretation.

PROOF We have to verify (2) and, given that $\mathsf{Th}(A_{TI}) = \mathbf{S5}$ and $\mathsf{Th}(A_{TPI}) = \mathbf{S4}$, to show that $\vdash_{\mathbf{S4}} \alpha \iff \vdash_{\mathbf{S4}} \Box\Diamond\Box\alpha$.

[$\Longrightarrow$:] If $\vdash_{\mathbf{S4}} \alpha$, then α is either an axiom or we got it by modus ponens or necessitation.

If α is an **S4** axiom, then it is either a **S4** axiom or $\alpha \equiv \Diamond\beta \to \Box\Diamond\beta$. In the first case, $\vdash_{\mathbf{S4}} \alpha$ and by necessitation $\vdash_{\mathbf{S4}} \Box\alpha$. We derive $\vdash_{\mathbf{S4}} \Box\alpha \to \Box\Diamond\Box\alpha$ as follows:

$\vdash_{\mathbf{S4}} \Box\Diamond\neg\alpha \to \Diamond\neg\alpha$ because it is a **T** axiom. By contraposition and necessitation $\vdash_{\mathbf{S4}} \Box\neg\Diamond\neg\alpha \to \Box\neg\Box\Diamond\neg\alpha$. By contraposition $\vdash_{\mathbf{S4}} \Diamond\Box\Diamond\neg\alpha \to \Diamond\Diamond\neg\alpha$ and by chaining with $\vdash_{\mathbf{S4}} \Diamond\Diamond\neg\alpha \to \Diamond\alpha$, which is the contraposition of an instance of an **S4** axiom, we get $\vdash_{\mathbf{S4}} \Diamond\Box\Diamond\neg\alpha \to \Diamond\neg\alpha$ and by contraposition $\vdash_{\mathbf{S4}} \Box\alpha \to \neg\Diamond\Box\Diamond\neg\alpha$, which is what we wanted. In the second case, where $\alpha \equiv \Diamond\beta \to \Box\Diamond\beta$, we have $\vdash_{\mathbf{S4}} \Box\neg\Box\Box\Diamond\beta \to \neg\Box\Box\Diamond\beta$, which is an instance of the **T** axiom. By contraposition, an instance of the **K** axiom and distribution of $\Diamond$ over disjunction, we get $\vdash_{\mathbf{S4}} \Diamond\Box(\Diamond\beta \to \Box\Diamond\beta)$ and we get the desired conclusion by necessitation.

If α is the result of an application of modus ponens, we apply the induction hypothesis to $\vdash_{\mathbf{S4}} \beta$ and $\vdash_{\mathbf{S4}} \beta \to \alpha$. The result follows if $\Box\Diamond\Box$ distributes over implication: $\vdash_{\mathbf{S4}} \Box\Diamond\Box(\beta \to \alpha) \to (\Box\Diamond\Box\beta \to \Box\Diamond\Box\alpha)$. We start with an instance of the **K** axiom $\vdash_{\mathbf{S4}} \Box(\beta \to \alpha) \to (\Box\alpha \to \Box\beta)$, contrapose the consequent, apply necessitation, distribute $\Box$ and contrapose again to get $\vdash_{\mathbf{S4}} \Box\Box(\beta \to \alpha) \to (\Diamond\Box\beta \to \Diamond\Box\alpha)$. By chaining with the **T** axiom, we weaken the antecedent from $\Box\Box$ to $\Box$, contrapose and apply necessitation again before contraposing back and get $\vdash_{\mathbf{S4}} \Diamond\Box(\beta \to \alpha) \to \Diamond(\Diamond\Box\beta \to \Diamond\Box\alpha)$. The consequent of this is equivalent to $\Box\Diamond\Box\beta \to \Diamond\Diamond\Box\alpha$ and this, by chaining with the **S4** axiom, implies $\Box\Diamond\Box\beta \to \Diamond\Box\alpha$. If we apply now necessitation to the whole, distribute $\Box$ and get $\vdash_{\mathbf{S4}} \Box\Diamond\Box(\beta \to \alpha) \to (\Box\Box\Diamond\Box\beta \to \Box\Diamond\Box\alpha)$, we strengthen the second antecedent by chaining with the **S4** axiom and get the desired result.

If $\alpha \equiv \Box\beta$ and was got by an application of necessitation, the induction hypothesis gives us $\vdash_{\mathbf{S4}} \Box\Diamond\Box\beta$. Because applying contraposition, necessitation, $\Box$ distribution, contraposition, necessitation and $\Box$ distribution again to the **S4** axiom gives us $\vdash_{\mathbf{S4}} \Box\Diamond\Box\beta \to \Box\Diamond\Box\Box\beta$, the result follows by modus ponens.

[$\Longleftarrow$:] Suppose $\vdash_{\mathbf{S4}} \Box\Diamond\Box\alpha$. Then $\vdash_{\mathbf{S4}} \Box\Diamond\Box\alpha$ by the trivial theory interpretation given by (5.1.27). $\vdash_{\mathbf{S4}} \neg\Box\alpha \to \Box\neg\Box\alpha$ because it is an axiom and so $\vdash_{\mathbf{S4}} \Diamond\Box\alpha \to \Box\alpha$ by contraposition. Chaining this with two instances of the **K** axiom, $\vdash_{\mathbf{S4}} \Box\Diamond\Box\alpha \to \Diamond\Box\alpha$ and $\vdash_{\mathbf{S4}} \Box\alpha \to \alpha$, we get $\vdash_{\mathbf{S4}} \Box\Diamond\Box\alpha \to \alpha$ and the result follows by modus ponens. $\Box$

6.1.3 Characterisation of regular theories

To get different characterisations of regular theories, we first have to define when a sequent extends another sequent:

Definition 6.1.12. *A sequent* $\langle\Sigma', \Delta'\rangle$ extends *or* is an extension *of another sequent* $\langle\Sigma, \Delta\rangle$, *denoted by* $\langle\Sigma, \Delta\rangle \leq_{\mathsf{ext}} \langle\Sigma', \Delta'\rangle$, *if* $\Sigma \subset \Sigma'$ *and* $\Delta \subset \Delta'$.

Another characterisation of regular theories is the following (Barwise and Seligman 1997: 120):

Theorem 6.1.13. *A theory $\langle \Sigma, \vdash \rangle$ is regular iff it satisfies the following:*

(**Finite Cut**) $\quad \Gamma, \alpha \vdash \Delta \quad \wedge \quad \Gamma \vdash \Delta, \alpha \quad \Longleftrightarrow \quad \Gamma \vdash \Delta$

(**Partition**) $\quad (\forall \langle \Gamma', \Delta' \rangle \in \mathbf{Part}(\Sigma)\ (\langle \Gamma, \Delta \rangle \leq \langle \Gamma', \Delta' \rangle \to \Gamma' \vdash \Delta')) \quad \Longrightarrow \quad \Gamma \vdash \Delta$

Let us call a partition $\langle \Gamma, \Delta \rangle$ "consistent" if $\Gamma \nvdash \Delta$. Here then is another characterisation of regular theories:

Theorem 6.1.14. *A theory T is regular iff the following condition holds:*

$\Gamma \vdash_T \Delta \quad \Longleftrightarrow \quad$ *there is no T-consistent partition extending* $\langle \Gamma, \Delta \rangle$

Proof [$\Longrightarrow$:] Suppose $\vdash$ is regular. If $\Gamma \vdash \Delta$, then every consistent partition extending $\langle \Gamma, \Delta \rangle$ must be inconsistent by (**Weakening**). If $\Gamma \nvdash \Delta$, then every partition extending $\langle \Gamma, \Delta \rangle$ must be inconsistent by (**Partition**).

[$\Longleftarrow$:] There is no partition extending $\langle \alpha, \alpha \rangle$, so $\alpha \vdash \alpha$ (**Identity**). If there is a consistent partition extending $\langle \Gamma \cup \Gamma', \Delta \cup \Delta' \rangle$, then there is also a consistent partition extending $\langle \Gamma, \Delta \rangle$. So we have (**Weakening**). If there is a consistent partition extending $\langle \Gamma, \Delta \rangle$, this can serve as counterexample to the antecedent of (**Global Cut**). □

This theory gives us an easy and universal way from partitions to theories:

Theorem 6.1.15. *Every set P of partitions of Σ is the set of consistent partitions of a unique regular theory on Σ.*

Proof We define $Th := \langle \Sigma, \vdash \rangle$ by

$\Gamma \vdash \Delta \quad \Longleftrightarrow \quad$ there is no consistent partition extending $\langle \Gamma, \Delta \rangle$

By (6.1.14), Th is regular and unique. □

We also have a notion of compactness for theories:

Definition 6.1.16. *A theory $T = \langle \Sigma, \vdash \rangle$ is* compact *if every constraint $\langle \Gamma, \Delta \rangle$ is an extension of a finite constraint.*

Compactness means that consistency can be characterised by finite sequents alone:

Theorem 6.1.17. *A theory is compact iff, whenever $\langle\Gamma,\Delta\rangle$ is a sequent and every finite sequent $\langle\Gamma_0,\Delta_0\rangle \leq \langle\Gamma,\Delta\rangle$ is consistent, then so is $\langle\Gamma,\Delta\rangle$.*

Proof This is just the contraposition of the definition. $\square$

Truth classifications of first-order languages are compact if their set of tokens is closed under logical equivalence.

For compact theories, (**Global Cut**) can be replaced by (**Finite Cut**):

Theorem 6.1.18. *A compact theory is regular iff it is closed under* (**Partition**), (**Weakening**) *and* (**Finite Cut**).

Proof [$\Longrightarrow$]: Trivial.

[$\Longleftarrow$]: We show (**Partition**), i.e. that every consistent sequent $\langle\Gamma,\Delta\rangle$ can be extended to a consistent partition. Let $(X,\leq_{\mathsf{ext}})$ be the set of all consistent extensions of $\langle\Gamma,\Delta\rangle$. We show that X is closed under suprema of $\leq_{\mathsf{ext}}$ chains. Let $\langle\langle\Gamma_0,\Delta_0\rangle,\langle\Gamma_1,\Delta_1\rangle\ldots\rangle$ be such a chain. Given that it is clearly an extension, we show that $\langle\cup\Gamma_i,\cup\Delta_i\rangle$ is consistent. Suppose the contrary. Given compactness, there are finite unions such that $\bigcup_{i=1}^n\Gamma_i \vdash \bigcup_{i=1}^m\Delta_i$. Suppose that $\bigcup_{i=1}^n\Gamma_i = \Gamma\cup\{a_1,\ldots a_n\}$ and $\bigcup_{i=1}^m\Delta_i = \Delta\cup\{b_1,\ldots b_n\}$. We now iterate the following procedure $n-1$ times:

$$\begin{array}{rl}
\Gamma, a_1,\ldots a_i \vdash \Delta, b_1,\ldots b_i & \\
\Gamma, a_1,\ldots a_i \vdash \Delta, b_1,\ldots b_i, a_i & (\textbf{Weakening}) \\
\Gamma, a_1,\ldots a_{i-1} \vdash \Delta, b_1,\ldots b_i & (\textbf{Finite Cut})
\end{array}$$

and then $m-1$ times the following steps:

$$\begin{array}{rl}
\Gamma \vdash \Delta, b_1,\ldots b_i & \\
\Gamma, b_i \vdash \Delta, b_1,\ldots b_i & (\textbf{Weakening}) \\
\Gamma \vdash \Delta, b_1,\ldots b_{i-1} & (\textbf{Finite Cut})
\end{array}$$

In the end, we arrive at $\Gamma\vdash\Delta$, contrary to our assumption. Because X is closed under suprema of $\leq_{\mathsf{ext}}$ chains, there exists, by Zorn's lemma, a $\leq_{\mathsf{ext}}$ maximal consistent sequent $\langle\Sigma_0,\Sigma_1\rangle$ extending $\langle\Gamma,\Delta\rangle$. We now show that it is a partition. $\Sigma_0\cap\Sigma_1=\emptyset$ follows from (**Identity**) because $\langle\Sigma_0,\Sigma_1\rangle$ is consistent. Suppose there is a $\alpha\in\Sigma : \alpha\notin\Sigma_0\cup\Sigma_1$. Then $\langle\Sigma_0\cup\{\alpha\},\Sigma_1\cup\{\alpha\}\rangle$ is an extension of $\langle\Gamma,\Delta\rangle$ and, by maximality of $\langle\Sigma_0,\Sigma_1\rangle$ has to be a constraint. But then $\Sigma_0 \vdash \Sigma_1$ follows by two applications of (**Finite Cut**). So $\langle \Sigma_0, \Sigma_1 \rangle$ is a consistent partition extending $\langle \Gamma, \Delta \rangle$. □

We can characterise regular theories also by algebraic properties of their consequence relations:

Theorem 6.1.19. *If T is a regular theory, the associated consequence relation $\vdash_T$ is a preorder. Every preorder $\leq$ on a set Σ gives rise to the consequence relation of a regular theory on Σ.*

Proof

[1^{st} claim:] We have to show that $\vdash_T$ is reflexive and transitive. Reflexivity follows from (**Identity**). For transitivity, suppose $\alpha \vdash_T \beta$ and $\beta \vdash_T \gamma$. From the first, we have $\alpha \vdash_T \beta, \gamma$, from the second $\alpha, \beta \vdash_T \gamma$ by (**Weakening**). From these two, $\alpha \vdash_T \gamma$ follows by (**Finite Cut**).

[2^{nd} claim:] Define the consequence relation as follows:

$$\Gamma \vdash \Delta \quad :\Longleftrightarrow \quad \alpha \leq \beta \text{ for an } \alpha \in \Gamma \text{ and } \beta \in \Delta$$

We show that $\Gamma \vdash \Delta$ iff there is no consistent partition extending $\langle \Gamma, \Delta \rangle$. If there is a consistent partition extending $\langle \Gamma, \Delta \rangle$, i.e. $\Sigma_0 \nvdash \Sigma_1$ then for all $\alpha \in \Sigma_0$ and hence for all $\alpha \in \Gamma$, there is no $\beta \in \Sigma_1$ and hence no $\beta \in \Delta$ such that $\alpha \leq \beta$. So $\Gamma \nvdash \Delta$. Suppose now $\Gamma \nvdash \Delta$. The consistent partition is $\langle \Sigma_0, \Sigma_1 \rangle$, where $\Sigma_0 := \{\alpha \in \Sigma \mid \exists \beta \in \Gamma \beta \leq \alpha\}$ and $\Sigma_1 = \Sigma \setminus \Sigma_0$. Because of reflexivity of $\leq$, $\Gamma \subset \Sigma_0$. Because of transitivity, $\Delta \subset \Sigma_1$. □

For extensional classifications, we have a stronger result:

Theorem 6.1.20. *Let A be a classification and $\mathsf{Th}(A)$ its theory:*

$$A \textit{ is extensional} \quad \Longleftrightarrow \quad \vdash_{\mathsf{Th}(A)} \textit{ is a partial order}$$

Let T be a regular theory and $\mathsf{Cla}(T)$ its classification:

$$Cla(T) \textit{ is extensional} \quad \Longleftrightarrow \quad \vdash_T \textit{ is a partial order}$$

Proof

[1^{st} claim, $\Longrightarrow$]: Due to (6.1.5) and (6.1.19), we only have to show antisymmetry. If $\alpha \vdash_{\mathsf{Th}(A)} \beta$ and $\beta \vdash_{\mathsf{Th}(A)} \alpha$, $\mathsf{tok}(\alpha) = \mathsf{tok}(\beta)$, but then $\alpha = \beta$ by extensionality.

[1^{st} claim, $\Longleftarrow$]: From $\mathsf{tok}(\alpha) = \mathsf{tok}(\beta)$ follows $\alpha \vdash_{\mathsf{Th}(A)} \beta$ and $\beta \vdash_{\mathsf{Th}(A)} \alpha$, thus $\alpha = \beta$ by antisymmetry.

[2^{nd} claim, $\Longrightarrow$]: Due to (6.1.19), we only have to show antisymmetry. If $\alpha \vdash_{\mathsf{Th}(\mathsf{Cla}(T))} \beta$ and $\beta \vdash_{\mathsf{Th}(\mathsf{Cla}(T))} \alpha$, $\mathsf{tok}(\alpha) = \mathsf{tok}(\beta)$, but then $\alpha = \beta$ by extensionality.

[2^{nd} claim, $\Longleftarrow$]: From $\mathsf{tok}(\alpha) \subset \mathsf{tok}(\beta)$ follows that if $\alpha \in \Gamma$ for a consistent partition $\langle \Gamma, \Delta \rangle$, then $\beta \in \Gamma$; but this just means $\alpha \vdash_{\mathsf{Cla}(\mathsf{Th}(A))} \beta$. Similarly, we establish $\beta \vdash_{\mathsf{Cla}(\mathsf{Th}(A))} \alpha$, thus $\alpha = \beta$ by antisymmetry. □

In state spaces, where every token a is of exactly one type, we have the following compact characterisation of the constraints of the theory of the corresponding event classification:

Theorem 6.1.21. *Let S be a state space, Ω the set of realised states (i.e. states which are the state of at least one token), and $T = \mathsf{Th}(\mathsf{Evt}(S))$. Then for each sequent $\langle \Gamma, \Delta \rangle$ of T the following holds:*

$$\Gamma \vdash_T \Delta \quad \Longleftrightarrow \quad (\bigcap \Gamma \cap \Omega) \subset \bigcup \Delta$$

Proof [$\Longrightarrow$]: Suppose $\Gamma \vdash_{\mathsf{Th}(\mathsf{Evt}(S))} \Delta$ and $\sigma \in \bigcap \Gamma \cap \Omega$. $\sigma \in \Omega$, so let $\sigma = \mathsf{state}_S(s)$. $\Gamma \vdash_s \Delta\rangle$ and $s \models_{\mathsf{Evt}(S)} \gamma$ for every $\gamma \in \Gamma$, so there is a $\delta \in \Delta$ such that $s \models_{\mathsf{Evt}(S)} \delta$, so $\mathsf{state}_S(s) \in \bigcup \Delta$.

[$\Longleftarrow$]: Suppose $(\bigcap \Gamma \cap \Omega) \subset \bigcup \Delta$, $s \in \mathsf{tok}(S)$ and $s \models_{\mathsf{Evt}(S)} \sigma$ for all $\sigma \in \Gamma$. We have to show that there is a $\delta \in \Delta$ such that $s \models_{\mathsf{Evt}(S)} \delta$. From $s \models_{\mathsf{Evt}(S)} \sigma$ it follows that $\mathsf{state}_S(s) \in \sigma$ (for every $\sigma \in \Gamma$) and thus that $\mathsf{state}_S(s) \in \bigcap \Gamma \cap \Omega$. But then $\mathsf{state}_S(s) \in \bigcup \Delta$, i.e. there is a $\delta \in \Delta$ such that $\mathsf{state}_S(s) \in \delta$. But the latter means that $s \models_{\mathsf{Evt}(S)} \delta$ and so we are done. □

We will later, in def. 7.1.3 of sct. 7.1.1, find further use for a generalised notion of realised states. For our definition of the regular theory of a state space, however, we just leave it out:

Definition 6.1.22. *The* regular theory $\mathsf{Th}(S)$ *of a state space S on arbitrary sets of states of S is given by:*

$$\Gamma \vdash_{\mathsf{Th}(S)} \Delta \quad \Longleftrightarrow \quad \bigcap \Gamma \subset \bigcup \Delta$$

As for classifications, we may also define sums and quotients for theories:

Definition 6.1.23. *The* sum $T + T'$ *of two regular theories is the regular theory which types are the disjoint union of* $\mathsf{typ}(T)$ *and* $\mathsf{typ}(T')$ *and whose consequence relation is given* *by:*

$$\Gamma_1, \Gamma_2 \vdash_{T_1+T_2} \Delta_1, \Delta_2 \quad :\Longleftrightarrow \quad \Gamma_1 \vdash_T \Delta_1 \vee \Gamma_2 \vdash_{T'} \Delta_2$$

We have the following theorem (Barwise and Seligman 1997: 133):

Theorem 6.1.24. *For any classifications A and B:* $\mathsf{Th}(A+B) = \mathsf{Th}(A) + \mathsf{Th}(B)$.

Making use of th. 6.1.15, we define quotients of theories as follows:

Definition 6.1.25. *Let $T = \langle \Sigma, \vdash \rangle$ be a regular theory and R a binary relation on Σ. The* dual quotient $T/R = \langle \Sigma_R, \vdash_{T/R} \rangle$ *of of T is the regular theory the types of which are the R-equivalence classes of Σ and and where $\vdash_{T/R}$ is given by:*

$$\Gamma \vdash_{T/R} \Delta \quad :\Longleftrightarrow \quad \{\alpha \in \Sigma \mid [\alpha]_R \in \Gamma\} \vdash_T \{\beta \in \Sigma \mid [\beta]_R \in \Delta\}$$

This means that the T/R-consistent partitions $\langle \Gamma', \Delta' \rangle$ are exactly those such that $\{\alpha \in \Sigma \mid [\alpha]_R \in \Gamma'\} \vdash_T \{\beta \in \Sigma \mid [\beta]_R \in \Delta'\}$ is T-consistent. Dual quotients may be seen as some kind of "transitive closure" in the following sense:

Theorem 6.1.26. *Let T be a theory with types $\alpha, \beta_1, \beta_2, \gamma$ and $R \subset \mathsf{typ}(T) \times \mathsf{typ}(T)$ a binary relation. Then the following holds:*

(3) $\quad \alpha \vdash_T \beta_1 \ \wedge \ \beta_1 R \beta_2 \ \wedge \ \beta_2 \vdash_T \gamma \quad \Longrightarrow \quad [\alpha]_R \vdash_{T/R} [\gamma]_R$

Proof Suppose $\alpha \vdash_T \beta_1$. We show that $[\alpha]_R \vdash_{T/R} [\beta_1]_R$. Suppose the contrary. Then there is a consistent partition $\langle \Gamma, \Delta \rangle$ with $[\alpha]_R \in \Gamma$ and $[\beta_1]_R \in \Delta$. Hence we have $\{\alpha \in \Sigma \mid [\alpha]_R \in \Gamma\} \nvdash_T \{\beta \in \Sigma \mid [\beta]_R \in \Delta\}$, which contradicts (**Weakening**). We have $[\beta_2]_R \vdash_{T/R} [\gamma]_R$ in an entirely parallel way and conclude to $[\alpha]_R \vdash_{T/R} [\beta_1]_R, [\gamma]_R$ and $[\alpha]_R, [\beta_2]_R \vdash_{T/R} [\gamma]_R$ from which $[\alpha]_R \vdash_{T/R} [\gamma]_R$ follows by (**Finite Cut**) and $[\beta_1]_R = [\beta_2]_R$. □

6.1.4 Moving and comparing regular theories

To see how it is possible to move theories between different classifications and how to compare different theories of the same classification with respect to their expressibility, we will here develop, again following Barwise and Seligman (1997), some further apparatus.

Definition 6.1.27. *Let T_1 and T_2 be two regular theories with the same set of types Σ. We define a* partial order on theories *as follows:*

(4) $\quad T_1 \sqsubseteq T_2 \quad :\Longleftrightarrow \quad \Gamma \vdash_{T_1} \Delta \Rightarrow \Gamma \vdash_{T_2} \Delta$

In this partial order, any two theories have both an upper and a lower bound.

Theorem 6.1.28. *Let T_1 and T_2 be two regular theories with the same set of types Σ. The least upper bound $T_1 \sqcup T_2$ is the theory $\langle \Sigma, \vdash_{T_1 \sqcup T_2} \rangle$ where $\vdash_{T_1 \sqcup T_2}$ is the smallest regular consequence relation that contains both $\vdash_{T_1}$ and $\vdash_{T_2}$. The greatest lower bound $T_1 \sqcap T_2$ is the theory $\langle \Sigma, \vdash_{T_1 \sqcap T_2} \rangle$ with $\Gamma \vdash_{T_1 \sqcap T_2} \Delta :\Longleftrightarrow \Gamma \vdash_{T_1} \Delta \wedge \Gamma \vdash_{T_2} \Delta$.*

PROOF

[1^{st} claim:] Obvious.

[2^{nd} claim:] We have to show that $\vdash_{T_1 \sqcap T_2}$ defines the largest regular theory contained in both T_1 and T_2. Any theory contained in both T_1 and T_2 must validate the left-to-right direction. By the right-to-left direction, $T_1 \sqcap T_2$ is minimal. It is clearly regular if T_1 and T_2 are. □

The set of theories on Σ is thus a lattice. By a generalisation of the proof above it can be shown that this lattice is complete. It also has both a top and a bottom element:

Theorem 6.1.29. *The smallest regular theory on Σ is $\langle \Sigma, \vdash_\perp \rangle$ with $\Gamma \vdash_\perp \Delta :\Longleftrightarrow \Gamma \cap \Delta \neq \emptyset$ for all sequents $\langle \Gamma, \Delta \rangle$. The largest regular theory on Σ is $\langle \Sigma, \vdash_\top \rangle$, where $\vdash_\top$ is the universal relation on all sequents of Σ.*

PROOF

[1^{st} claim:] We have to show that $\vdash_\perp$ defines the smallest regular theory on Σ. It is regular because of th. 6.1.14 and there is no consistent partition extending $\langle \Sigma, \Delta \rangle$ if $\Gamma \cap \Delta \neq \emptyset$. By (**Identity**), it is clearly the smallest regular theory.

[2^{nd} claim:] Obvious. □

What now is the interpretation of $\sqsubseteq$? It is a measure of expressibility and this can be seen by the following:

Theorem 6.1.30. *Let T_1 and T_2 be two regular theories with the same set of types Σ. Then we have:*

$T_1 \sqsubseteq T_2 \quad \Longleftrightarrow \quad$ *The inclusion map $T_1 \subset T_2$ is a theory interpretation*

PROOF We have to check (2), i.e. to validate $\Gamma \vdash_{T_1} \Delta \Rightarrow \Gamma \vdash_{T_2} \Delta$. But this is just (4). $\square$

We noticed in sct. 6.1.1 that theory interpretations (def. 6.1.9) may be interpreted as translations of the constraints of some theory into the set of types of another theory. If there is no translation needed, the first theory is straightforwardly embedded into the second. This is what is expressed by $T_1 \sqsubseteq T_2$.

Having now a measure of expressibility $\sqsubseteq$ of theories at hand, we may study different ways of moving theories around.

Definition 6.1.31. *The* inverse image $f^{-1}(T')$ *of a regular theory* $T' = \langle \Sigma', \vdash_{T'} \rangle$ *under* $f : \Sigma \to \Sigma'$ *is the theory with types* Σ *and the consequence relation given by:*

(5) $\quad \Gamma \vdash_{f^{-1}(T')} \Delta \quad :\Longleftrightarrow \quad f(\Gamma) \vdash_{T'} f(\Delta)$

This definition corresponds to the rule f-Elim we will discuss in detail below (cf. 212, sct. 7.1.2). We adopt it for the following reason:

Theorem 6.1.32. *The inverse image of a regular theory under* $f : \Sigma \to \Sigma'$ *is the largest regular theory on* Σ *such that* f *is an interpretation.*

PROOF We first have to show that $f^{-1}(T')$ is regular. Obviously, it satisfies (**Weakening**) for we have $f(\Gamma) \subset f(\Gamma \cup \Gamma')$ for any $\Gamma, \Gamma' \subset \mathsf{typ}(T)$ if f is a theory-interpretation. So we show (**Partition**). Suppose $\Gamma, \Delta \subset \mathsf{typ}(T)$ such that $\Gamma \nvdash_{f^{-1}(T)} \Delta$. So, we have $f(\Gamma) \nvdash_{T'} f(\Delta)$ by the definition of $\vdash_{f^{-1}(T')}$. By (**Partition**) in T', there is a consistent partition $\langle \Gamma', \Delta' \rangle$ of $\mathsf{typ}(T')$ extending $\langle f(\Gamma), f(\Delta) \rangle$. $\langle f^{-1}(\Gamma'), f^{-1}(\Delta') \rangle$ is a partition because f is a function. It extends $\langle \Gamma, \Delta \rangle$ and is consistent because f is an interpretation.

Suppose S is another regular theory on Σ that makes f an interpretation. We have to show that $\vdash_S \subset \vdash_{f^{-1}(T)}$. This is clear because $\vdash_S$ has to validate the left-to-right direction of (5) and possibly includes more constraints not forced by the right hand-side of (5). $\square$

The definition of images of regular theories is complicated by the fact that interpretations do not have to surjective. This is why the definition $\Gamma' \vdash \Delta' :\Longleftrightarrow f^{-1}(\Gamma) \vdash_T f^{-1}(\Delta)$ does not in general give us a regular consequence relation $\vdash$. The reason is that, for an interpretation $f : T \to T'$, it only follows from $\langle \Gamma, \Delta \rangle$ being a partition of Σ that $\langle f(\Gamma), f(\Delta) \rangle$ is a partition of Σ' if f is surjective (because f has to be surjective for $A \subset f(f^{-1}(A))$ to hold). If f is not surjective and $\alpha \in \Sigma' \setminus f(\Sigma)$, $f^{-1}(\alpha) \vdash_T f^{-1}(\alpha)$ becomes $\emptyset \vdash_T \emptyset$, which makes $\vdash_T$ inconsistent by th. 6.1.7. We will later, in sct. 7.1.2, see why this brings with it a certain disanalogy between def. 6.1.31 and the following definition:

Definition 6.1.33. *The* image $f(T)$ *of a regular theory* $T = \langle \Sigma, \vdash_T \rangle$ *under* $f : \Sigma \to \Sigma'$ *is the theory with types* Σ' *and the consequence relation given by its consistent partitions:*

$$(6) \quad \forall \langle \Gamma, \Delta \rangle \in \mathbf{Part}(\Sigma') : \quad \Gamma \nvdash_{f(T)} \Delta \iff f^{-1}(\Gamma) \nvdash_T f^{-1}(\Delta)$$

This definition corresponds to the rule f-Intro we will discuss in detail below (cf. 213, sct. 7.1.2). Though (6.1.33) is not equivalent with the $\vdash$ version, we have the following dual result:

Theorem 6.1.34. *The image of a regular theory* T *under* $f : \mathsf{typ}(T) \to \Sigma'$ *is the smallest regular theory on* Σ' *such that* f *is an interpretation.*

PROOF $f(T)$ is regular because it is defined in terms of consistent partitions (cf. th. 6.1.15).

Suppose S is another regular theory on Σ' that makes f an interpretation. We have to show that $\vdash_{f(T)} \sqsubset\!\!\!\vdash_S$. Suppose $\Gamma', \Delta', \subset \Sigma'$ and $\Gamma' \vdash_{f(T)} \Delta'$. Let $\langle \Gamma'', \Delta'' \rangle$ be a partition of Σ' extending $\langle \Gamma', \Delta' \rangle$. By (**Weakening**), we have $\Gamma'' \vdash_{f(T)} \Delta''$. By the definition of $\vdash_{f(T)}$, this gives us $f^{-1}(\Gamma'') \vdash_T f^{-1}(\Delta'')$. Because f is an interpretation, we get $f(f^{-1}(\Gamma'')) \vdash_{f(S)} f(f^{-1}(\Delta''))$. Because of $f(f^{-1}(\Gamma'')) \subset \Gamma''$ and $f(f^{-1}(\Delta'')) \subset \Delta''$, this gives us $\Gamma'' \vdash_{f(S)} \Delta''$ by (**Weakening**). Because $\langle \Gamma'', \Delta'' \rangle$ was arbitrary, we conclude $\Gamma' \vdash_S \Delta'$ by (**Partition**). □

6.1.5 From theories to classifications and back

For each classification A, we have the theory generated by it, $\mathsf{Th}(A)$. Conversely, given a theory T, we may define the corresponding classification $\mathsf{Cla}(T)$ as follows:

Definition 6.1.35. *Given a regular theory* T*, the* classification generated by T *is the classification* $\mathsf{Cla}(T)$ *with:*

$$\begin{array}{rcl} \mathsf{tok}(\mathsf{Cla}(T)) & := & \{\langle \Gamma, \Delta \rangle \mid \langle \Gamma, \Delta \rangle \in \mathbf{Part}(T) \wedge \Gamma \nvdash_T \Delta\} \\ \mathsf{typ}(\mathsf{Cla}(T)) & := & \mathsf{typ}(T) \\ \langle \Gamma, \Delta \rangle \models_{\mathsf{Cla}(T)} \alpha & :\iff & \alpha \in \Gamma \end{array}$$

Given an interpretation $f : T \to T'$*, we define* $\mathsf{Cla}(f) : \mathsf{Cla}(T) \rightleftarrows \mathsf{Cla}(T')$ *by:*

$$\begin{array}{rcl} \mathsf{Cla}(f)^\wedge(\alpha) & := & f(\alpha) \text{ for } \alpha \in \mathsf{typ}(T) \\ \mathsf{Cla}(f)^\vee(\langle \Gamma, \Delta \rangle) & := & \langle f^{-1}(\Gamma), f^{-1}(\Delta) \rangle \text{ for any token } \langle \Gamma, \Delta \rangle \in \mathsf{tok}(\mathsf{Cla}(T')) \end{array}$$

The tokens of $\mathsf{Cla}(T)$ are thus just the consistent partitions of $\mathsf{typ}(T)$. This definition is justified by the following result:

Theorem 6.1.36. $\mathsf{Cla}(f)$ *as defined above is well-defined and an infomorphism.*

Proof We have to show that $\mathsf{Cla}(f) : \mathsf{Cla}(T) \rightleftarrows \mathsf{Cla}(T')$. On tokens, $\langle f^{-1}(\Gamma), f^{-1}(\Delta)\rangle$ is a consistent partition of $\mathsf{typ}(T_1)$ if $\langle(\Gamma), (\Delta)\rangle$ is a consistent partition of $\mathsf{typ}(T_2)$, because f is a theory interpretation and (6.1.10). So we have to check (fund):

$$\mathsf{Cla}(f)^{\vee}(\langle\Gamma,\Delta\rangle) \models_{\mathsf{Cla}(T)} \alpha \overset{\text{def. }\mathsf{Cla}(f)^{\vee}}{\iff} \langle f^{-1}(\Gamma), f^{-1}(\Delta)\rangle \models_{\mathsf{Cla}(T)} \alpha \overset{\text{def. }\models_{\mathsf{Cla}(T)}}{\iff} \alpha \in f^{-1}(\Gamma)$$
$$\overset{ff^{-1}(\Gamma)=\Gamma}{\iff} f(\alpha) \in \Gamma \overset{\text{def. }\models_{\mathsf{Cla}(T')}}{\iff} \langle\Gamma,\Delta\rangle \models_{\mathsf{Cla}(T')} f(\alpha) \overset{\text{def. }\mathsf{Cla}(f)^{\wedge}}{\iff} \langle\Gamma,\Delta\rangle \models_{\mathsf{Cla}(T')} \mathsf{Cla}(f)^{\wedge}(\alpha)$$

□

The following crucial representation theorem now assures us that we can do the same things with theories than with classifications: the information "contained" in a classificatory scheme may be extracted and captured abstractly as some sort of logic, while every such logic can be embedded into a classificatory system.

Theorem 6.1.37. Representation theorem: *For any regular theory T and any interpretation f:*

$$T = \mathsf{Th}(\mathsf{Cla}(T)) \qquad f = \mathsf{Th}(\mathsf{Cla}(f))$$

Proof

[1^{st} claim:] Clearly $\mathsf{typ}(T) = \mathsf{typ}(\mathsf{Th}(\mathsf{Cla}(T)))$. We have to show that for any partition $\langle\Gamma,\Delta\rangle$:

$$\Gamma \vdash_T \Delta \iff \Gamma \vdash_{\mathsf{Th}(\mathsf{Cla}(T))} \Delta$$

For the direction from left to right, take any token $\langle\Gamma',\Delta'\rangle$ of $\mathsf{Cla}(T)$, i.e. a partition of $\mathsf{typ}(T)$ such that $\Gamma' \nvdash_T \Delta'$ and show that it satisfies $\langle\Gamma,\Delta\rangle$. Suppose it does not. Then $\langle\Gamma',\Delta'\rangle$ is of all the types in Γ, but of none of the types in Δ. This means $\Gamma \subset \Gamma'$ but $\Delta \cap \Gamma' = \emptyset$. Because $\langle\Gamma',\Delta'\rangle$ is a partition, the latter implies $\Delta \subset \Delta'$ and thus $\Gamma' \vdash_T \Delta'$ follows with (**Weakening**). We have a contradiction, so $\langle\Gamma',\Delta'\rangle$ satisfies $\langle\Gamma,\Delta\rangle$.

For the direction from right to left, suppose $\Gamma \nvdash_T \Delta$. By (**Partition**), there is a consistent partition $\langle\Gamma',\Delta'\rangle$ of $\mathsf{typ}(T)$ extending $\langle\Gamma,\Delta\rangle$. But then $\Gamma \nvdash_{\langle\Gamma',\Delta'\rangle} \Delta$.

[2^{nd} claim:] This is unproblematic because theory interpretations operate only on types: $f = \mathsf{Cla}(f)^\wedge = \mathsf{Th}(\mathsf{Cla}(f))$. $\square$

It follows from (6.1.37) that every theory is generated by some classification.

Theorem 6.1.38. Abstract Completeness Theorem: *Every regular theory is* $\mathsf{Th}(A)$ *for some classification* A*. Every interpretation is* $\mathsf{Th}(f)$ *for some infomorphism* f*.*

Proof Immediate from (6.1.37), for $\mathsf{Cla}(T)$ is a classification for a regular theory T and $\mathsf{Cla}(f)$ is an infomorphism for a theory interpretation f. $\square$

Theorem 6.1.39. *For any classification* A*,* $\mathsf{Sep}(A)$ *is isomorphic to* $\mathsf{Cla}(\mathsf{Th}(A))$*.*

Proof Define $f : \mathsf{Cla}(\mathsf{Th}(A)) \rightleftarrows \mathsf{Sep}(A)$ as $f^\wedge(\alpha) = \alpha$ on types and $f^\vee([a]_\sim) = \mathsf{state}(a)$. We have to show that f is an infomorphism and that it is bijective. Let us verify (fund):

$$f^\vee([a]_\sim) \models_{\mathsf{Cla}(\mathsf{Th}(A))} \alpha \overset{\text{def. of } f^\vee}{\iff} \langle \mathsf{typ}(a), \mathsf{typ}(A) \setminus \mathsf{typ}(a) \rangle \models_{\mathsf{Cla}(\mathsf{Th}(A))} \alpha \overset{\text{def. of} \models_{\mathsf{Cla}(\mathsf{Th}(A))}}{\iff}$$
$$\alpha \in \mathsf{typ}(a) \overset{\text{def. of } \mathsf{typ}(a)}{\iff} a \models_A \alpha \overset{\text{def. of} \models_{\mathsf{Sep}(A)}}{\iff} [a]_\sim \models_{\mathsf{Sep}(A)} \alpha \overset{\text{def. of } f^\wedge}{\iff} [a]_\sim \models_{\mathsf{Sep}(A)} f^\wedge(\alpha)$$

f is bijective because every indistinguishability class of tokens corresponds to exactly one state description and every state description describes the state of at least one token. $\square$

Theorem 6.1.40. *A classification* A *is isomorphic to* $\mathsf{Cla}(T)$ *for some regular theory* T *iff* A *is separated.*

Proof

[$\Longrightarrow$:] If A is isomorphic to the classification of some regular theory T then, by (6.1.37), it is isomorphic to $\mathsf{Cla}(\mathsf{Th}(\mathsf{Cla}(T)))$. But then, by (6.1.39), it is isomorphic to $\mathsf{Sep}(\mathsf{Cla}(T))$. Thus there is an bijective infomorphism $f : \mathsf{Sep}(\mathsf{Cla}(T)) \rightleftarrows A$ and A is separated by (5.1.11).

[$\Longleftarrow$:] Whenever A is separated, the identity infomorphism respects the invariant $\langle \mathsf{typ}(A), \sim_A \rangle$. By (5.1.13) there is an unique infomorphism g such that $\tau_\sim \circ g = \mathbf{1}_A$. So $g = \tau_I^{-1}$ and A is isomorphic to $\mathsf{Sep}(A)$ and thus, by (6.1.39), isomorphic to $\mathsf{Cla}(\mathsf{Th}(A))$. But $\mathsf{Th}(A)$ is a regular theory. $\square$

6.2 Information Flow

6.2.1 Information in and about a system

As indicated in sct. 6.1.1, theories give us a way to model informational dependencies in our system which is rather different from the consideration of channels. Channels are informational dependencies considered as entities, "tokens", in their own right. Theories, however, are relations on types, not bound to one particular classification and conditional. We saw on p. 142 that there is a crucial difference between information channels, which are informational connections between concrete situations to which we have independent epistemic access (by their being reified as tokens of a further limit classification) and informational connections which are given to us *in abstracto*, as regularities which hold of a given system. The latter are regular theories.

The distinction between channels and theories explains how information is both in the world and about the world. Information is *in* the world in so far as its bearers are concrete entities, which are not informational *per se* but depend for their informativity on some coding system and on informational regularities connecting them with other concrete tokens (we discussed this issue of the 'relativity' of information in sct. 1.4.7). Information, crucially, is not only in but also *about* the world: a given bearer of information is able to tell us something about something else than itself – this aboutness condition is what makes information bearers signs (cf. sct. 1.4.2). Regular theories are our means to capture this. They may themselves be considered as classifications and as every classification has a theory, every theory is *about* some classification (th. 6.1.39), namely the classification which it is itself (th. 6.1.37).

Regular theories, then, are not just things outside the classificatory framework and not just things used to describe it. They are also, as we will see in sct. 7.1 the means by which we reason *within a classification*: they allow us to deduce some information implicit in what we have. In Dretske's terminology, they give us access to *nested* information. We will in sct. 7.1 see how we can take advantage of this fact without committing us to omniscience. We will thus be able to draw the distinction we saw to be crucial in sct. 2.3.

Regular theories allow us to bridge the gulf between classifications: they enable reasoning at a distance and thus do justice to a crucial property of information:

> "Information typically involves a fact indicating something about the way things are elsewhere and elsewhen, and this is what makes information useful and interesting. [...] The informational content of a fact can concern remote things and situations." (Israel and Perry 1990: 4)

They account, in Lemon's words, for *what* information flows through a system, while channels explain *why* it flows (Lemon 1998: 399).

6.2.2 Reasoning at a distance

What picture now emerges of information flow? We saw, in sct. 6.2.1, that the distinction between channels and constraints allows us to distinguish between the informational regularities and the underlying connections in and between a distributed system. Consider the following channel $\mathfrak{C}$:

$$\begin{tikzcd}
 & C & \\
P \arrow[ur, "p"] & & D \arrow[ul, "d"']
\end{tikzcd}$$

Suppose an epistemic agent has classification P (for "*p*roximal") and reasons, using the means made available to him by P, about another classification D (for "*d*istal"). Here, then, is Barwise's and Seligman's "initial proposal" for information flow:

> "Suppose that the token a is of type α. Then a's being of type α *carries the information that b is of type β*, relative to the channel $\mathfrak{C}$, if a and b are connected in $\mathfrak{C}$ and if the translation α' of α entails the translation β' of β in the theory $\mathsf{Th}(C)$, where C is the core of $\mathfrak{C}$." (Barwise and Seligman 1997: 35)

So the picture is the following:

$$\begin{tikzcd}
 & p^\wedge(\alpha) \vdash_{\mathsf{Th}(C)} d^\wedge(\beta) & \\
a \models_P \alpha \arrow[ur, "p"] & & b \models_D \beta \arrow[ul, "d"']
\end{tikzcd}$$

One of the virtues of this proposal, as Barwise and Seligman (1997: 36) note, is that – due to th. 5.2.7 – it preserves the Xerox principle (2.2.1), i.e. makes information flow transitive because information flow relative to $\mathfrak{C}_1$ and information flow relative to $\mathfrak{C}_2$ automatically gives rise to information flow relative to the limit of these two channels.

This, however, brings a problem to the fore: the fact in virtue of which information flows is a global fact about the core C of the channel. The connecting fact, i.e. the token $c \in \mathsf{tok}(C)$ such that $p^\vee(c) = a$ and $d^\vee(c) = b$ plays no rôle at all, indeed the agent reasoning in P does not have to have epistemic access to it. If he has, however, he is able to make use of the information available to him in the following way:

(7)	$a \models_P \alpha$	premiss
(8)	$\exists c \in \mathsf{tok}(C) \, : \, p^\vee(c) = a \wedge d^\vee(c) = b$	a and b are connected
(9)	$c \models_C p^\wedge(\alpha)$	by (fund) from (7) and (8)
(10)	$p^\wedge(\alpha) \vdash_{\mathsf{Th}(C)} d^\wedge(\beta)$	constraint on the core
(11)	$c \models_C d^\wedge(\beta)$	c satisfies the constraint
(12)	$b \models_B \beta$	by (fund) from (11) and (8)

In (9), the agent needs some acquaintance with the connecting fact. If c is some token of the cover of a distributed system, this requirement does not have to be satisfied: the core may just be too remote for the agent to have an inkling of it. (Barwise and Seligman 1997: 36–37) point out two related shortcomings of their "initial proposal". The first, which they label "aesthetic", is that the constraints in questions are not constraints on the classifications P and D but on the core classification C. We will in sct. 7.1.2 see how the notion of an information context, linking classfications with regular theories, allows us to move the regular theories themselves. This will also help us to get a more realistic picture of the reasoning an agent can base on information available to him and thus alleviate our first worry. The second of Barwise's and Seligman's worry and the one they consider more important is that their initial proposal "gives a God's eye analysis of information flow" (1997: 37), for it presupposes complete information about the core of the channel – at least as far as the "exploitable" regularities are concerned. This will also be an issue of concern to us in sct. 7.1.2.

As noted above, regular theories give us the resources to model the crucial distinction between constraints and channels. We are now arrived at a stage of the theory represented by Barwise's 1993 article "Constraints, channels, and the flow of information". This theory is characterised by the following five principles (Barwise 1993: 6–8), where "$s_1 : \phi \Longrightarrow s_2 : \psi$" abbreviates "$s_1 : \phi$ carries the information that $s_2 : \psi$":

6.2.1 (Xerox Principle). *If $s_1 : \phi \Longrightarrow s_2 : \psi$ and $s_2 : \psi \Longrightarrow s_3 : \theta$ then $s_1 : \phi \Longrightarrow s_3 : \theta$.*

6.2.2 (Logic as Information Flow). *If ϕ entails ψ, then $s : \phi \Longrightarrow s : \psi$.*

6.2.3 (Addition of Information). *If* $s_1 : \phi \Longrightarrow s_2 : \psi$ *and* $s_1 : \phi' \Longrightarrow s_2 : \psi'$ *then* $s : (\phi \wedge \phi' \Longrightarrow s_2 : (\psi \wedge \psi')$.

6.2.4 (Exhaustive Cases). *If* $s_1 : \phi \Longrightarrow s_2 : (\psi \vee \psi')$, $s_2 : \psi \Longrightarrow s_3 : \theta$ *and* $s_2 : \phi' \Longrightarrow s_3 : \theta$ *then* $s_1 : \phi \Longrightarrow s_3 : \theta$.

6.2.5 (Contraposition). *If* $s_1 : \phi$ *would carry the information that* $s_2 : \psi$, *then* $s_2 : \neg\psi$ *would carry the information that* $s_1 : \neg\phi$.

(6.2.1) corresponds to Dretske's Xerox principle (2.2.1) discussed on p. 36. (6.2.2) (6.2.3) (6.2.4)

(6.2.5) stands out by being formulated subjunctively.[2]

[2]In the motivating example, however, Barwise formulates the consequent indicatively: "If the thermometer's mercury were 8 cm high, that would carry the information that the temperature were 40 F. Thus, the temperature's *not* being 40 F carries the information that the thermometer's mercury is *not* 8 cm high." (Barwise 1993: 8)

Chapter 7

Representing epistemic states by information contexts

7.1 Information contexts and epistemic states

7.1.1 Information contexts

We have seen how to use sequents or theories to represent the information available to an agent at a given time. While the information is defined with respect to some given set of tokens, these do not have to be the tokens the agent takes the information to be about. To keep track of these, we introduce the notion of *normal situations* and the associated notion of an information context, which in *Information Flow* were called "local logics":

Definition 7.1.1. *An* information context $\mathcal{C} = \langle \mathsf{cla}(\mathcal{C}), \vdash_{\mathcal{C}}, N_{\mathcal{C}} \rangle$ *consists of a classification* $\mathsf{cla}(\mathcal{C}) = \langle S, \Sigma, \models \rangle$ *of situations* S *and situation types* Σ*, a regular theory* $\mathsf{th}(\mathcal{C})$ *defined by a binary relation* $\vdash_{\mathcal{C}}$ *relating sets of situation types (which represents the information available in the context), and a set* $N_{\mathcal{C}}$ *of normal situations such that every sequent in* $\vdash_{\mathcal{C}}$ *is information about this set, i.e. such that* $\vdash_{\mathcal{C}} \subset \vdash_N$*: all normal situations satisfy all the constraints in* $\mathsf{th}(\mathcal{C})$.

The picture, then, is this: An information context represents the information available to an agent at a given time – by representing his picture of the world (the classification), the regularities he knows of (the regular theory) and the things he takes the information available to him to be about (the normal situations). It is not true, in general, that the agent has information about all the situations he recognises (so not all situations are normal): this happens when he recognises non-universal constraints as active. Some of them are excluded by information he has, i.e. are not considered relevant alternatives by him and hence are non-normal situations. Increase in information narrows down the class of normal situations. We do not assume, however, that every situation which satisfies all the constraints is normal: it may either satisfy them "by accident", as it were, or not be *taken to* satisfy them. So it will not be among the situations the available information is taken to be about and hence not count as normal.

Classifications, state spaces and regular theories give rise to information contexts:

Definitions 7.1.2. *The information context* $\mathsf{IC}(A)$ *generated by a classification* A *has classification* A*, regular theory* $\mathsf{Th}(A)$ *and all its tokens are normal. An information context is* natural *if it is generated by a classification. The information context* $\mathsf{IC}(S)$ *generated by a state space* S *has classification* $\mathsf{Evt}(S)$*, regular theory* $\mathsf{Th}(S)$ *and all its tokens are normal. The information context* $\mathsf{IC}(T)$ *generated by a regular theory* T *has classification* $\mathsf{Cla}(T)$*, regular theory* T *and all its tokens are normal. An information context is* formal *if it is generated by a regular theory.*

The dissociation of the available information from the situations the information is taken to be about allows us to speak of non-realised states, i.e. items of information which are not about any of the situations under consideration, and possible states, i.e. items of information which, for all we know in the given context, might be realized:

Definition 7.1.3. *Let* $\mathcal{C} = \langle \mathsf{cla}(\mathcal{C}), \vdash_{\mathcal{C}}, N_{\mathcal{C}} \rangle$ *be an information context with classification* $\mathsf{cla}(\mathcal{C}) = \langle S, \Sigma, \models \rangle$.

A state ω *of* $\mathcal{C}$ *is a partition* $\omega = \langle \Gamma, \Delta \rangle$ *of* Σ.
The state $\mathsf{state}(s)$ of situation $s \in S$ *is* $\langle \Gamma_s, \Delta_s \rangle := \langle \mathsf{typ}(s), \Sigma \setminus \mathsf{typ}(s) \rangle$.
A state ω *is* realized *by the situation* s *if* $\mathsf{state}(s) = \omega$.
A state $\omega = \langle \Gamma, \Delta \rangle$ *is* possible *iff* $\Gamma \nvdash_{\mathcal{C}} \Delta$.
A situation s *is* possible *iff its state* $\langle \Gamma_s, \Delta_s \rangle$ *is possible.*

As we did for the corresponding state-space notion on p. 189, we denote by "$\Omega_{\mathcal{C}}$" the set of all states of an information context $\mathcal{C}$. We introduced state descriptions of tokens and partitions in general in def. 5.1.35. Here they are used to model something very like valuations: a partition of Σ may be taken to correspond to truth assignments, i.e. functions mapping every primitive proposition to a truth-value (or a set of worlds) and then extended over Boolean compounds in the obvious way.

(7.1.3) is important enough to make some general comments: In Barwise (1993), a realised situation was called a *fact*.

MORE COMMENTS !!!

Modelling possible states of the system by partitions of the type set brings in a presupposition, namely that "distinct ways of settling the issues must settle some one issue in distinct ways" (Barwise 1997: 507). This means that states in which all situations carry the same information (so that they are of the same types) are identified. We will later, in sct. 8.1.3, discuss this assumption in more detail.

States that are realised by normal situations are automatically possible (for we have $\vdash_{\mathcal{C}} \subset \vdash_{N_{\mathcal{C}}}$ by definition).

Because all sequents are supposed to be information about the normal situations, every state realized by a normal situation is possible and hence every normal situation is a possible situation. We also have the following:

Theorem 7.1.4. *Let an information context $\mathcal{C}$ be given. A sequent $\langle \Gamma, \Delta \rangle$ holds in a situation s with* $\mathsf{state}(s) = \langle \Gamma_s, \Delta_s \rangle$ $(\Gamma \vdash_s \Delta)$ *iff the following holds: if $\Gamma \subset \Gamma_s$ then $\Delta \cap \Gamma_s \neq \emptyset$.*[1]

PROOF This follows from the fact that $\mathsf{state}(s)$ is a partition. □

Information contexts are meant to model the situations in which agents reason and to give us a handle on the information they have available. We will see later, e.g., how the acquisition of new information corresponds to a change in the corresponding information contexts. As theoreticians, however, we do not only reason *within* information contexts, but would like to be able to reason *about* them as well. We would like to uncover regularities, principles, structural features of the way *other* agents reason. This is why epistemic logic is interested in a *logic* of information contexts. A logic is concerned with structural or formal, i.e. topic-neutral relations between propositions (which may include propositions about the agents in question). But what are propositions in our framework? One possibility would be to go for Austinian propositions, i.e. entities of the kind $s \models \langle \Gamma, \Delta \rangle$. The problem with this approach, however, that it is not clear how to internally interpret sentence operators, i.e. to allow for the possibility that the agents themselves reason about what they or others know and believe. For a logic of epistemic behaviour, however, this possibility is crucial. It therefore seems more appropriate to go for Russellian propositions which may be conceived of as sets of states.

[1]There is a typo in the statement of the corresponding theorem in Barwise (1997: 508).

Within an information context, we may thus take a proposition to be a set of states or, equivalently, an element of the Boolean closure of the underlying classification. We may say that a proposition ϕ is *true of* a situation s iff $\mathsf{state}(s) \in \phi$. We may then define valuations of propositional letters to be subsets of the set of all states and extend this to Boolean compounds in the ordinary way:

Definition 7.1.5 (Truth in information contexts). *A* valuation V *of an information context* $\mathcal{C}$ *is a function from a set of propositional constants* $\mathbb{P}$ *to* $\mathcal{P}(\Omega_{\mathcal{C}})$. *Truth of a proposition in an information context under a valuation* V *is then defined as follows:*

$$\begin{array}{rcl}
(\mathcal{C}, s) \models_V p & :\Longleftrightarrow & \mathsf{state}(s) \in V(p) \\
(\mathcal{C}, s) \models_V \neg\phi & :\Longleftrightarrow & (\mathcal{C}, s) \not\models_V \phi \\
(\mathcal{C}, s) \models_V \phi \wedge \psi & :\Longleftrightarrow & (\mathcal{C}, s) \models_V \phi \ \textit{ and } \ (\mathcal{C}, s) \models_V \psi \\
(\mathcal{C}, s) \models_V \bigwedge \Phi & :\Longleftrightarrow & (\mathcal{C}, s) \models_V \phi \quad \textit{for all } \phi \in \Phi \\
(\mathcal{C}, s) \models_V \blacksquare\phi & :\Longleftrightarrow & (\mathcal{C}, t) \models_V \phi \quad \textit{for all possible situations } t
\end{array}$$

Validity of a formula ϕ under a valuation is truth in all situations, while validity *tout court* in an information context is validity under all valuations:

Definition 7.1.6 (Validity in information contexts). *A proposition* $\phi \in \mathcal{P}(\mathsf{tok}(\mathsf{cla}(\mathcal{C})))$ *is* valid *on information context* $\mathcal{C}$, *written* $\mathcal{C} \models_V \phi$, *iff it is true under* V *in all situations of that context, i.e. iff* $(\mathcal{C}, s) \models_V \phi$ *for all* $s \in \mathsf{tok}(\mathsf{cla}(\mathcal{C}))$. *It is valid in an information context, written* $\mathcal{C} \models \phi$, *iff it is valid under all valuations, i.e. iff* $(\mathcal{C}, s) \models_V \phi$ *for all* $V : \mathbb{P} \to \mathcal{P}(\Omega_{\mathcal{C}})$.

Please note that $\blacksquare$ is a rather peculiar kind of modality (hence its blackness): for one thing, it does not range over all states of the system, but only over the realised ones. It is "internal", registering which situations have states not excluded *by the information available to some agent at a given time.* There is no presumption, e.g., that $\mathcal{C} \models_V \blacksquare\phi$ whenever ϕ is true in all states or even all possible states. We will later, in ch. 8, discuss this modality in more detail.

Whenever we do not take the available information to cover more situations than it actually does, the information context is sound:

Definition 7.1.7. *An information context* $\mathcal{C}$ *is* sound *if every situation* $s \in S$ *is normal* $(N_{\mathcal{C}} = S)$. *The* sound part $\mathsf{Snd}(\mathcal{C})$ *of an information context* $\mathcal{C}$ *is defined as the restriction of the set of tokens* $\mathsf{cla}(\mathcal{C})$ *to the normal situations* $N_{\mathcal{C}}$.

It is clear that $\mathsf{Snd}(\mathcal{C})$ is sound and that $\mathcal{C}$ is sound iff $\mathsf{Snd}(\mathcal{C}) = \mathcal{C}$. Every normal situation is automatically possible. In an information context which is not sound, there are realized, but impossible situations. In a sound information context, every realised state is possible (i.e. not excluded by the available information): this means that the information available do not exclude too much or, equivalently, that there are not too many situations in focus. We will later see how this helps us to deal with the problem of logical omniscience (sct. 8.2).

Information contexts where every situation the available information speaks about is taken into consideration are called "complete":

Definition 7.1.8. *An information context $\mathcal{C}$ is* complete *if every possible state is realized by some normal situation (every sequence satisfied by every normal token is a constraint), i.e. if $\Gamma \nvdash_{\mathcal{C}} \Delta \Rightarrow \Gamma \nvdash_{N_{\mathcal{C}}} \Delta$: every consistent sequence has a normal counterexample.*
The completion $\mathsf{Cmp}(\mathcal{C})$ *of an information context $\mathcal{C}$ is obtained by adding a new normal token $n_{\langle\Gamma,\Delta\rangle}$ for each consistent partition $\langle\Gamma,\Delta\rangle$ that is not the state description of a normal token of $\mathcal{C}$. The new token $n_{\langle\Gamma,\Delta\rangle}$ has type $\alpha \in \mathsf{typ}(\mathsf{cla}(\mathcal{C}))$ iff $\alpha \in \Gamma$.*

It follows that for complete information contexts, the two relations $\vdash_{N_{\mathcal{C}}}$ and $\vdash_{\mathcal{C}}$ are identical. In a complete information context, every possible state is realised: there are enough counterexamples to witness all the unexcluded possibilities; no regularities are left out of the account and it is not possible to postulate further regularities without falling into inconsistency. It is clear that $\mathsf{Cmp}(\mathcal{C})$ is complete and that an information context $\mathcal{C}$ is complete iff $\mathsf{Cmp}(\mathcal{C}) = \mathcal{C}$.

As with regular theories (cf. def. 6.1.6), we can also speak of consistent information contexts:

Definition 7.1.9. *An information context $\mathcal{C}$ is* consistent *if it has at least one possible state, i.e. if there is a sequent $\langle\Gamma,\Delta\rangle$ with $\Gamma \nvdash_{\mathcal{C}} \Delta$.*

This definition is conservative in the following way:

Theorem 7.1.10. *An information context is consistent iff its theory is consistent.*

Proof Let $\mathcal{C}$ be any information context. We show that

$$\vdash_{\mathcal{C}} \text{ is consistent} \quad \Longleftrightarrow \quad N_{\mathcal{C}} \neq \emptyset$$

By (6.1.7) we know that $\vdash_{\mathcal{C}}$ is consistent iff $\emptyset \nvdash_{\mathcal{C}} \emptyset$. Every situation is of all the types in $\emptyset$, no situation is of one of the types in $\emptyset$. So $\emptyset \vdash_{\mathcal{C}} \emptyset$ iff there are no normal situations. □

What characterises natural and formal information contexts (i.e. contexts generated by classifications and regular theories)? Due to our representation theorem, we have the following relation between regular theories and and the information contexts of their classifications:

Theorem 7.1.11. $\mathsf{IC}(T) = \mathsf{IC}(\mathsf{Cla}(T))$, *thus every formal logic is natural.* $A \cong \mathsf{Cla}(\mathsf{Th}(A))$ *for a separated classification* A, *thus every natural logic on a separated classification is isomorphic to a formal logic.*

Proof The first claim is obvious. The second follows from th. 6.1.39. □

Information contexts on classifications are unique and automatically sound and complete:

Theorem 7.1.12. *For any information context* $\mathcal{C}$ *on a classification* A*:*

$$\mathcal{C} \textit{ is natural} \iff \mathcal{C} \textit{ is sound and complete} \iff \mathcal{C} = \mathsf{IC}(A).$$

Proof Obvious. □

Sound restrictions and completions of classifications preserve the completeness or soundness of the original classifications respectively:

Theorem 7.1.13. *For any information context* $\mathcal{C}$*:*

$$\begin{array}{lrcl}
(1) & \mathsf{Snd}(\mathcal{C}) \textit{ is complete} & \iff & \mathcal{C} \textit{ is complete} \\
(2) & \mathsf{Cmp}(\mathcal{C}) \textit{ is sound} & \iff & \mathcal{C} \textit{ is sound} \\
(3) & \mathsf{Snd}(\mathsf{Cmp}(\mathcal{C})) & = & \mathsf{Cmp}(\mathsf{Snd}(\mathcal{C}))
\end{array}$$

Proof

[(1):] For completeness, only the normal tokens are relevant which are not concerned by a restriction to the sound part.

[(2):] This holds because all tokens added in a completion are normal.

[(3):] Because every token we add to get the completion is normal, it does not matter whether we first throw away the non-normal tokens and then add counterexamples for every consistent sequence or the other way round. So the underlying classifications are the same. Because of (1), $\mathsf{Snd}(\mathsf{Cmp}(\mathcal{C}))$ is complete; because of (2), $\mathsf{Cmp}(\mathsf{Snd}(\mathcal{C}))$ is sound. By (7.1.12), there is only one sound and complete information context based on a fixed classification and hence the two are identical. □

The last identity of (7.1.13) justifies the following definition:

Definition 7.1.14. *The* sound completion $\mathsf{SC}(\mathcal{C})$ *of an information context* $\mathcal{C}$ *is given by* $\mathsf{Snd}(\mathsf{Cmp}(\mathcal{C}))$.

7.1.2 Moving and comparing information contexts

Channels, systems of classifications connected by infomorphisms, may be ordered with respect to a relation of refinement defined as follows:

Definition 7.1.15. *The channel* $\mathfrak{C}' = \{g_i : A_i \rightleftarrows C'\}_{i \in I}$ *is a* refinement *of the channel* $\mathfrak{C} = \{f_i : A_i \rightleftarrows C\}_{i \in I}$ *if there is a refinement isomorphism* $r : C' \rightleftarrows C$ *such that for each* $i \in I$ *it holds that* $f_i = rg_i$.

The channel $\{A_C \rightleftarrows A_{TPI}, A_T \rightleftarrows A_{TPI}\}$ of the embeddings of **K** and **T** into **S4**, for example, is a refinement by $A_{TI} \rightleftarrows A_{TPI}$ of the channel $\{A_C \rightleftarrows A_{TPI}, A_T \rightleftarrows A_{TII}\}$, embedding **K** and **T** into **S4**.

As there is a refinement ordering on channels, there is also an ordering of information contexts. Taken together, this two orderings allow us to distinguish between two quite different kinds of "grades of informativity". We use refinements of channels, e.g., in what sense the minimal cover of a channel is "minimal" (cf. th. 5.2.5). The more refined a channel, the less information it carries: if we add, e.g., new tokens to a channel and keep the same types, we have a refinement infomorphism from the richer to the poorer core classification. So the richer channel is more refined, i.e. covers more situations and thus exhibits less regularities, i.e. underwrites less flow of information.

As with regular theories (def. 6.1.27), we may compare information contexts with respect to their expressiveness:

Definition 7.1.16. *We define a* partial order on information contexts $\mathcal{C}_1$ *and* $\mathcal{C}_2$*:*

$$(4) \quad \mathcal{C}_1 \sqsubseteq \mathcal{C}_2 \quad :\Longleftrightarrow \quad (\Gamma \vdash_{\mathcal{C}_1} \Delta \Rightarrow \Gamma \vdash_{\mathcal{C}_2} \Delta) \;\wedge\; N_{\mathcal{C}_1} \subset N_{\mathcal{C}_2}$$

We read $\mathcal{C}_1 \sqsubseteq \mathcal{C}_2$ *as "*$\mathcal{C}_1$ *is less informative than* $\mathcal{C}_2$*".*

The first requirement is just that the $\mathcal{C}_1$-theory is contained in the $\mathcal{C}_2$-theory in the sense of $\sqsubseteq$ defined by (4) on p. 190. The following theorem justifies our reading of $\mathcal{C}_1 \sqsubseteq \mathcal{C}_2$:

Theorem 7.1.17. *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two information contexts on a classification A and $f : \mathcal{C}_1 \rightleftarrows \mathcal{C}_2$ a context infomorphism. If $\mathcal{C'}_1$ and $\mathcal{C'}_2$ are two information contexts with the same classifications than $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively, we have the following:*

$$(5) \quad \mathcal{C'}_1 \sqsubseteq \mathcal{C}_1 \wedge \mathcal{C}_2 \sqsubseteq \mathcal{C}'_2 \qquad \Longrightarrow \qquad f : \mathcal{C'}_1 \rightleftarrows \mathcal{C'}_2 \text{ is an context infomorphism}$$

Proof To be an context infomorphism, f has to be a theory interpretation, i.e. to preserve all constraints of $\mathcal{C}'$. By $\mathcal{C'}_1 \sqsubseteq \mathcal{C}_1$ every such constraint is also a constraint of $\mathcal{C}$ and thus preserved by $f : \mathcal{C}_1 \rightleftarrows \mathcal{C}_2$. By $\mathcal{C}_2 \sqsubseteq \mathcal{C'}_2$, every constraint of $\mathcal{C}_2$ is also a constraint of $\mathcal{C'}_2$. By $\mathcal{C}_2 \sqsubseteq \mathcal{C'}_2$, we have $N_{\mathcal{C'}_2} \subset N_{\mathcal{C}_2}$, hence $f(N_{\mathcal{C'}_2}) \subset f(N_{\mathcal{C}_2})$. Chaining this with $f(N_{\mathcal{C}_2}) \subset N_{\mathcal{C}_1}$ (because $f : \mathcal{C}_1 \rightleftarrows \mathcal{C}_2$ is an infomorphism) and $N_{\mathcal{C}_1} \subset N_{\mathcal{C'}_1}$ (because of $\mathcal{C'}_1 \sqsubseteq \mathcal{C}_1$), we get $f(N_{\mathcal{C'}_2}) \subset N_{\mathcal{C'}_1}$, the desired result about the normal tokens. □

Infomorphisms are supposed to preserve information: so if there is a way to preserve the information captured by an information context $\mathcal{C}_1$ in another information context $\mathcal{C}_2$, then this should equally be possible if we take a less informative context for $\mathcal{C}_1$ and a more informative context for $\mathcal{C}_2$: this is indeed what the theorem 7.1.17 states. The connection between the $\sqsubseteq$ and there being an infomorphism runs even deeper than this. We will come back to this in sct. 7.1.2.

As in the case of theories, we have the following theorem (Barwise and Seligman 1997: 158):

Theorem 7.1.18. *The ordering $\sqsubseteq$ on information contexts based on the same classification is a complete lattice.*

Barwise (1997: 509) distinguishes two ways of information update and shows how they correspond to changes in the subsequent information contexts of the agent. He discusses the first form of information update under the heading "information from impossibilies". We get information from impossibilities when we acquire the information that some state $\langle \Gamma, \Delta \rangle$ is not in fact possible. We may, e.g., note that $\langle \Gamma, \Delta \rangle$ is not realised by any normal situation and then add $\langle \Gamma, \Delta \rangle$ as a sequent to our regular theory. If $\mathcal{C}$ is our original information context and $\mathcal{C}'$ the least information context that contains the constraint $\langle \Gamma, \Delta \rangle$, we have $\mathcal{C} \sqsubseteq \mathcal{C}'$, i.e. the new information context contains more information than the old.

The second kind of information update is "throwing away misinformation". This happens when we encounter a situation made impossible (ruled out) by our regular theory. If we denote by "$\mathcal{C}[s]$" the least information context on the same classification which contains s as a normal situation, i.e. $\mathcal{C}[s] := \sqcap\{\mathcal{C}' \mid \mathcal{C}' \sqsubseteq \mathcal{C} \wedge s \in N_{\mathcal{C}'}\}$,[2] we have $\mathcal{C}[s] \sqsubseteq \mathcal{C}$, i.e. the new context contains less information than the old. This is not exactly true, for what has been left out (i.e. the sequents realised by s) did not constitute *information*, but *misinformation* (which, as we argued in sct. 1.4.5, is not a kind of information). This indicates that we have not yet captured the veracity requirement on information and we will have to return on this issue below (sct. 8.2).

A third way of information update, discussed not in Barwise (1997) but in Barwise and Seligman (1997) is the realisation that some type is universal among a given set of tokens and thus to dispense with it. This leads us to the following definition:

Definition 7.1.19. *Let an information context $\mathcal{C}$ and some set of types $\Theta \subset \Sigma_{\mathcal{C}}$ be given. The* conditionalisation $\mathcal{C}|\Theta$ of $\mathcal{C}$ on Θ *is the information context with the classification* $\mathsf{cla}(\mathcal{C}|\Theta) = \mathsf{cla}(\mathcal{C})$, *a regular theory given by:*

$$\Gamma \vdash_{\mathcal{C}|\Theta} \Delta \quad :\Longleftrightarrow \quad \Gamma, \Theta \vdash_{\mathcal{C}} \Delta$$

and the normal tokens $N_{\mathcal{C}|\Theta} := \{s \in N_{\mathcal{C}} \mid s \models_{\mathsf{cla}(\mathcal{C})} \theta$ for all $\theta \in \Theta\}$.

We have to justify that $\mathcal{C}|\Theta$ is indeed an information context.[3] We have the following theorem:

Theorem 7.1.20. *For every information context $\mathcal{C}$, there is a sound information context $\mathcal{C}'$ and a type θ such that $\mathcal{C} = \mathcal{C}'|\{\theta\}$.*

Proof Let $\theta \notin \Sigma_{\mathcal{C}}$ be a type such that $\mathsf{tok}(\theta) = N_{\mathcal{C}}$ and then define the required context by $\mathsf{cla}(\mathcal{C}') := \langle \mathsf{tok}(\mathsf{cla}(\mathcal{C})), \mathsf{typ}(\mathsf{cla}(\mathcal{C})) \cup \{\theta\}, \models_{\mathcal{C}}\rangle$. We show that the following defines a regular theory:

$$(1) \quad \Gamma \vdash_{\mathcal{C}'} \Delta \quad :\Longleftrightarrow \quad (\theta \notin \Gamma \wedge \Gamma \cap \Delta \neq \emptyset) \vee (\theta \in \Gamma \cap \Delta) \vee (\theta \in \Gamma \setminus \Delta \wedge \Gamma \setminus \{\theta\} \vdash_{\mathcal{C}} \Delta)$$

(**Weakening**) is clear. For (**Partition**), suppose $\Gamma \nvdash_{\mathcal{C}'} \Delta$. Thus we have $(\theta \in \Gamma \vee \Gamma \cap \Delta = \emptyset) \wedge (\theta \notin \Gamma \cap \Delta) \wedge (\theta \notin \Gamma \setminus \Delta \vee \Gamma \setminus \{\theta\} \nvdash_{\mathcal{C}} \Delta)$. Suppose $\theta \notin \Gamma$. Then $\Gamma \cap \Delta = \emptyset$ and

[2] Barwise (1997: 510) proves that $\mathcal{C}[s]$ is indeed the largest context such that s is normal in it, that its constraints are those constraints of $\mathcal{C}$ that are satisfied by s and that contains as normal situations the normal situations of $\mathcal{C}$ and s, i.e. that $\mathcal{C}[s] = \langle \mathsf{cla}(\mathcal{C}), \vdash_{\mathcal{C}} \cap \vdash_s, N_{\mathcal{C}} \cup \{s\}\rangle$.

[3] To show that $\vdash_{\mathcal{C}|\Theta}$ defines a regular theory, observe that $\Gamma \vdash_{\mathcal{C}|\Theta} \Delta \Rightarrow \Gamma, \Theta \vdash_{\mathcal{C}} \Delta \Rightarrow \Gamma, \Theta, \Gamma' \vdash_{\mathcal{C}} \Delta, \Delta' \Rightarrow \Gamma, \Gamma' \vdash_{\mathcal{C}|\Theta} \Delta, \Delta'$. Thus (**Weakening**) holds. For (**Partition**), suppose that $\Gamma \nvdash_{\mathcal{C}|\Theta} \Delta$. We thus have $\Gamma, \Theta \nvdash_{\mathcal{C}} \Delta$ and find, by (**Partition**) for $\vdash_{\mathcal{C}}$ a consistent partition $\langle \Gamma', \Delta' \rangle$ extending $\langle \Gamma \cup \Theta, \Delta \rangle$. Hence we have $\Gamma'', \Theta \nvdash_{\mathcal{C}} \Delta'$ for $\Gamma'' := \Gamma' \setminus \Theta$ and $\langle \Gamma'', \Delta' \rangle$ extends $\langle \Gamma, \Delta \rangle$.

$\langle\Gamma, \mathsf{typ}(\mathsf{cla}(\mathcal{C})) \setminus \Gamma\rangle$ is the required consistent partition extending $\langle\Gamma, \Delta\rangle$. If $\theta \in \Gamma$, on the other hand, we have $\theta \notin \Delta$ and there is a consistent partition $\langle\Gamma', \Delta'\rangle$ extending the consistent sequent $\Gamma \setminus \{\theta\} \vdash_{\mathcal{C}} \Delta$. Then $\langle\Gamma' \cup \{\theta\}, \Delta'\rangle$ is the required consistent partition. $\square$

(7.1.20) tells us that non-sound information contexts, i.e. contexts which include non-normal tokens and the constraints of which are too strong, may be seen as local versions (restrictions to some background type) of sound information contexts. Conditionalisation preserves the kl-ordering in the following way:

Theorem 7.1.21. *Let $\mathcal{C}$ be an information contexts and $\Theta_0 \subset \Theta_1 \subset \mathsf{typ}(\mathsf{cla}(\mathcal{C}))$ two sets of types. Then we have $\mathcal{C}|\Theta_0 \sqsubseteq \mathcal{C}|\Theta_1$.*

PROOF $\vdash_{\mathcal{C}|\Theta_0} \subset \vdash_{\mathcal{C}|\Theta_1}$ follows from (**Weakening**). We have $\{\alpha \in \mathsf{typ}(\mathsf{cla}(\mathcal{C}) \mid a \models_{\mathcal{C}} \alpha\ \forall \alpha \in \Theta_0\} \subset \{\alpha \in \mathsf{typ}(\mathsf{cla}(\mathcal{C}) \mid a \models_{\mathcal{C}} \alpha\ \forall \alpha \in \Theta_1\}$ and hence $N_{\mathcal{C}|\Theta_1} = N_{\mathcal{C}} \cap \{\alpha \mid a \models_{\mathcal{C}} \alpha\ \forall \alpha \in \Theta_1\} \subset N_{\mathcal{C}} \cap \{\alpha \mid a \models_{\mathcal{C}} \alpha\ \forall \alpha \in \Theta_0\} = N_{\mathcal{C}|\Theta_0}$. $\square$

As for classifications and theories, we may also define sums for information contexts.[4] Quotients are given by the following definition:

Definition 7.1.22. *The* quotient context $\mathcal{C}/I$ *of $\mathcal{C}$ by the invariant $I = \langle\Sigma, R\rangle$ is the information context with classification* $\mathsf{cla}(\mathcal{C})/I$, *theory* $\mathsf{th}(\mathcal{C}) \mid \Sigma$ *and* $N_{\mathcal{C}/I} = \{[a]_R \mid a \in N_{\mathcal{C}}\}$. *The* quotient context $\mathcal{C}/J$ *of $\mathcal{C}$ by the dual invariant $J = \langle A, R\rangle$ is the information context with classification* $\mathsf{cla}(\mathcal{C})/J$, *theory* $\mathsf{th}(\mathcal{C})/R$ *and* $N_{\mathcal{C}/J} = A \cap N_{\mathcal{C}}$.

This definition gives us indeed information contexts (cf. Barwise and Seligman 1997: 160). They behave in the following way:

Theorem 7.1.23. *Let $\mathcal{C}/J$ of $\mathcal{C}$ be an information context and $J = \langle A, R\rangle$ a dual invariant on* $\mathsf{cla}\mathcal{C}$. *Then we have the following:*

(6) $\quad \mathcal{C}/J$ *is sound* $\iff A \subset N_{\mathcal{C}}$

(7) $\quad \mathcal{C}/J$ *is complete* $\iff \forall\langle\Gamma, \Delta\rangle \in \mathbf{Part}(\mathsf{cla}(\mathcal{C}))\ \forall \alpha \in \mathsf{typ}(\mathsf{cla}(\mathcal{C}))\ \exists s \in A \cap N_{\mathcal{C}} :$
$((\Gamma \nvdash_{\mathsf{th}(\mathcal{C})} \Delta \wedge \alpha \in \Gamma \to [\alpha]_R \in \Gamma)) \to \langle\Gamma, \Delta\rangle = \mathsf{state}(s)$

[4] Barwise and Seligman (1997: 156–157) prove that for any classifications A and B: $\mathsf{IC}(A + B) = \mathsf{IC}(A) + \mathsf{IC}(B)$, where the sum of two contexts is the sum of the classifications and the theories and the Cartesian product of the two sets of normal tokens. They also prove that the pointwise sum of some context infomorphisms is a context infomorphism.

Proof

[(6):] The classification $\mathsf{cla}(\mathcal{C})/J$ has A as its set of tokens.

[(7):] By (**Partition**), if $\Gamma \nvdash_{\mathcal{C}/J} \Delta$, then there is a consistent partition of $\mathsf{cla}(\mathcal{C})/J$ extending $\langle \Gamma, \Delta \rangle$. It follows from the definition of $\mathsf{cla}(\mathcal{C})/J$ that such a partition of A/J exists iff there is a consistent partition $\langle \Gamma', \Delta' \rangle$ of A such that if $\alpha, \beta \in \mathsf{typ}(A)$ and $\alpha R \beta$, then $\alpha \in \Gamma \Leftrightarrow \beta \in \Gamma$. So $\mathcal{C}/J$ is complete iff any such partition has a normal counterexample. □

From (6) it follows that $\mathcal{C}/J$ is sound if $\mathcal{C}$ is sound. From (7), we may conclude that, if $\mathcal{C}$ is complete and $N_{\mathcal{C}} \subset A$, then $\mathcal{C}/J$ is complete too.

The next crucial definition shows us how we can "lift" classification, state space and regular theory infomorphisms in the natural way to the corresponding information contexts:

Definition 7.1.24. *A* context infomorphism $f : \mathcal{C}_1 \rightleftarrows \mathcal{C}_2$ *is a contravariant pair of functions such that:*

- $f : \mathsf{cla}(\mathcal{C}_1) \rightleftarrows \mathsf{cla}(\mathcal{C}_2)$ *is an infomorphism of classifications.*
- $f^{\wedge} : \mathsf{th}(\mathcal{C}_1) \to \mathsf{th}(\mathcal{C}_2)$ *is a theory interpretation.*
- $f^{\vee}(N_{\mathcal{C}_2}) \subset N_{\mathcal{C}_1}$.

Given a classification infomorphism $f : A \rightleftarrows B$, $\mathsf{IC}(f)$ *is the corresponding information context infomorphism* $\mathsf{IC}(f) : \mathsf{IC}(A) \rightleftarrows \mathsf{IC}(B)$.
Given a state space projection $f : S_1 \rightrightarrows S_2$, $\mathsf{IC}(f)$ *is the corresponding information context infomorphism* $\mathsf{IC}(f) : \mathsf{IC}(S_2) \rightleftarrows \mathsf{IC}(S_1)$, *which is the same as* $\mathsf{IC}(\mathsf{Evt}(f))$.
Given a regular theory interpretation $f : T_1 \to T_2$, $\mathsf{IC}(f)$ *is the corresponding information context infomorphism* $\mathsf{IC}(f) : \mathsf{IC}(T_1) \rightleftarrows \mathsf{IC}(T_2))$ *which is on types just* f *and maps every consistent partition* $\langle \Gamma, \Delta \rangle$ *of* T_2 *to* $\langle f^{-1}(\Gamma), f^{-1}(\Delta) \rangle$.

(7.1.24) shows us how we can "translate" the information represented by one information context into another information context, embedding the classification, translating the constraints and having normal situations whose corresponding tokens in the original classification were normal too. It thus describes the flow of information about information and gives us a *logic* of reasoning at a distance, where one information context is used to go proxy for another information context with whom it is linked by a context infomorphism.

Using the notion of the inverse image of a theory (def. 6.1.31 in sct. 6.1.4), we may now define the inverse image of an information context under a context infomorphism as follows:

Definition 7.1.25. *The* inverse image $f^{-1}(\mathcal{C})$ *of an information context* $\mathcal{C}$ *on* B *under an infomorphism* $f : A \rightleftarrows B$ *is the information context with classification* A*, regular theory* $f^{-1}(\mathsf{th}(\mathcal{L}))$ *and with normal tokens* $f(N_{\mathcal{C}}) = \{a \in \mathsf{tok}(A) \mid a = f(b) \text{ for some } b \in N_{\mathcal{L}}\}$.

It is clear that every $a \in f(N_{\mathcal{C}})$ satisfies every constraint of $f^{-1}(\mathsf{th}(\mathcal{C}))$ and thus that $f^{-1}(\mathcal{C})$ really is a local logic. We adopt this definition for the following reason:

Theorem 7.1.26. *The inverse image* $f^{-1}(\mathcal{C})$ *of an information context* $\mathcal{C}$ *under an infomorphism* $f : A \rightleftarrows B$ *is the* $\sqsubseteq$*-greatest information context on* A *such that* f *is a logic infomorphism.*

PROOF

$f : f^{-1}(\mathcal{C}) \to \mathcal{C}$ because f is a theory interpretation and we have th. 6.1.31.

Suppose that $f : \mathcal{C} \to \mathcal{C}'$ is a context infomorphism. We have to show that $\mathcal{C} \sqsubseteq f^{-1}(\mathcal{C}')$. Suppose $\Gamma \vdash_{\mathcal{C}} \Delta$. Because f is a context infomorphism, we have $f(\Gamma) \vdash_{\mathcal{C}'} f(\Delta)$. By (5) in def. 6.1.31, we have $\Gamma \vdash_{f^{-1}(\mathcal{C})} \Delta$. To check the condition on normal tokens, suppose $a \in N_{f^{-1}(\mathcal{C}')}$. Thus $a = f(b)$ for some $b \in N_{\mathcal{C}'}$. Because f is a logic infomorphism, we conclude that $a \in N_{\mathcal{C}}$. □

The operation of taking inverse images preserves completeness; but soundness is only preserved if the original infomorphism was token surjective:

Theorem 7.1.27. *The inverse image of a sound information context under a token surjective infomorphism is sound. The inverse image of a complete information context is complete.*

PROOF

[1^{st} claim:] This is immediate, because f is a context infomorphism only if $f(N_{\mathcal{C}}) \subset N_{f^{-1}(\mathcal{C})}$, $f(N_{\mathcal{C}}) = f(\mathsf{tok}(B))$ if $\mathcal{C}$ is sound and $f(\mathsf{tok}(B)) = \mathsf{tok}(A)$ if f is token surjective.

[2^{nd} claim:] To prove the contraposition, suppose that $f^{-1}(\mathcal{C})$ is not complete, i.e. that there is a consistent sequent $\langle \Gamma, \Delta \rangle$ of $\mathsf{th}(f^{-1}(\mathcal{C}))$ satisfied by every $a \in N_{f^{-1}(\mathcal{C})}$. Because f is an infomorphism, it follows from th. 6.1.8 that $\langle f(\Gamma), f(\Delta) \rangle$ is satisfied by every $b \in \mathsf{tok}(B)$ such that $a = f(b)$. But this means that $\langle f(\Gamma), f(\Delta) \rangle$ is satisfied by every $b \in N_{\mathcal{C}}$. But it is consistent by the definition of the inverse image of a theory (6.1.31). □

We saw in th. 6.1.8 on p. 183 that theory interpretations preserve the satisfaction of sequents. What about context infomorphisms? Would it be appropriate, e.g., to introduce the following “rule” f-Elim, whenever we have an infomorphism $f : A \rightleftarrows B$?

$$\frac{f(\Gamma) \vdash_B f(\Delta)}{\Gamma \vdash_A \Delta} \quad f\text{-Elim}$$

Though f-Elim does not preserve validity in general,[5], it does so if we use it to reason only about tokens in $f(\mathsf{tok}(B)) \subset \mathsf{tok}(A)$. In any case, f-Elim preserves non-validity.[6]

Using the notion of an image of a theory (def. 6.1.33 in sct. 6.1.4), we may now define the image of an information context under a context infomorphism as follows:

Definition 7.1.28. *The* image $f(\mathcal{C})$ *of an information context* $\mathcal{C}$ *on* A *under an infomorphism* $f : A \rightleftarrows B$ *is the local logic with classification* B*, regular theory* $f(\mathsf{th}(\mathcal{C}))$ *and with normal tokens* $f^{-1}(N_{\mathcal{C}}) = \{b \in \mathsf{tok}(B) \mid f(b) \in N_{\mathcal{C}}\}$*.*

While it was clear that inverse images of information contexts are information contexts too, it has to be verified that (7.1.28) gives us indeed an information context:

Theorem 7.1.29. *The image* $f(\mathcal{C})$ *of an information context* $\mathcal{C}$ *is an information context.*

Proof $f(\mathsf{th}(\mathcal{C}))$ is a regular theory by definition 6.1.33. We have to show that $\vdash_{f(\mathcal{C})} \subset \vdash_{N_{f(\mathcal{C})}}$, i.e. that every sequent that has a normal counterexample is consistent. By (**Partition**), it is enough to show that if every partition extending some sequent is a constraint, then the sequent in question holds of every normal situation. So suppose $\Gamma \vdash_{f(\mathsf{th}(\mathcal{C}))} \Delta$ and $b \in N_{f(\mathcal{C})}$. By (6) in our definition of images of theories (6.1.33), we have $f^{-1}(\Gamma) \vdash_{\mathsf{th}(\mathcal{C})} f^{-1}(\Delta)$. Because $\mathcal{C}$ is an information context, this partition holds of $f(b) \in N_{\mathcal{C}}$. By tn. 6.1.8, it then follows that $\langle f(f^{-1}(\Gamma)), f(f^{-1}(\Delta))\rangle = \langle \Gamma, \Delta\rangle$ holds of $b \in N_{f(\mathcal{C})}$. □

We adopt our definition (7.1.28) for the following reason:

Theorem 7.1.30. *The image of an information context* $\mathcal{C}$ *under an infomorphism* $f : A \rightleftarrows B$ *is the* $\sqsubseteq$*-least information context on* B *such that* f *is a context infomorphism.*

Proof To show that $f : \mathcal{C} \to f(\mathcal{C})$ is a context infomorphism, suppose $\Gamma \vdash_{\mathcal{C}} \Delta$ and $\langle \Gamma', \Delta'\rangle$ is a partition extending $\langle f(\Gamma), f(\Delta)\rangle$. Because f is a function, $\langle f^{-1}(\Gamma'), f^{-1}(\Delta')\rangle$ is a

[5] This means that there may well be classification A and B and an infomorphism $f : A \rightleftarrows B$ such that $f(\Gamma) \vdash_B f(\Delta)$ but $\Gamma \nvdash_A \Delta$. No counterexample $a \in \mathsf{tok}(A)$ to $\Gamma \nvdash_A \Delta$, however, may be in the image $f(\mathsf{tok}(B))$ of B.

[6] This means for any classifications A and B and infomorphisms $f : A \leftrightarrows B$, if $f(\Gamma) \nvdash_B f(\Delta)$ then $\Gamma \nvdash_A \Delta$. For if $b \in \mathsf{tok}(B)$ is a counterexample to the former sequent, then, by (fund), $f(b)$ will be a counterexample to the latter.

partition and it obviously extends $\langle\Gamma,\Delta\rangle$. So we have $f^{-1}(\Gamma')\vdash_{\mathcal{C}} f^{-1}(\Delta')$ by (**Weakening**) and $\Gamma'\vdash_{f(\mathcal{C})}\Delta'$ by the definition of the image of a theory (6.1.33). Because $\langle\Gamma',\Delta'\rangle$ was arbitrary, we conclude by (**Partition**) that $f(\Gamma)\vdash_{f(\mathcal{C})} f(\Delta)$

Suppose that $f:\mathcal{C}\to\mathcal{C}'$ is a context infomorphism. We have to show that $f(\mathcal{C})\sqsubseteq\mathcal{C}'$. Because f is a theory interpretation, the condition on the regular relations is satisfied. Suppose $b\in N_{\mathcal{C}'}$. Because f is a logic infomorphism, $f(b)\in N_{f(\mathcal{C})}$. □

Theorem 7.1.31. *The image of a sound information context is sound. The image of a complete information context under a token and type surjective infomorphism is complete.*[7]

PROOF

[1^{st} claim:] If the original information context is sound, every $f(b)$ for some token $b\in\mathsf{tok}(B)$ is normal in A. So every $b\in\mathsf{tok}(B)$ is in $f^{-1}(N_{\mathcal{C}})$, i.e. normal in $f(\mathcal{C})$.

[2^{nd} claim:] We have to show that every $f(\mathcal{C})$ consistent sequent has a normal counterexample. Suppose $\langle\Gamma,\Delta\rangle$ is a consistent sequent of $f(\mathcal{C})$. By (**Partition**), there is a consistent partition $\langle\Gamma',\Delta'\rangle$ extending it. Because f is a theory interpretation, we have $f^{-1}(\Gamma)\nvdash_{\mathcal{C}} f^{-1}(\Delta)$ by th. 6.1.10. By the completeness of $\mathcal{C}$, there is a counterexample $a\in N_{\mathcal{C}}$ for $\langle f^{-1}(\Gamma),f^{-1}(\Delta)\rangle$. By the token surjectivity of f, $a=f(b)$ for some $b\in\mathsf{tok}(B)$. By the definition of $f(\mathcal{C})$, $b\in N_{f(\mathcal{C})}$. By (6.1.8), b is a counterexample for $\langle f(f^{-1}(\Gamma)),f(f^{-1}(\Delta))\rangle$. Because f is type surjective, $\langle f(f^{-1}(\Gamma)),f(f^{-1}(\Delta))\rangle$ is a partition extending $\langle\Gamma,\Delta\rangle$. So b is also a counterexample to $\langle\Gamma,\Delta\rangle$. □

We are now prepared for a discussion of the "rule" f-Intro (for an infomorphism $f:A\rightleftarrows B$):

$$\frac{f^{-1}(\Gamma)\vdash_A f^{-1}(\Delta)}{\Gamma\vdash_B\Delta}\quad f\text{-Intro}$$

It may be seen that, due to the fundamental property of infomorphisms, f-Intro preserves

[7]The corresponding theorem 13.4 in Barwise and Seligman (1997: 166–167) is false, for they do not require type surjectivity. It seems that they had forgotten about the disanalogy we remarked on p. 193 between images and inverse images. If f is not type surjective, we just cannot conclude anything on $\langle\Gamma,\Delta\rangle$ from the fact that $\langle f(f^{-1}(\Gamma)),f(f^{-1}(\Delta))\rangle$ has a normal counterexample. To see that type surjectivity is necessary, consider two classifications $A:=\langle\{a\},\{\alpha\},\{\langle a,\alpha\rangle\}\rangle$ and $B:=\langle\{b\},\{\beta_1,\beta_2\},\{\langle b,\beta_1\rangle\}\rangle$ and an infomorphism given by $\alpha\mapsto\beta_1, b\mapsto a$. If we define a theory on A by the consistent partition $\langle\{\alpha\},\emptyset\rangle$ and have $N_A:=\{\alpha\}$, this gives us a complete information context (for the only consistent sequent has a normal counterexample). The image of this information context, with consistent partitions $\langle\{\beta_1,\beta_2\},\emptyset\rangle$ and $\langle\{\beta_1\},\{\beta_2\}\rangle$ (given by def. 6.1.33) and $N_B=\{b\}$ is not complete, for $\langle\{\beta_1,\beta_2\},\emptyset\rangle$ has no counterexample and hence no normal counterexample.

validity.[8] It does not, however, preserve non-validity:[9] Gödel proved the consistency of PA in ZFC, while the consistency of PA cannot be proven in PA itself, due to his second incompleteness theorem.

Let us now consider how our two defeasible "rules" f-Intro and f-Elim allow us to model the reasoning of an agent about a distributed system, i.e. the use he may make of information that is available to him in virtue of some channel $\mathfrak{C}$:

$$\begin{tikzcd}
 & C & \\
P \arrow[ur, "p"] & & D \arrow[ul, "d"']
\end{tikzcd}$$

We saw on p. 197 that Barwise's and Seligman's "initial proposal" presupposed complete information about the core C of the channel and thus gave us a "God's eye analysis of information flow". It therefore did not do justice to the imperfection of ordinary reasoning:

> "Ordinary reasoning is not logically perfect; there are logical sins of commission (unsound inferences) and omission (inferences that are sound but not drawn). Modeling this, AI has had to cope with logics that are both unsoound and incomplete. These are the sorts of logics we need in order to model our less than perfect information about the core of a channel." (Barwise and Seligman 1997: 37)

Suppose our agent uses classification P to reason about what happens elsewhere, in a region described by some different classification D and he does not have complete knowledge about the core of the channel C in virtue he has information about what happens at D. He will reason as follows:

$$\cfrac{\cfrac{\Gamma \vdash_P \Delta}{p(\Gamma) \vdash_C p(\Delta)}\ f\text{-Intro}}{d^{-1}(p(\Gamma)) \vdash_D d^{-1}(p(\Delta))}\ f\text{-Elim}$$

The problem in the first step, the application of f-Intro, is that due to def. 6.1.33, only manages to capture all the constraints on C iff every token in P is the image of some

[8]If c were a counterexample to $\Gamma \vdash_B \Delta$, i.e. if $c \models_B \gamma$ for all $\gamma \in \Gamma$ but $c \not\models_B \delta$ for all $\delta \in \Delta$, then we have, by (fund), $f(c) \models_A \gamma'$ for all $\gamma' \in f^{-1}(\Gamma)$ but $f(c) \not\models_A \delta'$ for all $\delta' \in f^{-1}(\Delta)$, hence $f^{-1}(\Gamma) \vdash_A f^{-1}(\Delta)$.

[9]This means that there may well be classification A and B and an infomorphism $f : A \leftrightarrows B$ such that $f^{-1}(\Gamma) \nvdash_A f^{-1}(\Delta)$ but $\Gamma \vdash_B \Delta$. By def. 6.1.31 we have, however: if $f^{-1}(\Gamma) \nvdash_{f^{-1}(B)} f^{-1}(\Delta)$ then $f(f^{-1}(\Gamma)) \vdash_B f(f^{-1}(\Delta))$. So if at least one counterexample to $f^{-1}(\Gamma) \nvdash_A f^{-1}(\Delta)$ comes from a token $a \in \mathsf{tok}(A)$ which is the image of some $b \in \mathsf{tok}(B)$ and if $\Gamma \subset f(f^{-1}(\Gamma))$ and $\Delta \subset f(f^{-1}(\Delta))$ (for which f has to be type surjective), we may conclude that $\Gamma \nvdash_B \Delta$.

token of C and if the resulting sequents $\langle p(\Gamma), p(\Delta)\rangle$ are partitions iff their originals were (i.e. if f is type-surjective). The first step, however, does preserve the validity of the original constraint and it does, due to th. 7.1.31, preserve soundness, i.e. it does not introduce further non-normal tokens. f-Intro may thus used to translate a constraint about a component of the system, i.e. the classification P in our example, into a constraint about the system as a whole (i.e. the core C of the channel $\mathfrak{C}$) (Barwise and Seligman 1997: 39).

The problem in the second step, the application of f-Elim, is that due to def. 6.1.31, it gives us constraints on D that only hold for tokens which are images of some connecting token of C:

> "This is certainly a reasonable rule of inference, even if it might lead one astray were one to apply it to something that was not a component of the system, that is, even though it is not sound." (Barwise and Seligman 1997: 39)

(7.1.27) tells us that f-Elim does preserve completeness, however. So we can be sure that every constraint on the connecting classification C gives us a (limitedly applicable) constraint on D.

Summing up, we have the following picture: if our agent starts with a complete theory of classification P, he may, using f-Intro, get a sound, though not necessarily complete, theory of C. Pulling this back to D, he looses completeness as well, ending up with a logic of D that is neither guaranteed to be sound nor guaranteed to be complete:

> "A sequent about distal tokens [of D] obtained from a constraint about proximal tokens [of P] in this way is guaranteed to apply to distal tokens that are connected to a proximal tokens in the channel, but about other distal tokens we have no guarantees." (Barwise and Seligman 1997: 40)

We will study some further properties of the resulting logic in sct. 7.1.6.

Images and inverse images allow us to relate any information context as follows to its sound restriction and its completion:

Theorem 7.1.32. *Let $\mathcal{C}$ be an information context. Then ICC is the inverse image of its sound restriction* $\mathsf{Snd}(\mathcal{C})$ *under the type identical inclusion infomorphism* $\iota : \mathsf{cla}(\mathcal{C} \rightleftarrows \mathsf{cla}(\mathsf{Snd}(\mathcal{C}))$. *$\mathcal{C}$ is also the image of its completion* $\mathsf{Cmp}(\mathcal{C})$ *under the type identical inclusion infomorphism* $\kappa : \mathsf{cla}(\mathsf{Cmp}(\mathcal{C})) \rightleftarrows \mathsf{cla}(\mathcal{C})$.

PROOF

[1^{st} claim:] This follows from th. 7.1.31.

[2^{nd} claim:] This follows from th. 7.1.27. $\square$

We have the following:

Theorem 7.1.33. *The operations of taking images and inverse images preserve the $\sqsubseteq$-order on information contexts.*

PROOF Suppose A and B are two classification and $f : A \rightleftarrows B$ is an infomorphism. We have to show the following:

$$\begin{array}{lllll} (1) & \mathcal{C}_1 \text{ and } \mathcal{C}_2 \text{ are contexts on } A \ \wedge \ \mathcal{C}_1 \sqsubseteq \in_2 & \Longrightarrow & f(\mathcal{C}_1) \sqsubseteq f(\mathcal{C}_2) \\ (2) & \mathcal{C}_1 \text{ and } \mathcal{C}_2 \text{ are contexts on } B \ \wedge \ \mathcal{C}_1 \sqsubseteq \in_2 & \Longrightarrow & f^{-1}(\mathcal{C}_1) \sqsubseteq f^{-1}(\mathcal{C}_2) \end{array}$$

For (1), note that by th. 7.1.26, f is an infomorphism from $\mathcal{C}_2$ to $f(\mathcal{C}_2)$. Because we have $\mathcal{C}_1 \sqsubseteq \in_2$, it is also an infomorphism from $\mathcal{C}_1$ to $f(\mathcal{C}_2)$ by th. 7.1.17. By th. 7.1.26, $f(\mathcal{C}_1)$ is the smallest context making f an context infomorphism from $\mathcal{C}_1$. So we have $f(\mathcal{C}_1) \sqsubseteq f(\mathcal{C}_2)$.

For (2), note that by th. 7.1.30, f is an infomorphism from $f^{-1}(\mathcal{C}_1)$ to $\mathcal{C}_1$. Because we have $\mathcal{C}_1 \sqsubseteq \in_2$, it is also an infomorphism from $f^{-1}(\mathcal{C}_1)$ to $\mathcal{C}_2$ by th. 7.1.17. By th. 7.1.30, $f^{-1}(\mathcal{C}_2)$ is the largest context making f an context infomorphism to $\mathcal{C}_2$. So we have $f^{-1}(\mathcal{C}_1) \sqsubseteq f^{-1}(\mathcal{C}_2))$. Recall that, by th. 7.1.17 $\square$

We now see that taking images and inverse images, while it may lead us to more or less informative information contexts in general, nevertheless preserves the informativeness ordering between classfications. It can also be shown (Barwise and Seligman 1997: 171) that the operations of taking images and inverse images preserves joins in the $\sqsubseteq$-ordering.

Before developing further however the general theory, we will discuss some examples of information contexts.

7.1.3 Truth information contexts

7.1.4 Modal information contexts

The simplest way to define a Lewis information context is the following:

Definition 7.1.34. *A* Lewis information context $\mathcal{L}$ *is an information context which consists of:*

- *the Lewis classification A_W on W,*

- $N_{\mathcal{L}} = \mathsf{tok}(A)$
- *a binary relation defined by:*

$$\Gamma \vdash_{\mathcal{L}} \Delta \quad :\Longleftrightarrow \quad \forall w \in \mathsf{tok}(A_W)\ \forall \alpha \in \Gamma\ \exists \beta \in \Delta \quad (w \models \alpha \ \longrightarrow\ w \models \beta)$$

The definition gives us a sound and complete information context, representing complete and non-redundant information about a set of situations, couched in a "language" which is maximally expressive.

We can do the same for Kripke classifications:

Definition 7.1.35. *A* Kripke information context $\mathcal{K}$ *is an information context which consists of:*

- *the Kripke classification $A_{\mathcal{C}}$,*
- $N_{\mathcal{K}} = \mathsf{tok}(A)$
- *a binary relation defined by:*

$$\Gamma \vdash_{\mathcal{K}} \Delta \quad :\Longleftrightarrow \quad \forall w \in \mathsf{tok}(A_{\mathcal{C}})\ \forall \alpha \in \Gamma\ \exists \beta \in \Delta \quad (w \models \alpha \ \longrightarrow\ w \models \beta)$$

These definitions, though they give us sound and complete information contexts, are not very interesting, for the additional machinery, the set of normal tokens and the binary relation, are not doing any extra work. One way to do is is to restrict the set of normal situations to a subclass of epistemic states and adjusting the binary relation accordingly:

Definition 7.1.36. $\mathcal{K}_{\mathbf{T}} = \langle A, \vdash_{\mathbf{T}}, N_{\mathcal{T}} \rangle$ *is an information context which consists of:*

- *the Kripke classification A_C,*
- $N_{\mathcal{K}_{\mathbf{T}}} = \mathcal{T}$
- *a binary relation defined by:*

$$\Gamma \vdash_{\mathbf{T}} \Delta \quad :\Longleftrightarrow \quad \forall w \in \mathcal{T}\ \forall \alpha \in \Gamma\ \exists \beta \in \Delta (w \models \alpha \ \longrightarrow\ w \models \beta)$$

We may, in a completely parallel way, define $\mathcal{K}_{\mathbf{S4}}, \mathcal{K}_{\mathbf{S4}}, \mathcal{K}_{\mathbf{K45}}$ etc. by restricting the normal situations to positively introspective and reflective, introspective and reflective and just introspective epistemic states accordingly.

Another way of putting to work the machinery of information contexts is to define Kripke classifications with respect to one epistemic state:

Definition 7.1.37. *A* Kripke information context with respect to a possible world w and an agent a $\mathcal{K}_{w,a} = \langle A, \vdash, N \rangle$ *is an information context which consists of:*

- *the Kripke classification $A_{\mathcal{C}}$,*
- *a set N_w of normal tokens such that $w' \in N_w$ whenever $w' \in w(a)$.*
- *a binary relation defined by:*

$$(8) \quad \Gamma \vdash_{K_w} \Delta \quad :\Longleftrightarrow \quad \forall w' \; \forall \alpha \in \Gamma \; \exists \beta \in \Delta (w' \in w(a) \wedge w' \models \alpha \;\longrightarrow\; w' \models \beta)$$

(8) ensures that the information available in a Kripke information context is information about the worlds considered possible (and hence normal) by a in that context.

Lewis information contexts, e.g., are sound and complete, for we assume, as in the case of Lewis classification, a maximally expressive set of types and interpret conjunction, negation and intersection algebraically by intersection, complement and union respectively.

7.1.5 Boolean information contexts

In this section, I want to introduce Boolean operations on contexts and state the conditions under which an information context is Boolean. In order to do this, we would have – in a way parallel to what we did in sct. 5.1.6 with classifications – to introduce, for any Boolean operator $\mathcal{B} \in \{\neg, \wedge, \vee\}$ and any regular theory $T = \langle \Sigma, \vdash \rangle$, $\mathcal{B}T := \mathsf{Th}(\mathcal{B}\mathsf{Cla}(T))$ and define $\mathcal{B} : \Sigma \to \Sigma$ such that it is a negation, conjunction or disjunction respectively iff it is a theory interpretation from $\mathcal{B}T$ to T (Barwise and Seligman 1997: 143).
$\neg : \Sigma \to \Sigma'$ is a negation on $T = \langle \Sigma, \vdash \rangle$ iff it satisfies the following closure conditions:

$$\frac{\Gamma \vdash_T \Delta, \alpha}{\Gamma, \neg\alpha \vdash_T \Delta} \;\neg\text{-Left} \qquad\qquad \frac{\Gamma, \alpha \vdash_T \Delta}{\Gamma \vdash_T \Delta, \neg\alpha} \;\neg\text{-Right}$$

$\wedge : \Sigma \to \Sigma'$ is a conjunction on T iff it satisfies the following closure conditions:

$$\frac{\Gamma, \Theta \vdash_T \Delta}{\Gamma, \bigwedge\Theta \vdash_T \Delta} \;\wedge\text{-Left} \qquad\qquad \frac{\Gamma \vdash_T \Delta, \alpha \quad \text{for each } \alpha \in \Theta}{\Gamma \vdash_T \Delta, \bigwedge\Theta} \;\wedge\text{-Right}$$

$\neg : \Sigma \to \Sigma'$ is a disjunction on T iff it satisfies the following closure conditions:

$$\frac{\Gamma, \alpha \vdash_T \Delta \quad \text{for each } \alpha \in \Theta}{\Gamma, \bigvee\Theta \vdash_T \Delta} \;\vee\text{-Left} \qquad\qquad \frac{\Gamma \vdash_T \Delta, \Theta}{\Gamma, \bigvee\Theta \vdash_T \Delta} \;\vee\text{-Right}$$

Given Boolean theories and Boolean classifications, we may define Boolean contexts as follows:

Definition 7.1.38. *Let $\mathcal{C} = \langle \mathsf{cla}(\mathcal{C}), \vdash_{\mathcal{C}}, N_{\mathcal{C}} \rangle$ be an information context. We define:*

$$\begin{aligned}
\neg\mathcal{C} &:= \langle \neg\mathsf{cla}(\mathcal{C}), \vdash_{\neg\mathsf{th}(\mathcal{C})}, N_{\mathcal{C}} \rangle \\
\wedge\mathcal{C} &:= \langle \wedge\mathsf{cla}(\mathcal{C}), \vdash_{\wedge\mathsf{th}(\mathcal{C})}, N_{\mathcal{C}} \rangle \\
\vee\mathcal{C} &:= \langle \vee\mathsf{cla}(\mathcal{C}), \vdash_{\vee\mathsf{th}(\mathcal{C})}, N_{\mathcal{C}} \rangle
\end{aligned}$$

An information context $\mathcal{C}$ is Boolean iff there are context infomorphisms from $\neg\mathcal{C}$, $\wedge\mathcal{C}$ and $\vee\mathcal{C}$ to $\mathcal{C}$ respectively.

These constructions preserve soundness and completeness of the original classification and permute with the IC functor (Barwise and Seligman 1997: 162–163).

Boolean information contexts represent epistemic situations "in which complete information about classical propositional logic is available" (Barwise 1997: 507).

Analogy with Nec: complete information about modal logic

Both Lewis information contexts (7.1.34) and Kripke information contexts (7.1.35) are Boolean, their negation, conjunction and disjunction being given by complement, intersection and union respectively.

7.1.6 From channels to contexts

To introduce the general theory of contexts incorporating channels, we discuss further the example of a particular information context we met in our discussion of the two inference "rules" f-Elim and f-Intro in sct. 7.1.2 (on p. 212 and p. 213 respectively):

Definition 7.1.39. *The information context $\mathsf{IC}(\mathfrak{C}, D)$ on D* induced by the binary channel *$\mathfrak{C} = \{p : P \rightleftarrows C, d : D \rightleftarrows C\}$ is the information context $\mathsf{IC}(\mathfrak{C}, D) = d^{-1}(p(\mathsf{IC}(P))$.*

Thus $\mathsf{IC}(\mathfrak{C}, D)$ is the information context representing the reasoning, in P, of an epistemic agent about what is represented by a distal classification D of which he receives information in virtue of connecting tokens of the core C of the channel. We now have the following theorem, summing up our discussion in sct. 7.1.2:

Theorem 7.1.40. *For an information context $\mathsf{IC}(\mathfrak{C}, D)$ on D induced by the binary channel $\mathfrak{C} = \{p : P \rightleftarrows C, d : D \rightleftarrows C\}$, where p is type-surjective,[10] we have the following (where $\langle \Gamma, \Delta \rangle$ is some partition of $\mathsf{typ}(D)$):*

(9) $\Gamma \nvdash_{\mathsf{th}(\mathsf{IC}(\mathfrak{C},D))} \Delta \;\Leftrightarrow\; \exists b \in \mathsf{tok}(P) \,:\, \langle p^{-1}(d(\Gamma)), p^{-1}(d(\Delta)) \rangle = \mathsf{state}(b)$

(10) $d \in \mathsf{tok}(D) \in N_{\mathsf{IC}(\mathfrak{C},D)} \;\Leftrightarrow\; \exists c \in \mathsf{tok}(C), p \in \mathsf{tok}(P) \,:\, p(c) = p \wedge d(c) = d$

Proof

[(9):] Let $\langle\Gamma,\Delta\rangle$ be some partition of $\mathsf{typ}(D)$. By our def. 6.1.31 of images of regular theories, $\langle\Gamma,\Delta\rangle$ is consistent in $\mathsf{th}(\mathsf{IC}(\mathfrak{C},D))$ iff $\langle d(\Gamma),d(\Delta)\rangle$ is consistent in $\mathsf{th}(p(\mathsf{IC}(P)))$. By (**Partition**), $\langle d(\Gamma),d(\Delta)\rangle$ is consistent iff there is a $\mathsf{th}(p(\mathsf{IC}P)$-consistent partition $\langle\Gamma',\Delta'\rangle$ of $\mathsf{typ}(C)$ extending it. If there is such a partition, then, by def. 6.1.33, $p^{-1}(\Gamma')\nvdash_{\mathsf{th}(\mathsf{IC}(P))}$ $p^{-1}(\Delta')$ is consistent. Because this sequent extends $\langle p^{-1}(d(\Gamma)),p^{-1}(d(\Delta))\rangle$, we have$p^{-1}(d(\Gamma))\nvdash_{\mathsf{th}(\mathsf{IC}(P))}$ $p^{-1}(d(\Delta)$ by (**Weakening**). If, conversely, we have $p^{-1}(d(\Gamma))\nvdash_{\mathsf{th}(\mathsf{IC}(P))} p^{-1}(d(\Delta)$, it follows from (**Partition**) that there is a consistent partition $\langle\Gamma'',\Delta''\rangle$ of $\mathsf{typ}(P)$ extending it. We have $\Gamma''=p^{-1}(p(\Gamma''))$, so we get $p(\Gamma'')\nvdash_{\mathsf{th}(p(\mathsf{IC}(P)))} p(\Delta'')$ by (6.1.33). This sequent extends $\langle d(\Gamma),d(\Delta)\rangle$, so we have $d(\Gamma)\nvdash_{\mathsf{th}(p(\mathsf{IC}(P)))} d(\Delta)$ by (**Weakening**). Because $\mathsf{IC}(P)$ is sound and complete, there is a token $b\in\mathsf{tok}(P)$ with state description $\mathsf{state}(b)=\langle\Gamma,\Delta\rangle$ iff $\Gamma\nvdash_{\mathsf{th}(\mathsf{IC}(P))}\Delta$.

[(10):] By (7.1.25), the normal tokens of $\mathsf{IC}(\mathfrak{C},D)$ are those tokens $d\in\mathsf{tok}(D)$ that are in the image of d, i.e. $N_{\mathsf{IC}(\mathfrak{C},D)}=d(N_{p(\mathsf{IC}(P))}$. Because $\mathsf{IC}(P)$ is sound, we have $d(N_{p(\mathsf{IC}(P))}=d(\mathsf{tok}(\mathsf{cla}(\mathsf{IC}(P))))$. □

(7.1.40) characterises induced information contexts in the following way: their consistent partitions are those sequents whose 'pull-back' is the state description of some token in P; their normal tokens are those that are connected to some token in P.

We also have the following representation theorem:

Theorem 7.1.41. *Every information context $\mathcal{C}$ on a classification A is of the form $\mathsf{IC}(\mathfrak{C},A)$ for a binary channel $\mathfrak{C}$ linking A to the classification $\mathsf{Cla}(\mathsf{SC}(\mathcal{C}))$ of the sound completion of $\mathcal{C}$.*

Proof

□

Definition 7.1.42. *The* idealisation $\mathsf{Idl}(\mathcal{C})$ *of an information context $\mathcal{C}$ is given by:*

$$\begin{array}{rcl} \mathsf{typ}(\mathsf{Idl}(\mathcal{C})) & := & \mathsf{typ}(\mathsf{cla}(\mathcal{C})) \\ \mathsf{tok}(\mathsf{Idl}(\mathcal{C})) & := & \{\langle\Gamma,\Delta\rangle \mid \langle\Gamma,\Delta\rangle\in\mathbf{Part}(\mathsf{cla}(\mathcal{C}))\wedge\Gamma\nvdash_{\mathsf{th}(\mathcal{C})}\Delta\} \\ \langle\Gamma,\Delta\rangle\models_{\mathsf{Idl}(\mathcal{C})}\alpha & :\Longleftrightarrow & \alpha\in\Gamma \end{array}$$

[10]This restriction is missing from the corresponding theorem 14.2 in (Barwise and Seligman 1997: 175). As noted in fn. 7, it is needed however, for the characterisation to hold.

Definition 7.1.43. *The channel* $\mathsf{Cla}(\mathcal{C})$ representing *an information context* $\mathcal{C}$ *is the binary channel* $\mathsf{Cla}(\mathcal{C}) := \langle h_{\mathsf{cla}(\mathcal{C})} : \mathsf{cla}(\mathcal{C}) \rightleftarrows \mathsf{cla}(\mathsf{Snd}(\mathcal{C})), h_{\mathsf{Idl}(\mathcal{C})} : \mathsf{Idl}(\mathcal{C}) \rightleftarrows \mathsf{cla}(\mathsf{Snd}(\mathcal{C}))\rangle$ *with type-identical infomorphisms given by:*

$$h_{\mathsf{cla}(\mathcal{C})}(a) = a$$
$$h_{\mathsf{Idl}(\mathcal{C})}(a) = \mathsf{state}_{\mathcal{C}}(a)$$

The next theorem states that every information context $\mathcal{C}$ on a classification A is induced by the channel representing it

Theorem 7.1.44. *Let* $\mathcal{C}$ *be an information context on a classification* A*. Then we have:*

$$\mathsf{IC}(\mathsf{Cla}(\mathcal{C}, A) = \mathcal{C}$$

PROOF

□

What distributed systems are for classifications, information systems are for information contexts:

Definition 7.1.45. *An* information system $\mathcal{S}$ *is an indexed family* $\mathsf{IC}(\mathcal{S}) = \{\mathcal{C}_i\}_{i\in I}$ *of information contexts together with a set* $\mathsf{inf}(\mathcal{C})$ *of logic infomorphisms with domain and codomain in* $\mathsf{IC}(\mathcal{S})$*.*

Definition 7.1.46. *The* a priori context $\mathsf{AP}(A)$ *on a classification* A *is the* $\sqsubseteq$*-smallest information context on* A*. Its theory* $\mathsf{th}(\mathsf{AP}(A))$ *is given by:*

$$\Gamma \vdash_{\mathsf{th}(\mathsf{AP}(A))} \Delta \quad \Longleftrightarrow \quad \Gamma \cap \Delta \neq \emptyset$$

We have to check that $\mathsf{AP}(A)$, so defined, really is the $\sqsubseteq$-smallest (least informative) context on A. So suppose there is another information context $\mathcal{C}$ on A.

Definition 7.1.47. *The* Lindenbaum quotient $\mathsf{Lind}(\mathcal{C})$ *of an information context* $\mathcal{C}$ *is the quotient of* $\mathcal{C}$ *by the dual invariant* $\alpha R \beta \;:\Leftrightarrow\; \alpha \vdash_{\mathsf{th}(\mathcal{C})} \beta$*, which identifies types which are equivalent in* $\mathsf{th}(\mathcal{C})$*.*

Definition 7.1.48. *Let* $\mathcal{S}$ *be an information system with information contexts* $\mathcal{C}_i = \langle A_i, \vdash_{\mathcal{C}_i}, N_{\mathcal{C}_i}\rangle$ *and logic infomorphisms* g_i*. Let* $\mathfrak{C} = \{g_i : A_i \rightleftarrows C\}_{i\in I}$ *be the limit channel of the distributed system* $\langle\{A\}_{i\in I}, \{g_i : A_i \rightleftarrows C\}_{i\in I}\rangle$ *underlying* $\mathcal{S}$*.*

The distributed context $\mathsf{Dist}(\mathcal{S})$ *of* $\mathcal{S}$ *is the following information context on the sum* $A = \sum_{i \in I} A_i$*:*

$$\mathsf{Dist}(\mathcal{S}) := (\sum_{i \in I} g_i)^{-1}(\bigsqcup_{i \in I} g_i(\mathcal{C}_i))$$

Theorem 7.1.49. *Construct the greatest logic on A such that $g = \sum_{i \in I} g_i$ is a logic infomorphism by: Its normal tokens $\{c_i\}_{i \in I}$ are those such that each c_i is a normal token of L_i and $c_i = f(c_j)$ for each infomorphism $f : A_i \rightleftarrows A_j$.*

Definition 7.1.50. *The distributed logic of an information system where each classification A has (only) its a priori logic $\mathsf{AP}(A)$ can be shown to be the Lindenbaum logic of the* systemwide logic $\mathsf{Log}(\mathcal{A})$ *of a distributed system $\mathcal{A}$ as follows:*

- $\mathsf{Cla}(\mathsf{Log}(\mathcal{A})) = \sum_{i \in I} A_i$.
- $\mathsf{Th}(\mathsf{Log}(\mathcal{A}))$ *is the regular closure of the logic with all constraints of the form $\alpha \vdash f(\alpha)$ and $f(\alpha) \vdash \alpha$ for all infomorphisms f of $\mathcal{A}$.*
- $\{c_i\}_{i \in I} \in N_{\mathsf{Log}(\mathcal{A})} \iff (f : A_i \rightleftarrows A_j \wedge c_j \in \mathsf{tok}(A_j) \Rightarrow c_i = f(c_j))$.

Definition 7.1.51. *The distributed logic $\mathsf{DLog}_{\mathfrak{C}}(\mathcal{L})$ of an information channel $\mathfrak{C} = \{f_i : A_i \rightleftarrows C\}_{i \in I}$* generated *by a local logic $\mathcal{L}$ on the core C is the pull-back of the distributed logic of the channel viewed as a distributed system from the core C to the sum $\sum_{i \in I} A_i$.*

Definition 7.1.52. *The distributed logic $\mathsf{DLog}(\mathcal{S})$ of a state space system $\mathcal{S} = \{f_i : S \rightrightarrows S_i\}_{i \in I}$ is the local logic on the sum $\sum_{i \in I} \mathsf{Evt}(S_i)$ pulled back from $\mathsf{Evt}(S)$. Its constraints can be shown to be given by:*

$$\Gamma \vdash_{DLog(\mathcal{S})} \Delta \iff (\forall \sigma \in \mathsf{typ}(S) \quad \forall k \in I : \quad (f_k(\sigma) \in \bigcap \Gamma_k \Rightarrow f_k(\sigma) \in \bigcup \Delta_k))$$

Its normal tokens can be shown to consist of those sequences of components that are connected by some token of the core S of $\mathcal{S}$.

Definition 7.1.53. *If S is a state space, an S-*logic *is a logic $\mathcal{L}$ on the event classification $\mathsf{Evt}(S)$ such that $Log(S) \sqsubseteq \mathcal{L}$. A state σ of S is $\mathcal{L}$-*consistent *if $\{\sigma\} \nvdash_{\mathcal{L}}$, if there is a token in $\mathsf{tok}(\mathcal{L})$ of type σ. The* subspace *$S_{\mathcal{L}}$ of S determined by the S-logic $\mathcal{L}$ has $\Omega_{\mathcal{L}}$ as types and $N_{\mathcal{L}}$ as tokens.*

Theorem 7.1.54. *For every state space S and every S-logic $\mathcal{L}$, the set $\Omega_{\mathcal{L}}$ of $\mathcal{L}$-consistent states of S is the smallest set of states of S such that $\vdash_{\mathcal{L}} \Omega_{\mathcal{L}}$.*

Theorem 7.1.55. *For every state space S and every S-logic $\mathcal{L}$:*

$$\Gamma \vdash_{\mathcal{L}} \Delta \quad \Longleftrightarrow \quad \Gamma, \Omega_{\mathcal{L}} \vdash_{Log(S)} \Delta$$

Theorem 7.1.56. *For every state space S and every S-logic $\mathcal{L}_1$ and $\mathcal{L}_2$:*

$$\mathcal{L}_1 \sqsubseteq \mathcal{L}_2 \quad \Longleftrightarrow \quad S_{\mathcal{L}_2} \subseteq S_{\mathcal{L}_2}$$

Theorem 7.1.57. *For any space state S, $\mathcal{L} \mapsto S_{\mathcal{L}}$ is an order inverting bijection between the family of all S-logics and the family of all subspace of S.*

Definition 7.1.58. *A subspace S_0 of a state space S is* sound in S *if* $\mathsf{tok}(S_0) = \mathsf{tok}(S)$.

Theorem 7.1.59. *An S-logic is sound iff the associated subspace is sound in S.*
The image of a complete subspace under a state-space projection is complete.
The inverse image of a sound subspace under a state space projection is sound.

Definition 7.1.60. *The* state space $\mathsf{Ssp}(\mathcal{L})$ generated by a local logic $\mathcal{L}$ on a classification A *is the subspace of* $\mathsf{Ssp}(A)$ *whose normal tokens are the normal tokens of $\mathcal{L}$ and whose types are the $\mathcal{L}$-consistent partitions. Given a logic infomorphism $f : \mathcal{L}_1 \rightleftarrows \mathcal{L}_2$,* $\mathsf{Ssp}(f) : \mathsf{Ssp}(\mathcal{L}_2) \rightrightarrows \mathsf{Ssp}(\mathcal{L}_1)$ *is the restriction of* $\mathsf{Cla}(f)$ *to* $\mathsf{Ssp}(\mathcal{L}_2)$.

Theorem 7.1.61. *A local logic $\mathcal{L}$ on a classification A is sound iff* $\mathsf{Ssp}(\mathcal{L})$ *is a sound subspace of* $\mathsf{Ssp}(A)$.
It is complete iff $\mathsf{Ssp}(\mathcal{L})$ *is a complete subspace.*

7.1.62. *For any classification A, $\mathcal{L} \mapsto \mathsf{Ssp}(\mathcal{L})$ is an order inverting bijection between the set of logics on A and the set of subspaces of the associated state space* $\mathsf{Ssp}(A)$.

7.2 Information Flow

7.2.1 Normal worlds

normal worlds: whih satisfy the constraints for principled reasons, not by accident (Barwise and Seligman 1997: 150)

WHAT I HAVE IN THE CONCLUSION Information contexts give us a context-dependent notion of normal situations and thus solve the problem of conditional constraints.

Channels allow for the crucial distinction between normal and non-normal situations and thus account for all the features of information flow which depend on such a distinction.

They satisfy, in other words, exactly the conditions Dretske required for what he calls "channel conditions" (cf. def. 2.2.6):

> "There are an enormous number of conditions whose permanence, whose lack of any (relevant) possible alternative state, qualifies them as channel conditions. [...] Even though the signal depends on these relationships being as they are, in the sense that *if* they were different, or *if* they changed erratically, the signal would be equivocal, these conditions contribute nothing to the actual equivocation, since they generate no information for the signal to carry. Since they are the origin of no information, these conditions qualify as the framework *within which* communication takes place, not as a source *about which* communication takes place." (Dretske 1981: 116)

Exactly the same is true of the distinction between normal and non-normal situation. Though it is no explicitly accessible to the agent whose epistemic behaviour we are modelling, it is not completely out of his reach neither. It can *become* relevant, though at present it is not.

Another crucial property of normal worlds is that they are *local*, in the sense of answering not to an absolute distinction between what is possible and what is impossible from the modeller's perspective, but to the epistemic situation of the agent itself. They may be interpreted as embodying what Seligman called the agent's *perspective*:

> "... we wish to consider [...] a defense of the situation theoretic approach to modeling a mole-like conception of the world against two failings. The first is the priority of individuation over classification and the second is the globalist view of law-like dependencies. Our solution is to make the notions of a situation type and a constraint relative to a *perspective*. A perspective limits the applicability of a collection of constraints to a specific domain of situations over which they are ubiquitous." (Seligman 1990: 150–151)

Normal worlds give us a way to say when a channel is 'active', i.e. supporting actual information flow. The problem of specifying the conditions under which a channel supports information flow (and the associated problem of misinformation, i.e. reliance on a channel which is not supporting information flow), emerged in the nineties as one of the major problems in situation theory:

> "This abundance of channels may at first seem somewhat worrying, but all that should be concluded is that, potentially, there are dependencies between any two situations. Only when a channel is active will there be consequences of any substance. Clearly then, much of the work needed to make sense of an analysis of information flow in terms of channels will go in to saying what it is for a channel to be active." (Seligman 1991: 279)

The combination of channels with normal worlds also solves another related problem discussed in sct. 4.8.3, namely to specify the background conditions on which conditional constraints depend.

SPECIAL CONCEPTION OF IMPOSSIBLE BECAUSE defined in informational terms.

> "As I analyze things, impossibilities are those states of the system under investigation that are ruled out by (i.e., incompatible with) the currently available information about the system. States not so ruled out are possibilities." (Barwise 1997: 491)

7.2.2 Jumping to conclusions

Nonmonotonicity

van Benthem / Israel:

sf

Chapter 8

Representing knowledge by information frames

8.1 Information frames and their modalities

8.1.1 Information frames

Information frames are for information contexts what distributed systems are for classifications:

Definition 8.1.1. *An* informational modal framework *or* information frame $\mathcal{M}$ *consists of a classification A, information contexts $\mathcal{C}_s$ for every situation $s \in S$ such that indistinguishable situations make the same states possible:*

$$(1) \quad state(s) = state(s') \quad \Longrightarrow \quad \Gamma \vdash_{\mathcal{C}_s} \Delta \Leftrightarrow \Gamma \vdash_{\mathcal{C}'_s} \Delta$$

*The situations that are normal in $\mathcal{C}_s$ are called s-*normal.
*The states possible in $\mathcal{C}_s$ are called s-*possible.

Given that infinitary modal logic characterises worlds in Kripke classifications up to bisimulation, (1) can be seen to capture the essential idea behind using epistemic states to represent Kripke models.

(1) may be seen to capture an important presupposition of *locality*, endorsed e.g. also by Fagin et al. (1995):

"...if s and s' are states of the system such that agent i has the same local state [i.e. has access to the same information] in both of them, i.e. $s \sim_i s'$, and agent i knows ϕ in state s, then i also knows ϕ in state s'." (Fagin et al. 1995: 315)

8.1.2 Modalities

On information frames, we can define an accessibility relation:

Definition 8.1.2. *Let $\mathcal{M}$ be an information frame. ω' is* accessible *from ω ($\omega \leadsto \omega'$) if $\exists s \in S$ with $state(s) = \omega$ and ω' is s-possible.*

In section (7.1), we defined valuations of propositions in information contexts. We now define modal propositions and extend these definitions to include modal formulae:

Definition 8.1.3. *A* valuation *V of an information frame $\mathcal{M}$ is a function from a set of propositional constants $\mathbb{P}$ to $\mathcal{P}(\Omega_{\mathcal{M}})$. We add the following to cover the modal case:*

- $V(\Diamond\phi) := \{\omega \mid \exists\omega' \in \Sigma : \omega \leadsto \omega' \wedge \omega' \in V(\phi)\}$
- $\Box\phi := \neg\Diamond\neg\phi$

Truth of a proposition under a valuation in an information frame $\mathcal{M}$ is defined as follows:

$$s \models_{\mathcal{M}_V} \phi \quad :\Longleftrightarrow \quad s \models_{\mathcal{C}_{sV}} \phi \quad \forall s \in S$$

Validity of a proposition p under a valuation is truth in all situations, while validity tout court *in an information frame is validity under all valuations:*

$$\mathcal{M} \models_V \phi \quad :\Longleftrightarrow \quad s \models_{\mathcal{M}_V} \phi \quad \forall s \in S$$
$$\mathcal{M} \models \phi \quad :\Longleftrightarrow \quad \mathcal{M} \models_V p \quad \forall\, V : \mathbb{P} \to \mathcal{P}(\Omega_{\mathcal{M}})$$

That these modalities behave in the way we would expect is shown by the following theorem:

Theorem 8.1.4.

(2) $\quad s \models \Diamond p \quad \Longleftrightarrow \quad \exists\omega \;\; (\omega \text{ } s\text{-possible} \wedge \omega \in p)$

(3) $\quad s \models \Box p \quad \Longleftrightarrow \quad \forall\omega \;\; (\omega \text{ } s\text{-possible} \Rightarrow \omega \in p)$

Proof Recall that $state(s) \rightsquigarrow \omega \iff \omega$ is s-possible.
(2, [$\Longrightarrow$]:) If $s \models \Diamond p$, there is a state ω' such that $state(s) \rightsquigarrow \omega'$ and $\omega' \in p$. This means that there is a situation t with the same state as s such that ω' is t-possible and $\omega' \in p$. If $\omega' = \rangle \Gamma', \Delta' \langle$, it's being t-possible means $\Gamma' \nvdash_t \Delta'$. Because $state(s) = state(t)$, it follows that $\Gamma' \nvdash_s \Delta'$, which means that ω' is s-possible.

(2, [$\Longleftarrow$]:) If there is an s-possible state ω such that $\omega \in p$, $state(s) \rightsquigarrow \omega$. This is just the definition of $state(s) \in \Diamond p$, i.e. $s \models \Diamond p$.

(3, [$\Longrightarrow$]:) Suppose $s \models \Box p$. Then $s \not\models \Diamond \neg p$, i.e. $state(s) \notin \Diamond \neg p$. This means that for all states ω which are s-possible $\omega \in p$.

(3, [$\Longleftarrow$]:) Suppose $\omega \in p$ for all s-possible states ω. So there is no state ω such that $state(s) \rightsquigarrow \omega$ and $\omega \in p$, that is $state(s) \notin \omega$, i.e. $s \models \neg \Diamond \neg p$. $\square$

8.1.3 Neighbourhood semantics

In this section, I would like to present and discuss the Montague-Scott structures and the accompanying so-called "neighbourhood semantics" first introduced by Richard Montague (1968 1970) and Dana Scott (1970) and discussed by Segerberg (1971). I will follow the presentation in (Chellas 1980: ch. 7) and (Fagin et al. 1995: 316–321).

We start with the following definition:

Definition 8.1.5 (Montague-Scott models). *Let a set $\mathbb{P}$ of propositional variables and a set $\mathcal{A}$ of agents be given. A* Montague-Scott model *M is a triple $M = \langle W, \pi, \{\mathcal{C}_i\}_{i \in \mathcal{A}} \rangle$, where W is a set of worlds, $\pi : \mathbb{P} \to \mathcal{P}(W)$ a valuation of $\mathbb{P}$ and, for every $i \in \mathcal{A}$, $\mathcal{C}_i$ is a function $\mathcal{C}_i : W \to \mathcal{P}(\mathcal{P}(W))$, representing the set of propositions known to i in s. A* Montague-Scott world *is a pair of a Montague-Scott Model $M = \langle W, \pi, \{\mathcal{C}_i\}_{i \in \mathcal{A}} \rangle$ and some $w \in W$.*

We change our definition of truth in Kripke worlds (4.5.6) for the following:

Definition 8.1.6 (Truth in Montague-Scott worlds). *We define for a Montague-Scott* *world* (M,w) *and* $\phi\in\mathcal{L}_\infty$*:*

$$\begin{array}{rcl}
(M,w)\models\top & :\Longleftrightarrow & w\in M\\
(M,w)\models p & :\Longleftrightarrow & w\in\pi(p)\\
(M,w)\models\neg\phi & :\Longleftrightarrow & (M,w)\not\models\phi\\
(M,w)\models\phi\wedge\psi & :\Longleftrightarrow & (M,w)\models\phi \quad \textit{and} \quad (M,w)\models\psi\\
(M,w)\models\bigwedge\Phi & :\Longleftrightarrow & (M,w)\models\phi \quad \textit{for all } \phi\in\Phi\\
(M,w)\models[a]\phi & :\Longleftrightarrow & \{v\in W \mid (M,v)\models\phi\}\in\mathcal{C}_i(w)
\end{array}$$

Every Kripke model gives rise to a Montague-Scott model validating the same formulae:

Theorem 8.1.7. *Let* $M = \langle W,\pi,\{R_i\}_{i\in\mathcal{A}}\rangle$ *be a Kripke model and define a Montague-Scott model* $M' = \langle W',\pi',\{\mathcal{C}_i\}_{i\in\mathcal{A}}\rangle$ *by* $W' := W$*,* $\pi' := \pi$ *and* $\mathcal{C}_i(w) := \{V\subset W \mid \{v\in W \mid wR_iv\}\subset V\}$*. Then we have for all formulae* $\phi\in\mathcal{L}_\infty$ *and all* $w\in W$*:*

$$(M,w)\models_K\phi \quad\Longleftrightarrow\quad (M',w)\models_{MS}\phi$$

Proof [$\Longrightarrow$:] We prove the claim by induction on ϕ:

- Let $\phi\in\mathbb{P}$. Then $w\models_{\mathrm{K}}\phi\leftrightarrow w'\models_{\mathrm{MS}}\phi$ by the sameness of $\pi=\pi'$.
- Let $\phi=\neg\psi$. Then $w\models_{\mathrm{K}}\psi\leftrightarrow w'\models_{\mathrm{MS}}\psi$ by induction hypothesis and $w\models_{\mathrm{K}}\phi\leftrightarrow w'\models_{\mathrm{MS}}\phi$ by definition of $\neg$.
- Let $\phi=\bigwedge\Phi$. Then $w\models_{\mathrm{K}}\psi$ for all $\psi\in\Phi$ and thus $w'\models_{\mathrm{MS}}\psi$ for all $\psi\in\Phi$ by induction hypothesis. This implies $w'\models_{\mathrm{MS}}\phi$ by definition of $\bigwedge$.
- Let $\phi=[a]\psi$ for some $a\in\mathcal{A}$. We have to show that $\{v\in W \mid (M,v)\models_{\mathrm{MS}}\psi\}\in\mathcal{C}_i(w)$. By the definition of $\mathcal{C}_i(w)$, it is enough to show that $\{v\in W \mid wR_iv\}\cap\{v\in W \mid (M,v)\models_{\mathrm{MS}}\psi\}\in\mathcal{C}_i(w)$. By the induction hypothesis, however, we have $\{v\in W \mid wR_iv\}\cap\{v\in W \mid (M,v)\models_{\mathrm{MS}}\psi\}=\{v\in W \mid wR_iv\}$ and the latter is included in $\mathcal{C}_i(w)$ by definition.

[$\Longleftarrow$:] Again, we prove the claim by induction on ϕ. The first three clauses are the same as above. For the fourth, we have:

- Let $\phi=[a]\psi$ for some $a\in\mathcal{A}$. Let $v\in W$ be some world such that wR_iv. We have to show that $v\models_{\mathrm{K}}\psi$. By definition of $\models_{\mathrm{MS}}$, we know that $\{v\in W \mid (M,v)\models_{\mathrm{MS}}\psi\}\in\mathcal{C}_i(w)$. By the definition of $\mathcal{C}_i(w)$, we have $\{v\in W \mid wR_iv\}\subset\{v\in W \mid (M,v)\models_{\mathrm{MS}}\psi\}$.

By the induction hypothesis, we have $\{v \in W \mid (M,v) \models_{\mathrm{MS}} \psi\} \subset \{v \in W \mid (M,v) \models_{\mathrm{K}} \psi\}$. Chaining these together, we get $v \in \{v \in W \mid (M,v) \models_{\mathrm{K}} \psi\}$.

□

The class of all formulae valid in all Montague-Scott structures is axiomatised by the following (Fagin et al. (1995) and Chellas (1980):

Theorem 8.1.8. *The class of all Montague-Scott structures is sound and complete with respect to the system of modal logic having all propositional tautologies,*

Almost all forms of logical omniscience discussed on p. 16 in sct. 1.3.2 fail in neighbourhood semantics. In particular, the knowledge attributed to agents does not have to be closed under theoremhood (7 / Nec), nor under logical (9) nor material (10) implication, nor material equivalence (11), nor conjunction (12) nor simplification (13). This is because we do not restrict in any way the knowledge sets $\mathcal{C}_i(w)$ attributed to our agents. To have these kinds of omniscience, we have to impose further conditions on our Montague-Scott models:

Theorem 8.1.9. *Let $M = \langle W, \pi, \{\mathcal{C}_i\}_{i \in \mathcal{A}} \rangle$ be a Montague-Scott model and $\phi, \psi \in \mathcal{L}_\infty$ any formulae . Then the following hold for every $a \in \mathcal{A}$:*

$$\begin{array}{lrcl}
(4) & \models \phi \Rightarrow \models [a]\phi & \iff & \forall w \in W : W \in \mathcal{C}_a(w) \\
(5) & \models \phi \to \psi \Rightarrow \models [a]\phi \to [a]\psi & \iff & \forall w \in W : T \in \mathcal{C}_a(w) \wedge T \subset U \;\Rightarrow\; T \in \mathcal{C}_a(w) \\
(6) & \models [a](\phi \to \psi) \to ([a]\phi \to [a]\psi) & \iff & \forall w \in W : U, (W \setminus U \cup T) \in \mathcal{C}_a(w) \Rightarrow T \in \mathcal{C}_a(w) \\
(7) & \models [a](\phi \leftrightarrow \psi) \to ([a]\phi \leftrightarrow [a]\psi) & \iff & \forall w \in W : \mathcal{C}_a(w) \\
(8) & \models ([a]\phi \wedge [a]\psi) \to [a](\phi \wedge \psi) & \iff & \forall w \in W : T, U \in \mathcal{C}_a(w) \;\Rightarrow\; T \cap U \in \mathcal{C}_a(w) \\
(9) & \models [a](\phi \wedge \psi) \to ([a]\phi \wedge [a]\psi) & \iff & \forall w \in W : T \in \mathcal{C}_a(w) \wedge T \subset U \;\Rightarrow\; U \in \mathcal{C}_a(w)
\end{array}$$

Proof

[(4):]

[(5):]

[(6):]

[(7):]

[(8):]

[(9):] □

This means that $[a](\phi \wedge \psi) \leftrightarrow ([a]\phi \wedge [a]\psi)$ only holds in Montague-Scott models in which the neighbourhoods $\mathcal{C}_a(w)$ are upwards closed and closed under finite intersections. If, additionally, closure under theoremhood (Nec) is to valid, they have to be filters. If (D), i.e. $\models \neg \mathsf{K}_i(\bot)$ (for every $i \in \mathcal{A}$) is to be valid, the filters must be proper, i.e. not contain $\emptyset$.

Neighbourhood semantics, however, does validate closure under valid equivalence, i.e. the principle 9 we discussed on p. 16 in sct. 1.3.2:

If $\vdash \phi \leftrightarrow \psi$ then $\vdash \mathsf{K}\phi \leftrightarrow \mathsf{K}\psi$

Though this principle is the weakest of the three stronger principles of logical omniscience (closure under theoremhood, i.e. (Nec); closure under valid implication (8) and closure under logical equivalence (9)), closure under valid implication follows from it if we also have closure under simplification (13):

$\vdash \mathsf{K}(\phi \wedge \psi) \rightarrow (\mathsf{K}\phi \wedge \mathsf{K}\psi)$

Fagin et al. (1995) have tried to defend closure under valid equivalence (9) as follows:

> "All forms of logical omniscience fail except for closure under logical equivalence. In other words, while agents need not know all logical consequences of their knowledge, *they are unable to distinguish between logically equivalent formulas.* This is as much as we can expect to accomplish in a purely semantic model, since logically equivalent sentences are by definition semantically indistinguishable." (Fagin et al. 1995: 319)

This defense, however, misses the point, as Egré (2002: 11) has noted. For either (9) is to be understood as an assertion about the actual knowledge had by the agent: then, as noted in sct. 1.3.2, we are forced to misdescribe them. The fact that we cannot do better is irrelevant. The other possibility is that (9) is not to be understood as a descriptive assertion, but as a modelling constraint and thus situated at what Egré (2002), following Recanati (2000), calls the "metarepresentational level":

> "...the substitution of logically equivalent sentences [does not have to] ascribe exorbitant logical capacities to the agent, but reflects only the fact that the *description* of her knowledge is likely to be made up to logical equivalence." (Egré 2002: 11)

This means that *we*, the modellers, constrain our ascriptions by our acceptance of (9). Is there any way of keeping (9) without having (8), i.e. of justifying a denial of (13)? Paul Egré thinks there is and denies the step from "a knows that p and q" to "a knows that p and that q" on the basis that, for $q \equiv \neg p$, we seem to ascribe a contradictory belief only with the former. He substantiates his point by an analogy between "a knows that p" and

"a (simply, i.e. non-epistemically) sees that p". Even when every F is necessarily a G (and hence "$\forall x(Fx \leftrightarrow Fx \wedge Gx)$" is necessary) I may well see every F (and hence everything that is both F and G) without seeing all the Fs and all the Gs:

> "... in possible-world semantics, a proposition has the structure of a universal quantifier, and the distinction between 'that $(\phi \wedge \psi)$' and 'that $\phi \wedge$ that ψ' corresponds precisely to the distinction between the quantified expressions 'every AB' and 'every A and every B'." (Egré 2002: 17)

Though the parallel is instructive, it is not entirely adequate: while it is true that, in order to see the Fs and the Gs, I need perceptual contact only with those things that are both Fs and Gs, it cannot be said that, in order to know that p and q, I only need epistemic contact with those worlds that are both p- and q-worlds. For the epistemic contact is needed, in the latter case, not with what corresponds, in the terminology of Barwise and Perry (1983: 209) introduced on p. 81 in sct. 3.2.4, to visual *options* but to visual *alternatives*. Epistemic alternatives are what *is not excluded* by what one knows, while perceptually registered scenes are what one actively sees.[1] If this 'negative' definition of epistemic alternatives is taken into account, the picture changes: for an epistemic situation that does not exclude the intersection of the p- and the q-worlds is automatically an epistemic situation that does not exclude either of these sets.

But what about the plausibility of (13) on its own? I think the strongest case against (13) can be made on the grounds that it may be very difficult for the agent to 'extract' one conjunct out of a long conjunction. It seems very difficult, however, to spell out this notion of 'extraction' in more detail. Notice that denying (13) rules out a conjunctive analysis of conditional knowledge. Suppose that I good at establishing whether p – but only in situations in which q is true. So when, in a given situation, I acquire the justified true belief that p, my knowledge is analysed as $p \wedge q$, (13) fails for I do not have to know that q. As this analysis of conditional knowledge, as we saw in sct. 3.3 and 4.8.3, is not a good one anyway, this may be taken to be not too strong an argument against (13).

While it seems difficult to adjucate the relative credentials of (13) and (9), it seems clear that we do not want to have both. We already saw on p. 16 in sct. 1.3.2 that (9), given (13), implies closure under implication (8) and under theoremhood (7/Nec). First-order knowledge may provide us with another family of putative counter-examples to (13): I know, let us suppose, that every man is mortal. But there are many men of which I do not know anything (except tautologies, perhaps). Let Sam be one of these. So I do not

[1] This is why they are not, as we noted in sct. 4.4, in general transparent to the agent.

know of Sam that if he is a man, he is mortal, even though this is a conjunct in the long conjunction equivalent with the universally quantified assertion.

(12) is not very plausible in neighbourhood semantics. If it is valid and any non-empty neighbourhood contains the universal, i.e. necessary proposition (which is unique by (9)). This means that any agent who knows something knows knows every necessary proposition.

8.1.4 Properties of information frames

Information frames are said to be sound, complete and consistent iff their information contexts are:

Definition 8.1.10. *Let $\mathcal{M} = \langle A, \{C_s \; s \in S\}\rangle$ be an information frame:*

$$\begin{aligned} \mathcal{M} \textit{ is } \text{complete} &\iff \forall s \in S : C_s \textit{ is complete} \\ \mathcal{M} \textit{ is } \text{sound} &\iff \forall s \in S : C_s \textit{ is sound} \\ \mathcal{M} \textit{ is } \text{consistent} &\iff \forall s \in S : C_s \textit{ is consistent} \end{aligned}$$

The next two obvious theorems show that the set of formulae valid in an information frame is closed under substitutions and that our valuations are classical, i.e. that any propositional tautology is valid in any information frame:

Theorem 8.1.11. *For any information frame $\mathcal{M}$ and any modal formulas ϕ, ψ_1, ... ψ_n:*

$$\mathcal{M} \models \phi(p_1, ...p_n) \iff \mathcal{M} \models \phi(\psi_1, ...\psi_n)$$

Theorem 8.1.12. *For any information frame $\mathcal{M}$, any modal formula ϕ and any valuation V of ϕ:*

$$PC \vdash \phi \iff \forall \omega \in \Omega_{\mathcal{M}} : \omega \in V(\phi)$$

The next, important, theorem shows that information frames validate the **K**-axiom:

Theorem 8.1.13. *For any information frame $\mathcal{M}$ and all propositions p, q:*

$$\mathcal{M} \models \Box(p \to q) \to (\Box p \to \Box q)$$

Proof We have to show that $\forall s \in S$, $s \models \Diamond(p \wedge \neg q) \vee (\neg\Box p \vee \Box q)$.

Suppose $s \not\models \Diamond(p \wedge \neg q)$. So for all s-possible states ω, $\omega \notin p \vee \omega \in q$. If $s \not\models \neg\Box p$, then all s-possible ω are in p, so they all have to be in q and so $s \models \Box q$. If, on the other hand, $s \not\models \Box q$, there is an s-possible ω not in q. So this ω is not in p, i.e. $s \not\models \Box p$.

Suppose $s \models \Box p \wedge \neg\Box q$. By the second conjunct, there is an s-possible state ω which is not in q but which is in p. So $s \models \Diamond(p \wedge \neg q)$. □

(Real) means that every possible state is realized, while (Poss) means that every realized state is possible:

Definition 8.1.14. *Let $\mathcal{M} = \langle A, \{C_s \; s \in S\}\rangle$ be an information frame:*

$$\mathcal{M} \text{ is (Real)} \quad \Longleftrightarrow \quad \forall \omega = \langle \Gamma, \Delta \rangle \;:\; \Gamma \nvdash \Delta \longrightarrow \exists s \in S \; (\omega = state(s))$$
$$\mathcal{M} \text{ is (Poss)} \quad \Longleftrightarrow \quad \forall \omega = \langle \Gamma, \Delta \rangle \;:\; \Gamma \vdash \Delta \longrightarrow \forall s \in S \; (\Gamma \vdash_s \Delta)$$

While the **K** axiom is valid in any information frame, the rule of necessitation $p/\Box p$ is sound only in frames where every possible state is realized:

Theorem 8.1.15.

$$\mathcal{M} \text{ is (Real)} \quad \Longleftrightarrow \quad (\mathcal{M} \models p \Rightarrow \mathcal{M} \models \Box p)$$

Proof

[$\Longrightarrow$]: Assume $\mathcal{M} \models p$ and choose any $s \in S$. We have to show that $s \models \Box p$, i.e. that $\omega \in p$ for any s-possible state ω. By (Real), ω is realized, say by $t \in S$. By the validity of p, $t \models p$. Hence $\omega = state(t) \in p$ as desired.

[$\Longleftarrow$]: Let $\omega = \langle \Gamma, \Delta \rangle$ be an s-possible, but unrealized state. Define $p := \{\omega\}$. We know that $\forall s \in S$, $state(s) \neq \omega$, that is $state(s) \notin p$. So $s \not\models p$ but $s \models \neg p$. So we have $\mathcal{M} \models \neg p$. Assume, for reductio, $\mathcal{M} \models \Box\neg p$. Then $s \models \Box\neg p$, i.e. $\omega \in \neg p$ because ω is s-possible. But then $\omega \notin p$ by definition of $\neg$. Contradiction. □

Theorem 8.1.16. *For any information frame $\mathcal{M}$:*

$$\mathcal{M} \text{ is complete} \quad \Longrightarrow \quad \mathcal{M} \text{ is (Real)}$$

Proof If every possible state is realized by a normal situation, it is realized *tout court.* □

Theorem 8.1.17. *For any information frame $\mathcal{M}$:*

$$\mathcal{M} \text{ is sound and (Real)} \quad \Longrightarrow \quad \mathcal{M} \text{ is complete}$$

Proof Let $\mathcal{M}$ be sound, (Real), but not complete. Being not complete, not every possible state is realized by a normal situation. Because every situation is normal (soundness), not every possible state is realized *tout court*. $\square$

Theorem 8.1.18. *For any information frame $\mathcal{M}$:*

$$\mathcal{M} \textit{ is (Real)} \iff \forall s \in S \ (\Gamma \vdash_s \Delta) \Rightarrow \Gamma \vdash \Delta$$

Proof

[$\Longrightarrow$]: Let every possible state in $\mathcal{M}$ be realized. Suppose that $\Gamma \vdash_s \Delta$ holds $\forall s \in S$. This means it holds of all situations and thus of all states these situations realize. Because all possible states are realized, it even holds of all possible states. This is just $\Gamma \vdash \Delta$.

[$\Longleftarrow$]: Suppose there is an unrealized possible state $\omega = \langle \Gamma, \Delta \rangle$ in $\mathcal{M}$ and $\Gamma \vdash_s \Delta$ holds $\forall s \in S$. $\square$

Definition 8.1.19. *We will study the following properties of an information frame $\mathcal{M} = \langle A, \{C_s \ s \in S\} \rangle$:*

- *(Refl)* *Every situation s is s-possible, i.e. $\forall s \in S(\Gamma_s \nvdash_s \Delta_s)$.*
- *(Trans)* *The accessibility relation is transitive: $\omega \rightsquigarrow \omega' \wedge \omega' \rightsquigarrow \omega'' \Rightarrow \omega \rightsquigarrow \omega''$.*
- *(Symm)* *The accessibility relation is symmetrical: $\omega \rightsquigarrow \omega' \Rightarrow \omega' \rightsquigarrow \omega$.*
- *(Sep)* *$\mathcal{M}$ is separated, i.e. $\forall s, s' \in S \ \ (state(s) = state(s') \Rightarrow s = s')$.*
  In separated frames, situations can be identified with their descriptions or states.
- *(Ext)* *$\mathcal{M}$ is extensional, i.e. $\forall \sigma, \sigma' \in \Sigma \ \ (\mathsf{tok}(\sigma) = \mathsf{tok}(\sigma') \Rightarrow \sigma = \sigma')$.*
  In extensional frames, situation types can be identified with their extensions.
- *(Boole)* *Σ is a Boolean algebra.*

Theorem 8.1.20. *For any information frame $\mathcal{M}$ which is (Sep):*

$$\textit{(Sound)} \iff (\Gamma \vdash_N \Delta \Rightarrow \forall s \in S \ \ (\Gamma \vdash_s \Delta).$$

Theorem 8.1.21. *For any information frame $\mathcal{M}$ which is (Sep):*

$$\textit{(Sound)} \wedge \textit{(Real)} \iff (\Gamma \vdash \Delta \Leftrightarrow \forall s \in S \ \ (\Gamma \vdash_s \Delta)).$$

Theorem 8.1.22. *For any information frame $\mathcal{M}$:*

$$\mathcal{M} \textit{ is (Cons)} \iff \mathcal{M} \models \Box p \rightarrow \Diamond p.$$

Theorem 8.1.23. $(\mathsf{typ}(.), \mathsf{tok}(.))$ *is a Galois relation between* S *and* Σ*, which means that for any* $G, G' \subset S$ *and any* $\Gamma, \Gamma' \subset \Sigma$*:*

- $G \subset G' \Rightarrow \mathsf{typ}(G) \supset \mathsf{typ}(G')$
  $\Gamma \subset \Gamma' \Rightarrow \mathsf{tok}(\Gamma) \supset \mathsf{tok}(\Gamma')$
- $G \subset \mathsf{tok}(\mathsf{typ}(G))$
  $\Gamma \subset \mathsf{typ}(\mathsf{tok}(\Gamma))$

Theorem 8.1.24. *For any information frame* $\mathcal{M}$*:*

$$\mathcal{M} \textit{ is (Boole)} \iff \forall G \subset S\ \exists \sigma \in \Sigma \quad (\mathsf{typ}(G) = \mathsf{typ}(\mathsf{tok}(\sigma)))$$

Theorem 8.1.25. *For any information frame* $\mathcal{M}$*:*

$$\mathcal{M} \textit{ is (Boole)} \wedge \textit{(Sep)} \Longrightarrow \forall\ G, G' \subset S \quad (\mathsf{typ}(G) = \mathsf{typ}(G') \Rightarrow G = G')$$

Theorem 8.1.26. *For any information frame* $\mathcal{M}$*:*

$$\mathcal{M} \textit{ is (Boole)} \wedge \textit{(Sep)} \iff \forall G \subset S \quad (G = \mathsf{toktyp}(G))$$

Theorem 8.1.27. *For any information frame* $\mathcal{M}$*:*

$$\mathcal{M} \textit{ is (Boole)} \wedge \textit{(Sep)} \iff \forall G \subset S\ \exists \sigma \in \Sigma \quad (\mathsf{tok}(\sigma) = G)$$

Theorem 8.1.28. *Any information frame* $\mathcal{M}$ *which is (Ext) is isomorphic to an information frame of sets* $\mathcal{M}^*$*.*
If $\mathcal{M} = \langle A = \langle S, \Sigma, \models \rangle, \{A, \vdash_s, N_s \mid s \in S\} \rangle$ *is (Ext), then it is isomorphic to the information frame:*

$$\mathcal{M}^* = \langle A^* = \langle S, \Sigma^*, \in \rangle, \{A^*, \vdash_s^*, N_s \mid s \in S\} \rangle \quad \textit{with}$$

$$\Sigma^* := \{\mathsf{tok}(\sigma) \mid \sigma \in \Sigma\} \subset \mathcal{P}(S) \quad \textit{and}$$

$$\Gamma^* \vdash_s^* \Delta^* \ \Leftrightarrow\ \Gamma \vdash_s \Delta$$

8.2 Epistemic interpretations of modalities on information frames

Deal with omniscience

Veracity requirement (defended in sct. 1.4.5) lack of recognised on p. 208

We now have to answer the obvious question whether and in what way our modalities can be given an epistemic interpretation. Barwise saves (1.4.1) by making the very notion of possibility relative to an informational state: given a state characterised by information Γ, some further informational state Δ is possible iff $\Gamma \cup \Delta$ is consistent.

We noted in sct. 3.2.4 that "knows that ϕ" and "does not know that ϕ" are best construed as contraries, not contradictories. EPISTEMIC COMMAND NEEDED; ATTUNEMENT

8.2.1 Lewis frames

Definition 8.2.1. *A modal formula A is* valid under a valuation V *in an information frame* $\mathcal{M}$ $(\mathcal{M} \models_V A)(\mathcal{M} \models_V A)$ *if it holds in all situations under* V*, i.e. if* $\forall s \in S\ (state(s) \in V(A))$*. A modal formula A is* valid *in* $\mathcal{M}$ $(\mathcal{M} \models A)$ *if it is valid under any valuation.*

Definition 8.2.2. *The* Lewis information frame $\mathcal{M}_{\mathcal{F}}$ associated with a Kripke frame $\mathcal{F} = \langle W, R\rangle$ *consists of the Lewis classification A and Lewis information contexts* $\mathcal{C}_w$ *for every possible world* $w \in W$.

Definition 8.2.3. *The* canonical Lewis frame associated with a Kripke frame *is the Lewis information frame with* $wRw' \Leftrightarrow w' \in N_w$ *for any Lewis information context* $\mathcal{C}_w$.

Theorem 8.2.4. *Any Lewis information frame is (Real), (Sep), (Ext) and (Boole). In addition, canonical Lewis information frames are (Compl).*

8.2.2 Kripke frames

Definition 8.2.5. *The* Kripke frame $\mathcal{F}_{\mathcal{M}} = \langle W, R\rangle$ associated with a extensional information frame $\mathcal{M} = \langle A = \langle S, \Sigma \models\rangle, \{\langle A, \vdash_s, N_s\rangle \mid s \in S\}\rangle$ *is the Kripke frame with* $W = S$ *and* $aRb \Leftrightarrow b$ *is a-possible.*

Theorem 8.2.6. *For a given Kripke model* $\langle W, R, V\rangle$*, we define a valuation* $V_{\mathcal{F}}$ *of propositional constants over* $\mathcal{M}_{\mathcal{F}}$ *as follows:*

$$V_{\mathcal{F}}(p) := \{state(s) \mid s \in W, \langle W, R, V, s\rangle \models p\}$$

Then we have for any modal formula A:

$$\langle W, R, V, s\rangle \models A \iff state(s) \in V_{\mathcal{F}}(A)$$

Theorem 8.2.7. *For a given information frame $\mathcal{M}$ which is (Ext) and (Real), i.e. $\mathcal{M} = \langle\langle S, \Sigma^*, \in\rangle, \{A^*, \vdash_s^*, N_s \mid s \in S\}\rangle$ with $\Sigma^* \subset \mathcal{P}(S)$ and a given valuation V, we define a valuation $V_{\mathcal{M}}$ of propositional constants over $\mathcal{F}_{\mathcal{M}}$ as follows:*

$$V_{\mathcal{M}}(p) := \{s \mid state(s) \in V(p)\}$$

Then we have for any modal formula A:

$$\langle \mathcal{M}, V, s\rangle \models A \iff s \in V_{\mathcal{M}}(A)$$

8.2.3 Information frames as models of modal logics

Theorem 8.2.8. *For any modal formula A and any information frame $\mathcal{M}$:*

$$(\mathcal{M} \textit{ is (Real)} \Rightarrow \mathcal{M} \models A\ ) \iff K \vdash A$$

Theorem 8.2.9. *For any modal formula A and any information frame $\mathcal{M}$:*

$$(\mathcal{M} \textit{ is (Real)} \wedge \textit{(Refl)} \Rightarrow \mathcal{M} \models A\ ) \iff T \vdash A$$

Theorem 8.2.10. *For any modal formula A and any information frame $\mathcal{M}$:*

$$(\mathcal{M} \textit{ is (Real)} \wedge \textit{(Trans)} \Rightarrow \mathcal{M} \models A\ ) \iff K4 \vdash A$$

Theorem 8.2.11. *For any modal formula A and any information frame $\mathcal{M}$:*

$$(\mathcal{M} \textit{ is (Real)} \wedge \textit{(Refl)} \wedge \textit{(Trans)} \Rightarrow \mathcal{M} \models A\ ) \iff S4 \vdash A$$

Theorem 8.2.12. *For any modal formula A and any information frame $\mathcal{M}$:*

$$(\ M \textit{ is (Real)} \wedge \textit{(Refl)} \wedge \textit{(Trans)} \wedge \textit{(Symm)} \Rightarrow \mathcal{M} \models A\ ) \iff S5 \vdash A$$

Theorem 8.2.13. *For any modal formula A and any information frame $\mathcal{M}$:*

$$(\mathcal{M} \textit{ is (Real)} \wedge \textit{(Cons)} \Rightarrow \mathcal{M} \models A\ ) \iff D \vdash A$$

8.3 Information Flow

WHAT I HAVE IN THE CONCLUSION Information frames allow for the introduction of agent-relative epistemic modalities not subject to most of the problems plaguing epistemic modal logic. Information frames also afford a plausible picture of informational alternatives, bringing them back within the cognitive reach of the agents modelled.

We noticed in (4.8.3) the importance of ruling out, as not supporting information flow, accidental regularities. Indeed, this was one of the major drawbacks of Dretske's reliance on communication theory in his definition of knowledge (def. 2.2.5). The theory presented in this chapter presents a solution to this problem.

The basic idea is to reverse the notion of explanation. Instead of taking the system under consideration to be a vast assemblage of individually isolated events or situations, it is characteristic for the situation-theoretic approach as described in ch. 3 to take constraints and informational dependencies (channels) to be situations themselves, i.e. as something an adequate theory may take as given and try to *describe*, rather than to reconstruct it. This approach is fore-shadowed in the following remark of Jerry Seligman:

> "Our approach to the problem [of defining channels as necessary co-occurrence of situations, thereby making channels with impossible antecedents automatically 'active'] is to reverse the direction of explanation. Instead of trying to explain the law-likeness of information flow along channels in terms of a global notion of possibility, we will reconstruct possibilities from the notion of an active channel. Instead of supposing that certain worlds are, as a matter of fact, possible whilst others are not, [we] will suppose that, in certain situations, certain channels are active whilst in other situations they are not." (Seligman 1991: 281)

As indicated in sct. 7.2.2, information flow may provide us with the adequate tools to deal with the problem of epistemic dynamics (cf. sct. 4.8.4).

Not only is dynamics of belief, the logic of belief change and accommodation of new information, of intrinsic interest, it may be seen to be crucial even for the more modest goal of modelling static belief and knowledge states in two respects. It may be plausibly argued, first, that even an adequate characterisation even of the static doxastic and epistemic state of an agent must include his *dispositions* to change his beliefs or retreat his claims to knowledge. This may be seen as accounting for the efficiency of the mental (cf. p. 70 in sct. 3.1):

> "Knowing the doxastic state of an agent requires more than knowing his or her current belief set, it also requires knowing how he or she would respond to new information about the world. [...] A belief state is a "belief set cum dispositions for belief change"." (Gochet and Gribomont 2002: 27)

Second, knowledge and belief differ in the dispositions they rationalise and this difference may be plausible taken to characterise them.[2]

[2] We noted the difficulty of developing a joint framework for knowledge and belief in sct. 1.3.7. I think that something like this was the motivation behind Hintikka's use of "defensibility" mark of knowledge, on which – as we saw in sct. 1.3.4 – he relied in his defence of (S4).

Chapter 9

Conclusion

What, now, are the main merits of Barwise's and Seligman's theory of information flow? Have we found *the* (or, at least, *a*) logic of information?

I do not think so. Even though the theory, as we saw, has some clear advantages over other work done in the field, it is both too general and not worked out enough to deserve this honorific title. It is a start, and probably – as I tried to argue – a start in the right direction. Much work has too be done and the theory may still – given that it is unlikely that many people will work on it – go the way of oblivion.

The main advantages, as I see them, of the theory is that it is a theory of information and not a theory of knowledge or meaning. I hope to have shown that – even though the concept is elusive and there is not much to be hoped from a 'philosophy' of information – information is a legitimate and interesting subject matter for people hitherto concerned with epistemic logic.

More specifically, I find the following properties of Barwise's and Seligman's theory attractive:

- The basic elements of the theory are situations, not models and possible worlds.
- Set theory is treated just as a special kind of classification, not as all-embracing meta-theory.
- The basic ontology is vehicle is that of a classification, and not of a (possible) state of the world, which allows for perspectivism and a flexible way to deal with indistinguishable situations.
- The information-carrying items are conceived of as types, properties of situations, instead of propositions; the theory is therefore much freer in the compounding of such items.

- The distinction between channels and theories explains how information is both in the world and about the world.
- Information contexts give us a context-dependent notion of normal situations and thus solve the problem of conditional constraints.
- Information frames allow for the introduction of agent-relative epistemic modalities not subject to most of the problems plaguing epistemic modal logic.
- Information frames also afford a plausible picture of informational alternatives, bringing them back within the cognitive reach of the agents modelled.

If only some of these advantages would be preserved by some of the theories succeeding to the one of Barwise and Seligman, we would have a clear case of scientific progress.

Bibliography

Anderson, Anthony C., 2001. "The Lesson of Kaplan's Paradox about Possible World Semantics". Talk given at the 2001 Kaplan conference in San Marino.

Bar-Hillel, Yoshua and Carnap, Rudolf, 1953. "Outline of a Theory of Semantic Information". In: Yoshua Bar-Hillel, editor, "Language and Information", London: Addison Wesley.

Barwise, Jon, 1981. "Scenes and other Situations". *Journal of Philosophy* 78:369–397. Reprinted in (Barwise 1989d: 5-33).

Barwise, Jon, 1986a. "Conditionals and Conditional Information". In: Traugott, editor, "On Conditionals", Cambridge: Cambridge University Press. Reprinted in (Barwise 1989d: 97–136).

Barwise, Jon, 1986b. "Information and Circumstance: A reply to Fodor". *Notre Dame Journal of Formal Logic* 27:324–338. Reprinted in (Barwise 1989d: 137–154).

Barwise, Jon, 1986c. "Logic and Information". In: Ruth Barcan Marcus, editor, "Logic, Methodology and Philosophy of Science VII", Dordrecht: Elsevier. Reprinted in (Barwise 1989d: 37–58).

Barwise, Jon, 1986d. "Review of Stalnaker, Inquiry". *Philosophical Review* 45:429–434.

Barwise, Jon, 1986e. *Situations, Sets, and the Axiom of Foundation.* Technical Report, CSLI Stanford. Reprinted in (Barwise 1989d: 177–200).

Barwise, Jon, 1987. *Situations and Small Worlds*. Technical Report, CSLI Stanford. Reprinted in (Barwise 1989d: 79–97) and as "Situationen und kleine Welten" in von Stechow and Wunderlich (1991).

Barwise, Jon, 1988. *On the Model Theory of Common Knowledge*. Technical Report, CSLI Stanford. Reprinted in (Barwise 1989d: 201–220).

Barwise, Jon, 1989a. “Notes on Branch Points in Situation Theory”. In: Barwise (1989d), pages 255–276.

Barwise, Jon, 1989b. “Situated Set Theory”. In: Barwise (1989d), pages 289–292.

Barwise, Jon, 1989c. “Situations, Facts, and True Propositions”. In: Barwise (1989d), pages 221–254.

Barwise, Jon, 1989d. *The Situation in Logic*. Stanford: CSLI.

Barwise, Jon, 1989e. “Toward a Mathematical Theory of Meaning”. In: Barwise (1989d), pages 293–298.

Barwise, Jon, 1993. “Constraints, Channels, and the Flow of Information”. In: Peter Aczel, David Israel, Katagiri, and Peters, editors, “Situation Theory and its Applications vol.3”, Stanford: CSLI.

Barwise, Jon, 1997. “Information and Impossibilities”. *Notre Dame Journal of Formal Logic* 38:488–515.

Barwise, Jon, Gavron, Jean Mark, Plotkin, Gordon, and Tutiya, Syun, editors, 1991. *Situation Theory and its Applications vol.2*. Stanford: CSLI.

Barwise, Jon and Moss, Lawrence S., 1996. *Vicious Circles – On the Mathematics of Non-Wellfounded Phenomena*. Stanford: CSLI.

Barwise, Jon and Perry, John, 1981a. “Semantic Innocence and Uncompromising Situations”. In: French, Uehling, and Wettstein, editors, “Midwest Studies in Philosophy 6”, Minneapolis: University of Minnesota Press.

Barwise, Jon and Perry, John, 1981b. “Situations and Attitudes”. *Journal of Philosophy* 78:668–691.

Barwise, Jon and Perry, John, 1983. *Situations and Attitudes*. Cambridge MA: MIT Press, repr. Stanford: CSLI Publications 1999 edition. Reissued 1999 in the *David Hume Series of Philosophy and Cognitive Science Reissues* series of CSLI Press, Stanford.

Barwise, Jon and Perry, John, 1985. “Shifting Situations and Shaken Attitudes”. *Linguistics and Philosophy* 8:105–161. Reprinted in (Barwise and Perry 1983: xxxiii-lxxxviii), CSLI reedition 1999.

Barwise, Jon and Perry, John, 1999. "Introduction to the Reissue Edition". In: Barwise and Perry (1983). Reissued 1999 in the *David Hume Series of Philosophy and Cognitive Science Reissues* series of CSLI Press, Stanford.

Barwise, Jon and Seligman, Jerry, 1997. *Information Flow: The Logic of Distributed Systems*. (Cambridge Tracts in Theoretical Computer Science, 44). Cambridge: Cambridge University Press.

Benthem, Johan van, 1991. *Language in Action: Categories, Lambdas, and Dynamic Logic*. Cambridge MA: MIT Press.

Blackburn, Patrick, Rijke, Maarten de, and Venema, Yde, 2001. *Modal Logic*. Cambridge: Cambridge University Press.

Bynum, Terrell Ward and Moor, James H., 1998. *The Digital Phoenix – How Computers are Changing Philosophy*. Oxford: Blackwell.

Carnap, Rudolf, 1942. *Introduction to Semantics*. Cambridge MA: Harvard University Press.

Castañeda, Hector-Neri, 1964. "Review of Hintikka, Knowledge and Belief". *Journal of Symbolic Logic* 29:132–134.

Chellas, Brian F., 1980. *Modal Logic: an Introduction*. Cambridge: Cambridge University Press.

Church, Alonzo, 1943. "Carnap's introduction to Semantics". *Philosophical Review* 52:298–304.

Church, Alonzo, 1950. "On Carnap's Analysis of Statements of Assertion and Belief". *Analysis* 10:97–99.

Cooper, Robin and Kamp, Hans, 1991. "Negation in Situation Semantics and Discourse Representation Theory". In: Barwise et al. (1991).

Cooper, Robin, Mukai, Kuniaki, and Perry, John, editors, 1990. *Situation Theory and its Applications vol.1*. Stanford: CSLI.

Davey, B.A. and Priestley, H.A., 1990. *Introduction to Lattices and Order*. Cambridge: Cambridge University Press.

Davidson, Donald, 1968. “On Saying That”. *Synthese* 19:130–146. Reprinted in Davidson (1984).

Davidson, Donald, 1984. *Truth and Interpretation*. Oxford: Clarendon Press.

Degen, Wolfgang, Heller, Barbara, and Herre, Heinrich, 2002. “GOL: A Framework for Building and Representing Ontologies”. In: Heller et al. (2002). A reader for the introductory workshop in ontology, Naumburg April 17-20, 2002.

Devlin, Keith, 1990. “Infons and Types in an Information-Based Logic”. In: Cooper et al. (1990).

Devlin, Keith, 1991. *Logic and Information*. Cambridge: Cambridge University Press.

Devlin, Keith, 2001. *InfoSense: Turning Information into Knowledge*. New York: W.H. Freeman.

Donnellan, Keith S., 1966. “Reference and Definite Description”. *Philosophical Review* 75. Reprinted in Steinberg and Jakobovits (1971).

Donnellan, Keith S., 1968. “Putting Humpty Dumpty Together Again”. *Philosophical Review* 77:203–215.

Donnellan, Keith S., 1979. “Speaker Reference, Descriptions and Anaphora”. In: Peter Cole, editor, “Syntax and Semantics: Pragmatics (Vol. 9)”, New York: ???? Reprinted in French et al. (1978).

Dretske, Fred, 1970. “Epistemic Operators”. *Journal of Philosophy* 67:1007–1023. Reprinted in Dretske (2000).

Dretske, Fred, 1971. “Conclusive Reasons”. *Australasian Journal of Philosophy* 49:1–22. Reprinted in Dretske (2000).

Dretske, Fred, 1981. *Knowledge and the Flow of Information*. Cambridge MA: MIT Press. Reissued 1999 in the *David Hume Series of Philosophy and Cognitive Science Reissues* series of CSLI Press, Stanford.

Dretske, Fred, 1983. “Précis of Knowledge and the Flow of Information””. *Behavioral and Brain Sciences* 6:55–63. Reprinted in Kornblith (1985).

Dretske, Fred, 1988. *Explaining Behaviour: Reasons in a World of Causes*. Cambridge MA: MIT Press.

Dretske, Fred, 1989. *Seeing and Knowing*. London: Routledge.

Dretske, Fred, 1990. "Putting Information to Work". In: Hanson (1990). Reprinted in revised form in Dretske (2000).

Dretske, Fred, 2000. *Perception, Knowledge and Belief: Selected Essays*. Cambridge: Cambridge University Press.

Dretske, Fred, 2001a. "Externalism and Self-Knowledge". Draft, distributed at 2001 summer school in Heidelberg, Germany.

Dretske, Fred, 2001b. "How Do You Know You Are Not A Zombie?" Draft, distributed at 2001 summer school in Heidelberg, Germany.

Dubucs, Jacques, 1991. "On logical omniscience". *Logique et Analyse* 133–134:41–55.

Duc, Ho Ngoc, 2001. *Resource-bounded reasoning about knowledge*. Ph.D. thesis, Faculty of Mathematics and Informatics, University of Leipzig.

Egré, Paul, 2002. "A pragmatic approach to the problem of logical omniscience". Unpublished draft available from `l.freches@magic.fr`.

Fagin, Ronald and Halpern, Joseph Y., 1992. "Two views of belief: belief as generalized probability and belief as evidence". *Artificial Intelligence* 54:275–317.

Fagin, Ronald, Halpern, J.Y., Moses, Yoram, and Vardi, Moshe Y., 1995. *Reasoning About Knowledge*. Cambridge MA: MIT Press.

Floridi, Luciano, 2002a. "Introduction: What is the Philosophy of Information?" In: Floridi (2002c).

Floridi, Luciano, 2002b. "On the Intrinsic Value of Information Objects and the Infosphere". `http://www.wolfson.ox.ac.uk/~floridi/pdf/oivioi.pdf`.

Floridi, Luciano, editor, 2002c. *The Blackwell Guide to the Philosophy of Computing and Information*. Oxford: Blackwell.

Floridi, Luciano, 2003. "Outline of a Theory of Strongly Semantic Information". *Minds and Machines* 13.

Foley, Richard, 1987. "Dretske's Information-Theoretic Account of Knowledge". *Synthese* 70:159–184.

Føllesdal, Dagfinn, 1983. "Situation Semantics and the 'Slingshot' Argument". *Erkenntnis* 19:91–98.

Fox, C.J., 1983. *Information and Misinformation – An Investigation of the Notions of Information, Misinformation, Informing, and Misinforming*. Westport CO: Greenwood Press.

French, Uehling, and Wettstein, editors, 1978. *Midwest Studies in Philosophy 3*. Minneapolis: University of Minnesota Press.

Galgemi, Aldo, Pisanelli, Domenico M., and Steve, Geri, 2002. "An Overview of the ONIONS Project: Applying Ontologies to the Integration of Medical Terminologies". In: Heller et al. (2002). A reader for the introductory workshop in ontology, Naumburg April 17-20, 2002.

Gettier, Edmund, 1963. "Is Justified True Belief Knowledge?" *Analysis* 23:121–123.

Gochet, Paul and Gribomont, P., 2002. "Epistemic Logic". Draft available from `pgochet@ulg.ac.be`, to be published in **?**.

Goldman, Alvin I., 1967. "A Causal Theory of Knowing". *Journal of Philosophy* 64:357–372.

Goodman, Nelson, 1961. "About". *Mind* 70:1–24.

Grice, Paul, 1948. "Meaning". *Philosophical Review* 66. Reprinted in Grice (1989b).

Grice, Paul, 1969. "Utterer's Meaning and Intentions". *Philosophical Review* 78. Reprinted in Grice (1989b).

Grice, Paul, 1989a. "Retrospective Epilogue". In: Grice (1989b).

Grice, Paul, 1989b. *Studies in the Way of Words*. Cambridge MA: Harvard University Press.

Hair, S.G, 1969. "A Definition of Informational Content". *Journal of Philosophy* 66:132–133.

Hanson, Philip P., editor, 1990. *Information, Language and Cognition*. Vancouver Studies in Cognitive Science, Vol. I. Vancouver: University of British Columbia Press.

Heller, Barbara, Herre, Heinrich, and Smith, Barry, editors, 2002. *Ontological Spring*. IFOMIS. A reader for the introductory workshop in ontology, Naumburg April 17-20, 2002.

Hintikka, Jaakko, 1962. *Knowledge and Belief*. Ithaca NY: Cornell University Press.

Hintikka, Jaakko, 1968. “Epistemic Logic and the Methods of Philosophical Analysis”. *Australasian Journal of Philosophy* 46:37–51. Reprinted in Hintikka (1969).

Hintikka, Jaakko, 1969. *Models for Modalities*. Dordrecht: Reidel.

Hintikka, Jaakko, 1970. “Knowledge, Belief, and Logical Consequence”. *Ajatus* 32:32–47. Reprinted in revised form in (Hintikka 1975: ch.∼9).

Hintikka, Jaakko, 1975. *The Intentions of Intentionality and Other New Models for Modalities*. Dordrecht: Reidel.

Hintikka, Jaakko, 1983. “Situations, Possible Worlds and Attitudes”. *Synthese* 54:153–162. Reprinted in Hintikka and Hintikka (1989).

Hintikka, Jaakko and Hintikka, Merill B., 1989. *The Logic of Epistemology and The Epistemology of Logic*. Dordrecht: Reidel.

Hoek, Wiebe van der, 1993. “Systems for Knowledge and Belief”. *Journal of Logic and Computation* 3:173–198.

Israel, David, 1989. “Concepts of information: Comparative axiomatics”. In: Richmond Thomason, editor, “Philosophical Logic and Artificial Intelligence”, pages 35–72. Dordrecht: Kluwer.

Israel, David and Perry, John, 1990. “What is Information?” In: Hanson (1990).

Israel, David and Perry, John, 1991. “Information and Architecture”. In: Barwise et al. (1991).

Kakuda, Yuzuru and Kikuchi, Makoto, 2001. “Topology on Classifications in Abstract Design Theory”. `http://kurt.scitec.kobe-u.ac.jp/∼kikuchi/papers/tcadt.pdf` .

Kaplan, David, 1989. “Afterthoughts”. In: Almog, Perry, and Wettstein, editors, “Themes from Kaplan”, Oxford: Oxford University Press.

Kaplan, David, 1994. "A Problem in Possible-World Semantics". In: Sinnott-Armstrong, Raffman, and Asher, editors, "Modality, Morality, and Belief: Essays in Honor Ruth Barcan Marcus", Cambridge: Cambridge University Press.

Kent, Robert E., 2001. "A KIF Formalization for the IFF Category Theory Ontology". `citeseer.nj.nec.com/kent01kif.html`.

Kornblith, Hilary, editor, 1985. *Naturalizing Epistemology*. Cambridge: Cambridge University Press.

Kripke, Saul A., 1972. "Naming and Necessity". In: Donald Davidson and Gilbert Harman, editors, "Semantics of Natural Language", pages 253–355. Dordrecht: Reidel. Reprinted as Kripke (1980).

Kripke, Saul A., 1980. *Naming and Necessity*. Oxford: Blackwell.

Krivine, Jean-Louis, 1971. *Introduction to Axiomatic Set Theory*. Synthese Library. Dordrecht: Reidel. English translation of "Théorie axiomatique des ensembles".

Lemmon, E.J., 1965. "Review of Hintikka, Knowledge and Belief". *Philosophical Review* 74:381–384.

Lemon, Oliver, 1998. "Review of Barwise/Seligman, Information Flow". *Erkenntnis* 49:397–401.

Lenzen, Wolfgang, 1978. "Recent work in epistemic logic". *Acta Philosophica Fennica* 30.

Lenzen, Wolfgang, 1979. *Glauben, Wissen und Wahrscheinlichkeit*. Wien: Springer.

Lewis, David, 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

Lewis, David, 1996. "Elusive Knowledge". *Australasian Journal of Philosophy* 74:549–567. Reprinted in Lewis (1999).

Lewis, David, 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lindström, Sten, 1991. "Critical Study of Barwise/Perry, Situations and Attitudes". *Noûs* 25:743–770.

Maloney, Christopher, 1983. "Dretske on Knowledge and Information". *Analysis* 43:25–28.

Martin, Richard M., 1962a. "On Knowing, Believing, Thinking". *Journal of Philosophy* 59:586–600.

Martin, Richard M., 1962b. "Toward an Extensional Logic of Belief". *Journal of Philosophy* 59:169–172.

Milner, Robin, 1990. "Operational and Algebraic Semantics of Concurrent Processes". In: J. van Leeuwen, editor, "Handbook of Theoretical Computer Science vol. B.", pages 1203–1242. Dordrecht: Elsevier.

Montague, Richard, 1968. "Pragmatics". In: Jaakko Hintikka, Julius M. Moravcsik, Patrick Suppes, and Klinabsky, editors, "Contemporary Philosophy: A Survey", Florence: La Nuova Italia Editrice. Reprinted in Thomason (1974).

Montague, Richard, 1970. "English as a Formal Language". In: Bruno et al Vistenti, editor, "Linguaggi nella Societa e nella Tecnica", Milan: Edizioni di Comunita. Reprinted in Thomason (1974).

Morris, William Edward, 1990. "Knowledge and the Regularity Theory of Information". *Synthese* 82:375–398.

Nozick, Robert, 1981. *Philosophical Explanations*. Cambridge MA: Harvard University Press.

Perry, John, 1979. "The Problem of the Essential Indexical". *Noûs* 13:3–21. Reprinted in (Perry 2000: 27–44).

Perry, John, 2000. *The Problem of the Essential Indexical and Other Essays*. Stanford: CSLI. Expanded Edition.

Perry, John, 2001a. *Knowledge, Possibility, and Consciousness*. Cambridge MA: MIT Press.

Perry, John, 2001b. *Reference and Reflexivity*. Stanford: CSLI.

Popper, Karl Raimund, 1935. *Logik der Forschung*. Tubingen: J.C.B. Mohr. Translated as "Logic of Scientific Discovery" and published by Basic Books, New York, in 1959.

Quine, Willard van Ornam, 1950. *Methods of Logic*. Cambridge MA: Harvard University Press.

Quine, Willard van Ornam, 1960. *Word and Object*. Cambridge MA: MIT Press.

Recanati, François, 2000. *Oratio Obliqua, Oratio Recta: An Essay on Metarepresentation*. Cambridge MA: MIT Press.

Ryle, Gilbert, 1954. "Formal and Informal Logic". In: "Dilemmas", pages 111–129. Cambridge: Cambridge University Press. The Tarner Lectures 1953.

Savitt, Steve, 1987. "Absolute Informational Content". *Synthese* 70:185–190.

Scott, Dana, 1970. "Advice on modal logic". In: Karel Lambert, editor, "Philosophical Problems in Logic", pages 143–173. Dordrecht: Reidel, 2nd ed. 1980 edition.

Segerberg, Krister, 1971. *An Essay in Classical Modal Logic*. Philosophical Studies, published by the Philosophical Society and the Departement of Philosophy, No. 13, University of Uppsala. Uppsala: Filosofiska Studier.

Seligman, Jerry, 1990. "Perspectives in Situation Theory". In: Cooper et al. (1990), pages 147–192.

Seligman, Jerry, 1991. "Physical Situations and Information Flow". In: Barwise et al. (1991), pages 257–292.

Seligman, Jerry and Moss, Lawrence S., 1997. "Situation Theory". In: Johan van Benthem and Alice ter Meulen, editors, "Handbook of Logic and Language", Elsevier.

Shannon, Claude E., 1948. "A Mathematical Theory of Communication". *Bell System Technical Journal* Reprinted in Shannon and Weaver (1949).

Shannon, Claude E. and Weaver, Warren, 1949. *Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Shoham, Yoav and Moses, Yoram, 1989. "Belief as Defeasible Knowledge". *Journal of Artificial Intelligence* 64:299–321.

Smith, Barry, 2002. "Ontology". In: Floridi (2002c). Cited after preprint in Heller et al. (2002).

Smokler, Howard, 1966. "Informational Content: A Problem of Definition". *Journal of Philosophy* 63:201–211.

Stalnaker, Robert, 1971. "Pragmatics". *Synthese* 22:272–289. Reprinted in (Stalnaker 1999a: 31–46).

Stalnaker, Robert, 1986. "Possible worlds and situations". *Journal of Philosophical Logic* 15:109–123.

Stalnaker, Robert, 1991. "The Problem of Logical Omniscience, I". *Synthese* 89:425–440:. Reprinted in (Stalnaker 1999a: 241–254).

Stalnaker, Robert, 1996. "Impossibilities". *Philosophical Topics* 24,1:193–204.

Stalnaker, Robert, 1999a. *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.

Stalnaker, Robert, 1999b. "The Problem of Logical Omniscience, II". In: Stalnaker (1999a), pages 255–273.

von Stechow and Wunderlich, editors, 1991. *Semantik – Semantics: Ein internationales Handbuch der zeitgenössischen Forschung*. Berlin: de Gruyter.

Steinberg and Jakobovits, editors, 1971. *Semantics: An interdisciplinary reader*. Cambridge: Cambridge University Press.

Sterelny, Kim, 1983. "Review of Dretske, Knowledge and the Flow of Information". *Australasian Journal of Philosophy* 61:207–211.

Stine, Gail C., 1974. "Quantified Logic for Knowledge Statements". *Journal of Philosophy* 71:127–140.

Stine, Gail C., 1976. "Skepticism, Relevant Alternatives, and Deductive Closure". *Philosophical Studies* 29:249–260.

Sturdee, David, 1997. "The Semantic Shuffle: Shifting Emphasis in Dretske's Account of Representational Content". *Erkenntnis* 47:89–103.

Thomason, Richmond, editor, 1974. *Formal Philosophy – Selected Papers of Richard Montague*. New Haven: Yale University Press.

Weaver, Warren, 1949. "Recent Contributions to the Mathematical Theory of Communication". *Scientific American* Reprinted in Shannon and Weaver (1949).

Williams, Bernard, 1978. *Descartes – The Project of Pure Enquiry*. London: Penguin, repr. 1990 edition.

Williamson, Timothy, 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.

Williamson, Timothy, 2002. “Knowledge and Scepticism”. In: Frank Jackson and Michael Smith, editors, “The Oxford Handbook of Contemporary Analytic Philosophy”, Oxford: Oxford University Press.

Index