Chapter 2 EPISTEMIC LOGIC AND INFORMATION AS RANGE

Epistemic logic was developed originally for the analysis of the philosophical notion of knowledge (Hintikka 1962), and as such, it has generated controversy. In this book, we use it in a completely neutral manner, as a formalization of the notion 'information as range' in the Introduction. The hard information which an agent currently has is given by a set of possible worlds, and what it 'knows' is that which is true in all worlds of that 'range' (van Benthem 2005). This picture seems close to the common sense notion of information, and it makes knowledge a standard modality. We present some basics in this chapter, stressing points of methodology that will recur. We refer to Blackburn, de Rijke & Venema 2000, Blackburn, van Benthem & Wolter, eds., 2006 for all details in what follows. [11] Occasionally, we will return to issues having to do with the philosophical interpretation of the system, when this makes sense from our general informational point of view.

This first chapter presents some known features of epistemic logic, where the selection of topics reflects the two main uses of any logical system. First, its language and semantics provide a way of describing situations, checking assertions for truth or falsity, engaging in communication, and the like. This leads to semantic issues of *definition* and *expressive power*, and these will be central in this book. The second use of a logical system is as a syntactic *calculus of reasoning*, providing valid inference forms, and deductive axiomatic systems have tended to dominate modal logic, with completeness theorems tying them to the semantics. In this book, the second aspect will be present, but not dominant. To me, modal formalisms are not primarily about inferential 'life-style choices' like *K, S4* or *S5*, the way many people think of the field, but about ways of saying things about agency simply and perspicuously. And tied up with this, there is a third motive of *computational complexity* for key tasks in the two earlier perspectives. We want to strike a good balance between expressive power and computational complexity of logical tasks (cf. Blackburn & van Benthem 2006), and this will also provide our first dynamic angle, in the form of *logic games* for evaluation and other tasks. Our presentation is not a textbook for epistemic logic,

[11] Epistemic logic is flourishing today in computer science, game theory, and other areas beyond its original habitat: cf. Fagin et al. 1995, van der Hoek & Meijer 1995, van der Hoek & Pauly 2007.

[12] but a view of the kind of standard logical system which we will 'dynamify' in this book – not to replace it with something else, but to make it do even further things.

At this point, readers who know their epistemic logic might skip to the next chapters.

1.1 The basic language

We start with the simplest epistemic base language, describing knowledge of individual agents, taken from some set of agents I which is relevant to the application at hand. But let us first illustrate what sort of situation this language is typically supposed to describe.

Example Questions and Answers.
A stranger approaches you in Amsterdam, and asks

Q *Is this building the Rijksmuseum?*

As a well-informed and helpful Dutch citizen, you answer truly

A *Yes.*

As explained in the Introduction, this scenario involves a mix of two factors: (a) factual information, and (b) information about the information of others. In particular, the answer produces *common knowledge* between ***Q*** and ***A*** of the relevant topographical fact. ■

It is worth having a logic that gets clear on these matters, and epistemic logic, though created to describe cogitation by lonesome thinkers, is jus right for this interactive purpose.

Definition Basic epistemic language.
The *basic epistemic language EL* has a standard propositional language with proposition letters and Boolean operators, plus modal operators $K_i\phi$ ('*i knows that* ϕ'), for each $i \in I$, and $C_G\phi$: 'ϕ is common knowledge in the group G'. The inductive syntax rule is as follows, where the '*p*' stands for any choice of atoms from some set of proposition letters:

$$p \mid \neg\phi \mid \phi \vee \psi \mid K_i\phi \mid C_G\phi$$

We write $<i>\phi$ for the existential modality $\neg K_i \neg\phi$: which says intuitively that 'agent i considers ϕ possible'. The existential dual modality of $C_G\phi$ is written as $<C_G>\phi$. [13] ■

[12] A textbook presentation of modal logic in this spirit are the lecture notes van Benthem 2008.

Through its agent-indexed modalities, this basic language describes the preceding scenario.

Example Questions and Answers, continued.

Let $\boldsymbol{Q}$ ask a factual question "$\varphi?$", to which $\boldsymbol{A}$ answers truly: "*Yes*". The presupposition for a normal truthful answer is that $\boldsymbol{A}$ knows that φ: which may be written as $K_A\varphi$. The question itself, if it is a normal co-operative one, conveys at least the presuppositions (i) $\neg K_Q\varphi\ !\ \neg K_Q\neg\varphi$ ('$\boldsymbol{Q}$ does not know if φ') and (ii) $<\boldsymbol{Q}>(K_A\varphi \vee K_A\neg\varphi)$ ('$\boldsymbol{Q}$ thinks that $\boldsymbol{A}$ may know the answer'). After the whole two-step communication episode, φ is known to both agents: $K_A\varphi\ !\ K_Q\varphi$, while they also know this about each other: $K_QK_A\varphi\ !\ K_AK_Q\varphi$, etcetera. Indeed, they achieve the 'limit notion' of common knowledge, written as $C_{\{Q,A\}}P$. ■

Common knowledge is a notion whose importance has been recognized in many areas: from philosophy to game theory, computer science, linguistics, and psychology.

2.2 Models and semantics

The preceding assertions only make precise sense when backed up by a formal semantics. Here is the formal version of the earlier intuitive idea of 'information as range'.

Definition Models.

Models $\boldsymbol{M}$ for the language are triples $(W, \{\rightarrow_i \mid i\in G\}, V)$, where W is a set of worlds, the $\rightarrow_i$ are binary accessibility relations between worlds, and V is a valuation assigning truth values to proposition letters at worlds. In what follows, our primary semantic objects are *pointed models* $(\boldsymbol{M}, s)$ where s is the actual world representing the true state of affairs. ■

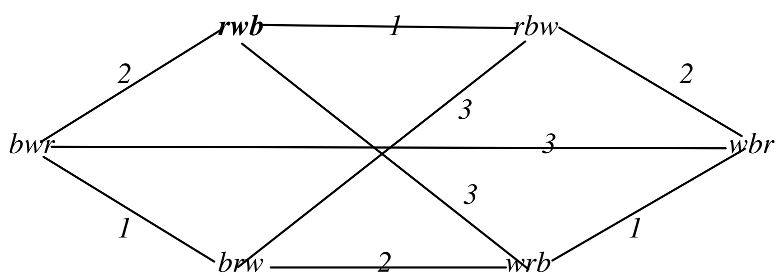
These models stand for collective information states of a group of agents. We impose no general conditions on the accessibility relations, such as reflexivity or transitivity – leaving a 'degree of freedom' for a modeler using the system. Many of our examples work well with equivalence relations (reflexive, symmetric, and transitive) – and such special settings may help the reader. But in some chapters, we will need arbitrary accessibility relations allowing end-points without successors, to leave room for false or misleading information.

---

[13] Common knowledge is a higher-order notion, while individual modalities are first-order universal quantifiers. But in this initial phase of our presentation, this technical difference will not matter.

Example Setting up realistic models.

In what follows, we shall mainly be using very simple epistemic models to illustrate our points. Even so, a real feel for the elegance and utility of the approach only comes from the 'art of modeling', i.e., finding models for real scenarios. Doing so also has the virtue of dispelling delusions of grandeur about 'possible worlds'. Consider this simple game: three cards 'red', 'white', 'blue' are distributed over three players: *1, 2, 3*, who get one each. Each player can see her own card, but not that of the others. The real distribution over the players 1, 2, 3 is *red, white, blue* (***rwb***). Here is the resulting information state:



This pictures the 6 relevant states of the world (the 'hands', or distributions of the cards), with the appropriate accessibilities (equivalence relations in this case) pictured by the uncertainty lines between hands. E.g., the single 1-line between *rwb* and *rbw* indicates that player *1* cannot distinguish these two situations, whereas *2* and *3* can (they have different cards in them). In particular, the diagram says the following, intuitively. Though they are in ***rwb*** (as an outside observer might notice), no player knows this. Of course, the game itself is a dynamic process yielding further information, which will be our theme in Chapter 3. ■

Over these structures, which may often be pictured graphically and concretely as a sort of 'information diagrams', we can now interpret the epistemic language:

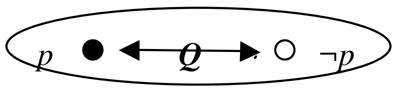
Definition Truth conditions.

$\boldsymbol{M}, s \models p$	iff	*V makes p true at s*
$\boldsymbol{M}, s \models \neg\phi$	iff	*not* $\boldsymbol{M}, s \models \phi$
$\boldsymbol{M}, s \models \phi \wedge \psi$	iff	$\boldsymbol{M}, s \models \phi$ *and* $\boldsymbol{M}, s \models \psi$
$\boldsymbol{M}, s \models K_i\phi$	iff	*for all t with* $s \rightarrow_i t$: $\boldsymbol{M}, t \models \phi$

$\boldsymbol{M}, s \models C_G\phi$ iff *for all t that are reachable from s by some finite sequence of* $\rightarrow_i$ *steps* ($i \in G$): $\boldsymbol{M}, t \models \phi$ [14] ■

Example A model for a question/answer scenario.

Here is how a question answer episode might start (this is just one of many appropriate initial situations!). In the following diagram, reflexive arrows are presupposed, but not drawn. Intuitively, agent $\boldsymbol{Q}$ does not know whether p, but $\boldsymbol{A}$ is fully informed about it:

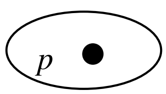


In the black world, the following formulas are true:

$p, K_Ap, \neg K_Qp \,!\, \neg K_Q\neg p, K_Q(K_Ap \vee K_A\neg p),$

$C_{\{Q, A\}}(\neg K_Qp \,!\, \neg K_Q\neg p), C_{\{Q, A\}}(K_Ap \vee K_A\neg p)$

This is an excellent situation for $\boldsymbol{Q}$ to ask $\boldsymbol{A}$ whether p is the case: he even knows that she knows the answer. Once the answer "Yes" has been given, intuitively, this model changes to the following one-point model where maximal information has been achieved:



Now, of course $C_{\{Q, A\}}p$ holds at the black world. ■

Over such models, an epistemic language sharpens distinctions. For instance, that everyone in a group knows ϕ is not common knowledge, but universal knowledge $E_G\phi$, being the conjunction of all formulas $K_i\phi$ for all $i \in G$. Now let us broaden things still further.

From communication to observation and learning Many of the above scenarios can also be viewed as ways of getting information from an arbitrary source: e.g., 'Nature' in the case of observation. In fact, performing observations, or even more general *learning scenarios* are also excellent ways of interpreting the semantics that we have given here. We have just emphasized the conversational setting because it is lively and intuitive.

---

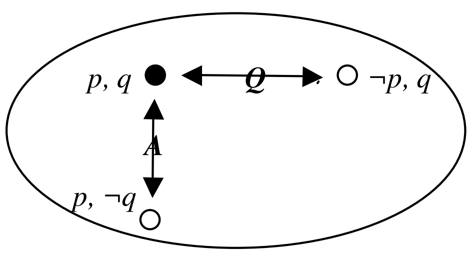
[14] Common knowledge acts as a *dynamic logic* modality $[(\cup_{i \in G} \rightarrow_i)^*]\phi$. This will return below.

From semantics to new language design There is something conservative to 'providing a semantics'. A language is given, and we seek an interpretation that 'fits' it to some model class. But epistemic models are appealing in their own right, as geometrical structures for *informational situations*. And then, there is a question what language best describes these structures, perhaps changing the given formalism. We return to this issue below.

Social information is still richer To illustrate the preceding point, but also a further topic, we notice that the *social character* of group information suggests additional operators.

Example From implicit to explicit group knowledge.

Consider a setting where both agents have information that the other lacks, say as follows:



Here, the black dot is the actual world. Now the most cooperative scenario would be for $\boldsymbol{Q}$ to tell $\boldsymbol{A}$ that q is the case (after all, this is something he knows), while she can tell him that p is the case. Intuitively, this reduces the initial three-point model to the one-point model where $p \wedge q$ is common knowledge. Other three-world examples model other interesting, sometimes surprising conversational settings (cf. van Benthem 2006). ■

Another way of describing what happens in the preceding example is that, when $\boldsymbol{Q}$, $\boldsymbol{A}$ do the best they can in informing each other, they will cut things down to the *intersection* of their individual accessibility relations. This suggests a new natural notion for groups:

Definition Distributed knowledge.

Intuitively, a formula ϕ is *implicit* or *distributed knowledge* in a group, written $D_G\phi$, when agents could come to see it by pooling their information. More technically, extending our language and the above truth definition, this involves intersection of accessibility relations:

$$\boldsymbol{M}, s \models D_G\varphi \quad \text{iff} \qquad \text{for all } t \text{ with } s \bigcap_{i\in G} \rightarrow_i t\text{: } \boldsymbol{M}, t \models \varphi$$ ■

Intuitively, groups of agents can turn their implicit knowledge into common knowledge (modulo some technicalities) by communication. We will pursue this in Chapters 3, 10.

Digression: epistemic models once more Though our system works over arbitrary models, in practice, equivalence relations are common. Our card examples were naturally described as follows: each agent has possible 'states' it can be in (say, the card deals it may receive), and worlds are global states consisting of vectors X, Y with each agent in some local state. Then, clearly, the natural accessibility relation goes via component-wise equality:

$$X \to_i Y \text{ iff } (X)_i = (Y)_i.$$

Another case are the models for games in Chapters 9, 14, with worlds vectors of strategies for players. The latter application is a sort of 'normal form' (van Benthem 1996):

Fact Every multi-S5 model is representable as a sub-model of a vector model.

Proof Each agent has for its states the equivalence classes of its accessibility relation, and this is an isomorphism since the i- equivalence classes of worlds s, t are equal iff $s \to_i t$. ■

2.3 Validity and axiomatic systems

Validity of formulas φ in epistemic logic is defined as usual in semantic terms, as truth of φ in all models at all worlds. Consequence may be defined through validity of conditionals.

Minimal logic The following completeness result says that the validities over arbitrary models may be described purely syntactically by the following calculus of deduction: [15]

Theorem The valid formulas are precisely the theorems of the minimal epistemic logic axiomatized by (a) all valid principles of propositional logic, (b) the definition $<>\varphi \leftrightarrow \neg K\neg\varphi$, (c) modal distribution $K(\varphi \to \psi) \to (K\varphi \to K\psi)$, together with the inference rules of Modus Ponens (from $\varphi \to \psi$ and φ to ψ') and Necessitation ('if φ is a theorem, then so is $K\varphi$').

---

[15] We will often drop agent subscripts for K-operators when they play no essential role.

Proof The proof for this basic result can be found in any good textbook. Modern versions employ Henkin-style constructions with maximally consistent sets over some finite set of formulas only, producing a finite counter-model for a given non-derivable formula. ■

One axiom of this simple calculus has sparked continuing debate, viz. the distribution law $K(\varphi \to \psi) \to (K\varphi \to K\psi)$. This seems to say that agents' knowledge is closed under logical inferences, and this 'omniscience' seems unrealistic. At stake here is our earlier distinction between semantic and inferential information. Semantically, with processes of observation, an agent who has the hard information that $\varphi \to \psi$ and that φ, also has the hard information that ψ. But in a more finely-grained perspective of inferential information, geared toward processes of elucidation, this need not be the case at all. We will not join the fray here, but Chapter 5 shows one way of introducing inferential information into basic epistemic logic.

Another point to note is this. Deductive systems may be used in two modes. They describe (a) agents' own explicit reasoning 'inside' scenarios, or (b) 'outside' reasoning by theorists about them. In some settings, the difference will not matter: the modeler is 'one of the boys' – but sometimes, it may. In this book, we will not pursue discussions between 'first person' and 'third person' perspectives on epistemic logic, as we can accommodate both.

Stronger epistemic logics and frame correspondence On top of this minimal system, two further steps go hand in hand: helping ourselves to stronger axioms endowing agents with further features, and imposing further structural conditions on accessibility in our models. For instance, here are three more axioms with vivid epistemic interpretations:

$K\phi \to \phi$	*Veridicality*
$K\phi \to KK\phi$	*Positive Introspection*
$\neg K\phi \to K\neg K\phi$	*Negative Introspection*

The former seems uncontroversial (knowledge is 'in synch' with reality), but the latter two have been much-discussed, since they assume that, in addition to their logical omniscience, agents now also have capacities of unlimited introspection into their own epistemic states. Formally, these axioms correspond to the following structural conditions on accessibility:

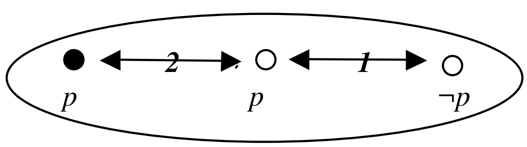
$K\phi \to \phi$	*reflexivity*	$\forall x: x \to x$
$K\phi \to KK\phi$	*transitivity*	$\forall xyz: (x \to y \wedge y \to z) \Rightarrow x \to z$
$\neg K\phi \to K\neg K\phi$	*euclidity*	$\forall xyz: (x \to y \wedge x \to z) \Rightarrow y \to z$

The term 'correspondence' can be made precise using 'frame truth' of modal formulas under all valuations on possible worlds with accessibility (modaL Correspondence Theory, cf. van Benthem 1984). Powerful general results exist describing which axioms match up with what conditions: first-order like here, or definable in higher-order languages. We will occasionally refer to such techniques, but refer to the literature for details.

The complete deductive system with all the above axioms is called *S5*, or 'multi-*S5'* when we have more than one agent. Taking the preceding conditions together, it is the logic of equivalence relations over possible worlds. What may be surprising is that this logic has no 'interaction axioms' relating different modalities K_i, K_j. But none are plausible.

Example Your knowledge and mine do not commute.

The following model provides a counter-example to the implication $K_1 K_2 p \to K_2 K_1 p$. Its antecedent is true in the black world to the left, but its consequent is false:



Such implications only hold when agents stand in special informational relationships. ■

In the dynamic-epistemic logics of Chapter 3 and later ones, some operator commutation principles will hold, but then between epistemic modalities and action modalities.

In between the minimal logic *K* and *S5*, many other epistemic logics have a following, such as *'KD45'* for knowledge. In this book, however, we will mostly use intuitions from both extremes, though the reader will see that our results apply much more generally.

Once again: deductive versus expressive power The traditional emphasis in modal logic is on different axiomatic systems for the same language. This is the dimension of *deductive* *power*, and it corresponds to modal theories of different model classes. But there is an independent dimension of *expressive power*, i.e., which language to use for describing our structures, weaker or stronger. The basic epistemic language is one candidate, but there are many extensions, adding 'universal modalities' ranging over all worlds (accessible or not), or 'nominals' picking out single worlds. The most striking addition to the base language in the epistemic setting has been the earlier-mentioned modality of *common knowledge*.

Theorem The complete epistemic logic with common knowledge is axiomatized by the following two principles in addition to the minimal epistemic logic, where E_G is the earlier modality for 'everybody in the group knows':

$$C_G\,\phi \leftrightarrow \phi \,\&\, E_G\, C_G\, \phi \qquad \textit{Fixed-Point Axiom}$$

$$(\phi \,\&\, C_G\,(\phi \rightarrow E_G\,\phi)) \rightarrow C_G\,\phi \qquad \textit{Induction Axiom}$$

These axioms are also of independent interest for what they say. The Fixed-Point Axiom expresses an intuition which has been called 'reflective equilibrium': common knowledge of ϕ is a proposition implying ϕ of which every group member knows that *it* is true. On top of this, the Induction Axiom says that it is not just any equilibrium state of this kind, but the largest one: technically, we have a 'greatest fixed-point' (cf. Chapters 3, 4 below).

2.4 Invariance and expressive power

Expressive power of languages is measured by their powers of *distinguishing* different models, or equivalently, by what they consider indistinguishable. Thus, betweenness is fundamental to geometry because this configuration of objects is invariant under the natural spatial movements of translation and rotation. For first-order logic, a basic invariance is *isomorphism*: structure-preserving bijection between models: first-order formulas cannot distinguish between a tuple of objects $\boldsymbol{a}$ in one model $\boldsymbol{M}$, and its image $f(\boldsymbol{a})$ in another model $\boldsymbol{N}$ related to $\boldsymbol{M}$ via an isomorphism f. [16] This style of analysis is quite widespread.

First-order translation For a start, viewed as a description of the above epistemic models, our language is weaker than a first-order logic whose variables range over worlds:

[16] Another candidate is 'potential isomorphism': van Benthem 2002 has some extensive discussion.

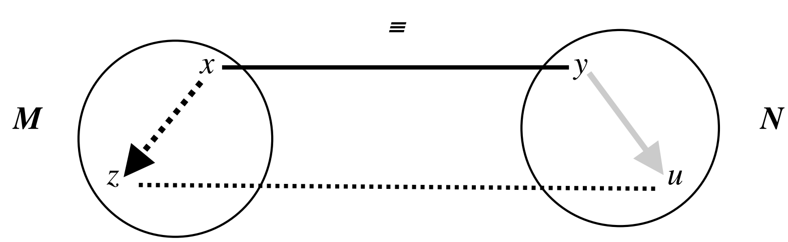
Fact There exists an effective translation from the basic epistemic language into first-order logic yielding equivalent formulas.

Proof This is the well-known 'standard translation', taking, e.g., an epistemic formula $p \wedge K\Diamond q$ into an equivalent first-order formula $Px \wedge \forall y(Rxy \rightarrow \exists z(Ryz \wedge Qz))$. The latter formula is a straightforward transcription of the truth conditions of the epistemic one. ■

These modal translations have only special 'bounded' or guarded' quantifiers over accessible worlds, making them a rather special subclass of first-order logic as a whole. This also shows in powers of distinction. For instance, the full first-order language can distinguish between a one-point reflexive model and an irreflexive 2-cycle - but it should be intuitively clear that these two models verify the same epistemic formulas everywhere.

Bisimulation and information equivalence Here is a notion of semantic invariance which fits the modal language like a Dior gown:

Definition A *bisimulation* between two models ***M, N*** is a binary relation $\equiv$ between their states *s, t* such that, whenever $s \equiv t$, then (a) *s, t* satisfy the same proposition letters ('local harmony'), (b1) if $s\ R\ s'$, then there exists a world t' with $t\ R\ t'$ and $s' \equiv t'$, and (b2) the same 'zigzag' or 'back-and-forth clause' holds in the opposite direction.



■

Clause (1) expresses 'local harmony', the zigzag clauses (2) the dynamics of simulation. Bisimulation is usually given a 'procedural' spin: it identifies processes that run through similar states with similar local choices. Thus, it gives one answer to the crucial question:

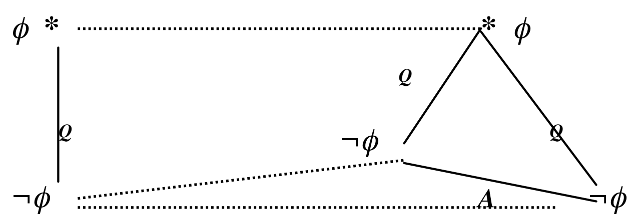
When are two processes the same?

We say 'one answer': as there are often different natural ways of providing a criterion of identity between structures, [17] and process theory is no exception (cf. van Benthem 1996). We will meet the same fundamental issue with games in Chapter 14 below, but right here, we just note that structural identity also makes sense for information structures:

When are two information structures the same?

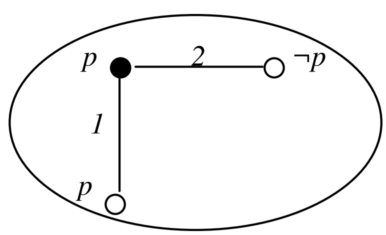
Example Bisimulation-invariant information structures

Our earlier question-answer example has a bisimulation with the following variant:

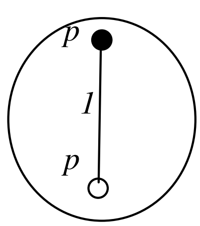


In a natural sense, both models represent *the same information state* for the agents.

Bisimulation also occurs naturally in information update changing a current model. Suppose that the initial model is like this, with the actual world indicated by the black dot:

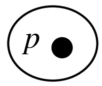


All three worlds satisfy different epistemic formulas. Now, despite her uncertainty, in the actual world, agent *1* does know that *p*, and can say this – updating to the model



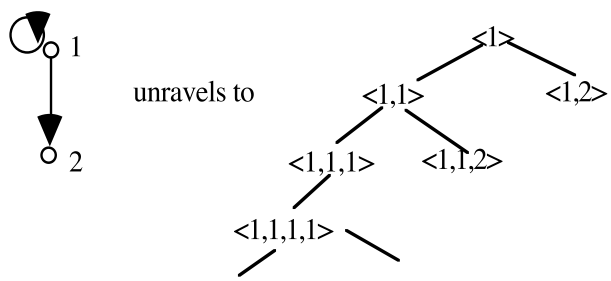
[17] This diversity is sometimes called 'Clinton's Principle': *it all depends on what you mean by 'is'*.

But here the two worlds are intuitively redundant, and indeed this information state for the two agents has an obvious bisimulation to just the one-point model



■

More generally, each model has a smallest *bisimulation contraction* satisfying the same modal formulas. But bisimulation can also make models larger, turning them into *trees* through the method of '*unraveling*', providing a geometrically perspicuous 'normal form':



Thus, informational equivalence can work both ways, depending on what we find useful.

Invariance and definability Some basic model-theoretic results tie bisimulation closely to truth of modal formulas. For convenience, we restrict attention to *finite models* – but this is easily generalized in the theory of modal logic. We formulate results for general relations:

Invariance Lemma The following two assertions are equivalent:

(a) $\boldsymbol{M}, s$ and $\boldsymbol{N}, t$ are connected by a bisimulation,

(b) $\boldsymbol{M}, s$ and $\boldsymbol{N}, t$ satisfy the same epistemic formulas.

Proof The direction from (a) to (b) is a straightforward induction on the construction of epistemic formulas. In the opposite direction, for finite models, we can even make do with the basic modalities. Let $(\boldsymbol{M}, x)\ Z\ (\boldsymbol{N}, y)$ be the relation which holds if the worlds x and y satisfy the same epistemic formulas. For a start, then, $s\ Z\ t$. Clearly also, Z-related worlds satisfy the same proposition letters. Now, assume that $x\ Z\ y$, and let $x \to x'$. Suppose for contradiction, that there is no world y' with $y \to y'$ satisfying the same modal formulas. Then there is a modal formula $\alpha_{y'}$ true in x' and false in y'. Taking the conjunction α of all these formulas $\alpha_{y'}$ for all successors y' of y, we see that α is true in x', and so $\Diamond\alpha$ is true in world x, whence $\Diamond\alpha$ is also true in y, as $x\ Z\ y$. But then there must be a successor y_* of y satisfying α: which contradicts the construction of α. [18] ■

This connection between the epistemic language and bisimulation-invariant structure raises the issue to which extent these are two sides of the same coin. This issue is relevant to the choice of 'information states' that we will be using later on in this book. The modal theory of a world w in a model $\boldsymbol{M}$ is an explicit record of everything that is true 'internally' at w about the facts, agents' knowledge of these, and their knowledge of what others know. By contrast, models $(\boldsymbol{M}, w)$ themselves locate the same information in the local valuation of a world plus its pattern of interaction with other worlds – just as category theorists describe a mathematical structure 'externally' by its connections to other structures in a category through the appropriate morphisms. The following result says essentially that states in an epistemic model and maximally consistent epistemic theories are indeed equivalent.

State Definition Lemma For each model $\boldsymbol{M}, s$ there exists an epistemic formula β (involving common knowledge) such that the following are equivalent:

(a) $\boldsymbol{N}, t \models \beta$

(b) $\boldsymbol{N}, t$ has a bisimulation $\equiv$ with $\boldsymbol{M}, s$ such that $s \equiv t$

Proof This result from Barwise & Moss 1997 adapts the technique of 'Scott sentences' (our version follows van Benthem 1997). This time, we show the proof for the case of S5, but it works for arbitrary accessibility relations. Any finite multi-*S5* model $\boldsymbol{M}, s$ falls into maximal 'zones' of worlds that satisfy the same epistemic formulas in our language.

Claim 1 There exists a finite set of formulas ϕ_i $(1 \leq i \leq k)$ such that (a) each world satisfies one of them, (b) no world satisfies two of them (i.e., they define a partition of the model), and (c) if two worlds satisfy the same formula ϕ_i, then they agree on all epistemic formulas.

To see this, take any world s, and find 'difference formulas' $\delta^{s,t}$ between it and any t not satisfying the same epistemic formulas: say s satisfies $\delta^{s,t}$ and t does not. The conjunction

---

[18] The lemma even holds for arbitrary models, provided we take epistemic formulas from a language

of all $\delta^{s, t}$ is a formula ϕ_i true only in s and the worlds sharing its epistemic theory. We may assume ϕ_I also lists all information about proposition letters true and false throughout their 'zone'. We also make a quick observation about uncertainty links between these zones:

\# If any world satisfying ϕ_I is $\sim_a$-linked to a world satisfying ϕ_i,
then all worlds satisfying ϕ_I also satisfy $<a>\phi_i$

Next take the following description $\beta_{M, s}$ of $\boldsymbol{M}$, s:

(a) all (negated) proposition letters true at s plus the unique ϕ_i true at $\boldsymbol{M}$, s

(b) common knowledge for the whole group of

- (b1) the disjunction of all ϕ_i
- (b2) all negations of conjunctions $\phi_I \,\&\, \phi_j\ (i{\neq}j)$
- (b3) all implications $\phi_i \rightarrow <a>\phi_j$ for which situation # occurs
- (b4) all implications $\phi_i \rightarrow [a] \vee \phi_j$ where the disjunction runs over all situations listed in the previous clause.

Claim 2 $\boldsymbol{M}, s \models \beta_{M, s}$

Claim 3 If $\boldsymbol{N}, t \models \beta_{M, s}$, then there is a bisimulation between $\boldsymbol{N}$, t and $\boldsymbol{M}$, s

To prove Claim 3, let $\boldsymbol{N}$, t be any model for $\beta_{M, s}$. The ϕ_i partition $\boldsymbol{N}$ into disjoint zones Z_i of worlds satisfying these formulas. Now relate all worlds in such a zone to all worlds that satisfy ϕ_i in the model $\boldsymbol{M}$. In particular, t gets connected to s. We check that this gives a bisimulation. The atomic clause is clear by construction. But also, the zigzag clauses follow from the given description. (a) Any $\sim_a$-successor step in $\boldsymbol{M}$ has been encoded in a formula $\phi_i \rightarrow <a>\phi_j$ which holds everywhere in $\boldsymbol{N}$, producing the required successor there. (b) Conversely, if there is no $\sim_a$-successor in $\boldsymbol{M}$, this shows up in the limitative formula $\phi_i \rightarrow [a] \vee \phi$, which also holds in $\boldsymbol{N}$, so that there is no 'excess' successor there either. ■

The Invariance Lemma says that bisimulation has the right fit with the modal language. The State Definition Lemma strengthens this to say that each semantic state is captured by

---

with arbitrary *infinite* conjunctions and disjunctions. We forego such technicalities here.

one epistemic formula. This result extends to arbitrary models provided we are willing to use formulas from a language allowing arbitrary infinite conjunctions and disjunctions.

Example Defining a model up to bisimulation.
Consider the two-world model for our earlier basic question-answer episode. Here is an epistemic formula which defines its ϕ-state up to bisimulation:

$$\phi \,\&\, C_{\{Q,\,A\}}\, ((K_A\phi \vee K_A\neg\phi) \,\&\, \neg K_Q\phi \,\&\, \neg K_Q\neg\phi) \qquad \blacksquare$$

These results allows us to switch, in principle, between semantic and syntactic accounts of information states. Syntactic approaches have been dominant in areas like belief revision theory, and semantic ones in the dynamic logics we will discuss later on. The reader might want to keep this duality in mind, even as we will pursue the semantic road.

2.5 Computation and the complexity profile of a logic

While derivability and definability are the main pillars of logic, in recent years, there has been a growing awareness that issues of *task complexity* form a natural complement. Given that information has to be recognized or extracted to be of use to us, it is natural to ask how complex such extraction processes really are. We start with a well-known observation.

Decidability In traditional modal logic, the interest in complexity has gone no further than just asking the following question. Validity in first-order logic is *undecidable*, validity in propositional logic is *decidable*: what about modal logic, which sits in between?

Theorem Validity in the minimal modal logic is decidable. So is that for multi-*S5*.

Proof One proof is via the *effective finite model property*: each satisfiable modal formula has a finite model whose size can be computed effectively from the length of the formula. [19] One can first unravel the given model modulo bisimulation to a tree, and then cut the tree at a depth equal to the operator depth of the formula, while cutting width to a size matching the number of sub-formulas in the formula. Another proof notes that the earlier *translation* needs only *two variables*, and the 2-variable fragment of first-order logic is decidable. Further proofs use *filtration*, semantic tableaux, or analysis of sequent calculi. ■

---

[19] Note that this property typically fails for the full language of first-order logic.

Validity is decidable for many well-known modal logics. [20] But things change rapidly when we consider logics with combinations of modalities that show some 'interaction':

Theorem The minimal modal logic of two modalities *[1], [2]* satisfying the axiom $[1][2]\varphi \rightarrow [2][1]\varphi$ plus a *universal modality U* over all worlds is undecidable.

Proof In this logic, one can encode undecidable *tiling problems* on the structure *INxIN* (Harel 198x, Marx 2007). The reason is essentially that the given special axiom defines a *grid structure* satisfying the following first-order convergence property:

$$\forall xyz: (xR_1y \wedge yR_2z) \rightarrow \exists u: (xR_2u \wedge uR_1z). \qquad \blacksquare$$

In a slogan: modal logics of trees are harmless, modal logics of grids are dangerous!

Tree *grid*

Since modal logics with such interaction axioms tend to occur in theories of agency, where we pile up descriptions of their separate abilities in terms of modalities, undecidability is always around the corner there (cf. Chapter 11). But in this chapter, we first look the other way, and enquire into the fine-structure of the decidable tasks that we do have.

Complexity: a sketch Complexity theory studies the computation time or memory space needed for some computational task, measured as a function of input size. In particular, inside the decidable problems, it distinguishes 'feasible' rates of growth such as linear, quadratic, or general polynomial (the complexity class called ***P***) from 'non-feasible' ones like non-deterministic polynomial time problems in ***NP***, and beyond to polynomial space (***Pspace***), exponential time (***Exptime***), and ever higher up. To get some feeling for natural tasks calibrated in this way: roughly speaking, sorting a given sequence by magnitude, or finding a path from a given node of a graph to another 'goal node', takes quadratic time,

---

[20] Validity is also decidable for many other modal logics over more specialized model classes. In particular, *S5* may be translated into *monadic first-order logic*, which is obviously decidable.

parsing a sentence grammaticality takes cubic time, solving a Traveling Salesman problem is in ***NP***, while finding a winning strategy in many parlour games is in ***Pspace***. Making these notions precise is a bit tedious, and often puts people off by its accidental details of the workings of Turing machines or other devices. We will therefore stick with this basic outline, and rely on the reader's intuitive understanding of the discussion to follow.

Complexity profile of a logic To really understand how a logical system works, it helps to check the complexity for some basic tasks that it would be used for. Determining validity, or equivalently, *testing for satisfiability* of given formulas, is one of these:

Given a formula φ, determine whether φ has a model.

But there are other, equally important tasks, such as *model checking*:

Given a formula φ and a finite model $(\boldsymbol{M}, s)$, check whether $\boldsymbol{M}, s \models \varphi$.

And here is a third key task, related to our discussion of invariance and expressive power, which one might call *testing for model equivalence*:

Given two finite models $(\boldsymbol{M}, s)$, $(\boldsymbol{N}, t)$, check if they satisfy the same formulas.

Let us draw a table of the complexity profiles for two well-known classical logics around modal ones. Here entries mean that the problems are in the class indicated, and no higher:

	Model Checking	Satisfiability	Model Comparison
Propositional logic	*linear time (**P**)*	***NP***	*linear time (**P**)*
First-order logic	***Pspace***	*undecidable*	***NP***

Where does the basic modal language fit? Its model-checking problem, i.e., its best performance using the above truth definition, turns out to be efficient, and fast. While first-order model-checking requires a complete search through the whole model, repeated with every quantifier iteration and hence generating exponential growth, there are much faster modal algorithms. Also, close-reading the above decidability argument tells us where to locate satisfiability. Finally, the third basic task, of testing for modal equivalence, or equivalently, for the existence of a bisimulation, has turned out relatively efficient, too:

Fact The complexity profile for the minimal modal logic is as follows:

Model Checking	Satisfiability	Model Comparison
Ptime	***Pspace***	***Ptime***

Proof (a) A best model-checking algorithm needs time steps of order $length(\varphi)*size(\boldsymbol{M})^2$. One computes truth values for all of φ's sub-formulas (there are $length(\varphi)$ of these), bottom-up, in all worlds of the model. For each sub-formula, this takes one pass through the model: $size(\boldsymbol{M})$ steps. The number of steps per world is bounded: we compute a truth value for (i) atoms (1 step look-up), (ii) a Boolean operator (≤ 3 steps), or (iii) <> with ψ's truth values already marked in each world. The latter subroutine takes at worst $size(\boldsymbol{M})$ steps, searching through all successors of the current world. (b) For satisfiability, the crux is that using the finite model property need not produce a whole model for a given formula, but one can check for its existence branch by branch, erasing 'work space', making do with polynomial space. (c) Finally, finding a bisimulation is a simple process of starting 'from above', with all links between all worlds in the two models, and then successively eliminating all links that could never make it into any bisimulation. ■

This is still sloppy. Our analysis of algorithms found an *upper bound* on the complexity of the logical task. To show a *lower bound*, one must prove that *all* algorithms solving the task have at least this complexity. One common method uses a set of 'calibration problems' whose precise complexity has been determined. One now *reduce* some such problem, say in ***NP*** or ***Pspace,*** to satisfiability in modal logic in polynomial time, for a lower bound.

For epistemic logic, results are the same. But with *S5*, there is one difference. Single-agent satisfiability is in ***NP***, since *S5* allows for a normal form without iterated modalities, close to propositional logic. But with two agents, and two equivalence relations in the models, satisfiability jumps back to ***Pspace***: social life is more complicated than being alone.

Complexity results are affected by expressive power of the language. When we add our *common knowledge* modality $C_G\varphi$ to epistemic logic, the above profile changes as follows:

Model Checking	Satisfiability	Model Comparison
P	***Exptime***	***P***

Best algorithms for model checking with common knowledge are computationally clever; and the analysis of decidability changes. For our purposes, we just draw a general lesson.

'The Balance': expressive power versus computational effort In logic just as in physics, there is a 'Golden Rule' of design: what you gain in one desirable dimension, you lose in another. In particular, high expressive power of a system means high complexity. Thus, first-order logic itself is expressively weaker than second-order logic, the natural medium for defining many mathematical notions. But its poverty has a striking reward, viz. the recursive enumerability of valid consequence, and hence the Completeness Theorem, as well as powerful model-theoretic existence properties such as the Compactness Theorem. A major methodological issue in logic is then to strike a *Balance* between expressive power and computational complexity. Many modal logics are good compromises on this road.

2.6 Games for logical tasks

Computation is not just a routine 'chore' which measures difficulty of tasks: it is also a fundamental process in its own right. This comes out particularly well with *game versions* of logical tasks, which do not just 'get the results' extensionally, but also model the 'intensional' procedural activity behind them in a vivid manner (van Benthem 2007).

Evaluation games We cast the process of evaluating modal formula φ in model $(\boldsymbol{M}, s)$ as a *two-person game* between a 'Verifier', claiming that φ is true, and a 'Falsifier' claiming that it is false. The game starts at some world s. Each move is dictated by the main operator of the formula at hand (the total length is again bounded by its modal depth):

Definition The modal evaluation game $game(\boldsymbol{M}, s, \varphi)$.

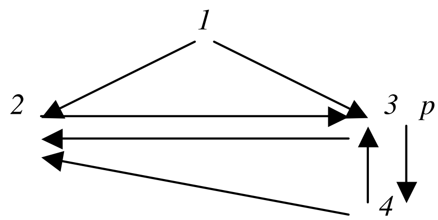
The rules of the games endow the logical operators with an interactive dynamic meaning:

atom p	test p at s in $\boldsymbol{M}$: if true, then V wins – otherwise, F wins,
disjunction	V *chooses* a disjunct, and play continues with that,
conjunction	F chooses a conjunct and play continues with that,
$<>\varphi$	V picks an R-successor t of the current world; play continues with φ at t,
$[]\varphi$	F picks an R-successor t of the current world; play continues with φ at t.

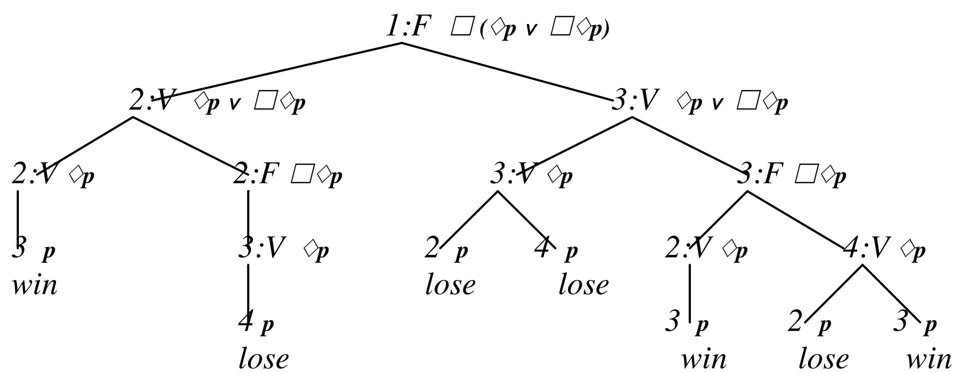
A player also loses when (s)he must pick a successor, but cannot do so. ■

Example A complete game tree.

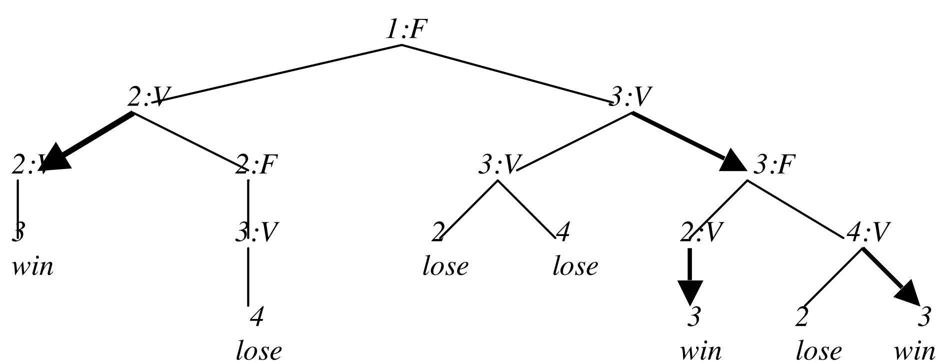
Again, we give an illustration in general modal logic. Here is the complete game tree for the modal formula *□(◇p v □◇p)* played starting at state *1* in the following model:



We draw game nodes, plus the player which is to move, plus the relevant formula. Bottom leaves have ‘win’ if Verifier wins (the atom is true there), ‘lose’ otherwise.



Each player has three winning runs, but the advantage is for Verifier, who can always play to ensure a win, whatever Falsifier does. Her *winning strategy* is indicated in black here:



■

A strategy can encode subtle dynamic behaviour: Verifier must hand the initiative to Falsifier at state *3* on the right if she is to win! The reason why Verifier has a winning strategy is that she is defending a true statement. The background for these games and their properties is in Chapter 14, but the main point for us here is this simple observation:

Fact For any modal evaluation game, the following two assertions are equivalent:

(a) formula φ is true in model $\boldsymbol{M}$ at world s,

(b) player V has a winning strategy in $game(\boldsymbol{M}, s, \varphi)$.

Proof The proof is a straightforward induction on the formula φ. This is a useful exercise to get a feel for the workings of strategies, and the 'game dynamics' of the modalities. ■

Model comparison games The next game provides fine-structure behind our invariance of bisimulation between models $(\boldsymbol{M}, s)$, $(\boldsymbol{N}, t)$. They involve a Duplicator (maintaining there is an analogy) and Spoiler (claiming a difference), playing over pairs (x, y) in the models.

Definition Modal comparison games.

In each round of the game, Spoiler chooses a model, and a state u which is a successor of x or y, and Duplicator responds with a corresponding successor v in the other model. Spoiler wins if u, v differ in their atomic properties, or Duplicator cannot find a successor in 'her' model. The game continues over some finite number of rounds, or infinitely. ■

There is a tight connection between these games, and modal properties of the two models:

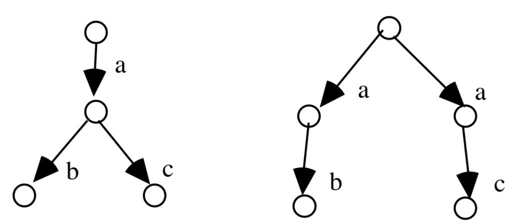
Fact (a) Spoiler's winning strategies in a k-round game between $(\boldsymbol{M}, s)$ and $(\boldsymbol{N}, t)$ match exactly the modal formulas of *operator depth k* on which s, t disagree.

(b) Duplicator's winning strategies in the *infinite* round game between $(\boldsymbol{M}, s)$ and $(\boldsymbol{N}, t)$ match the bisimulations between $\boldsymbol{M}$, $\boldsymbol{N}$ linking s to t.

It is not hard to prove this result generally, but we merely give some illustrations.

Example 'Choosing now or later'.

Consider the game between the following two models starting from their roots:



Spoiler can win the comparison game in *2* rounds – and different strategies do the job. One stays inside the same model, exploiting the modal difference formula $<a>(<b>T \;\&\; <c>T)$ of depth *2* with uniform modalities. Another winning strategy for Spoiler switches models, but it can make do with the modal formula $[a]<b>T$ which contains only *2* operators in all. The switch is signaled by the interchange in the modalities. ■

Games for satisfiability or model construction Games for model construction sometimes use semantic tableaux (van Benthem, to appear). Here we state a more abstract version:

Definition A modal satisfiability game.

Let $G = (S, R_I, R_{II})$, with S being all sets of (negated) sub-formulas of some given formula φ, and $s_0 = \{\varphi\}$. Player I defends the *SAT*-claim; *II* challenges it. From a winning strategy for player *I*, one gets a model. Call a state $s \in S$ *full* if (a) for each formula in s, each of its sub-formulas is either in s, or has its negation in s (never both), (b) conjunctions in s have both conjuncts in s, disjunctions in s have at least one disjunct in s. The *modal rank* of s is the maximum of all modal ranks of formulas in s. As for moves, R_I has all transitions (s, s') where s' is a full state extending s of the same modal rank – and R_{II} has all transitions from s to s' where some formula θ has $\Diamond\theta \in s$ and $s' = \{\theta\} \cup \{\delta \mid []\delta \in s\}$. [21] ■

Fact (a) Player *I* has a winning strategy in the game from G, s_0 iff $\varphi \#$ *SAT*,

(b) G, s_0 always terminates within m rounds, with m the modal depth of φ,

(c) all states s are polynomially-sized bounded in *length*(φ). [22]

This concludes our first introduction to logic games, which show modal logic, and hence also epistemic logic *itself* in a dynamic interactive light. We discuss this issue of logics themselves being dynamic objects at greater length in (van Benthem, to appear). For here, just observe that that the above *game trees themselves are modal models*. This is a foothold for applications of modal logic to game theory, as we will see starting from Chapter 10.

[21] To understand these moves, it helps if you have seen a Henkin-style completeness argument.

[22] This is an alternative proof that satisfiability for the minimal modal logic is in ***Pspace***.

2.7 Summary and conclusion

We have surveyed epistemic logic in a somewhat unusual manner. Our purpose has been three-fold. First, we needed to lay the groundwork for the notions and techniques that will be used later on in this book. Next, we wanted to shake off the philosophical dust which makes people think that epistemic logic is a failed attempt at grasping the epistemological notion of knowledge, whereas it is better viewed as a logic for the notion of semantic information. And finally, we have presented it as a role model for what we take a logical system to be, a combination of three main aspects: (a) *expressive power of definition*, (b) *inferential power of deduction*, and (c) *computational task performance*, eventually leading to a suggestive interactive embodiment of the logic itself in dynamic *games*.

2.8 Appendix: further topics and open problems

There are many further issues about knowledge, information, and logic that we have left out here. A few were mentioned already, such as connections with *dynamic logic* and modal fixed-point logics, or the contrast between *first person and third person* perspectives in interpreting our models. We mention a few more, with pointers to the literature.

Internal versus external logical views of knowledge Epistemic logic builds on standard classical propositional logic, while adding explicit operators for knowledge. By contrast, intuitionistic logic treats knowledge 'internally' by 'epistemic loading' of the interpretation of standard logical constants like negation and implication. Van Benthem 1993, 2008 discuss the contrast between the two systems, and draw deeper comparisons. The dynamic content of intuitionistic logic involves observational update (Chapter 3), awareness-raising actions (Chapter 5), as well as 'procedural information' in the sense of Chapter 11. The final section of Chapter 5 supplies some details. Even so, there is a general question what a more intuitionistic-style version would look like for the dynamics explored in this book.

Agent powers and diversity Epistemic logic incorporates a number of idealizations about agents having information. They possess unlimited deductive powers, and there are also more hidden assumptions of perfect memory, that will return in Chapters 3, 4. How can we lift this, and make our logics an account of actual agents with bounded rationality? Liu 2007 is a discussion of the main issues, with first attempts a finding 'parameters' inside dynamic-epistemic logics that allow for cooperation between diverse agents. But in general, the theory of this book is about idealized agents still, and we need extensions.

Topological models Relational possible worlds models are a special case of a more general, and historically older semantic paradigm for modal logic, viz. *topological models*. Let ***M*** be a topological space $(X, \mathbb{O}, V)$ with points X, a family of open sets $\mathbb{O}$, and a valuation V as in modal models. Now we can interpret a modal language as follows:

Definition Topological semantics.

*ϕ is true at point s in **M***, written $\boldsymbol{M}, s \models \Box\phi$, if s is in the *topological interior* of $[[\phi]]^{\boldsymbol{M}}$: the set of points satisfying ϕ in ***M***. Formally, $\exists O\in\mathbb{O}: s\in O\ \&\ \forall t\in O: \boldsymbol{M}, t \models \phi$. ■

The existential $\Diamond$ denotes topological *closure*. Typical topologies are metric spaces like the real line, or trees with the closed sets in the tree order as opens. The latter coincides with relational semantics. In this setting, the modal axioms of *S4* state topological properties: $\Box\phi \to \phi$ is inclusion, $\Box\Box\phi \leftrightarrow \Box\phi$ is idempotence, and $\Box(\phi \wedge \psi) \leftrightarrow \Box\phi \wedge \Box\psi$ is closure of opens under intersections. Infinitary distribution, valid in relational semantics, fails here:

Example $\Box \bigwedge_{i\in I} p_i \leftrightarrow \bigwedge_{i\in I} \Box p_I$ fails on metric spaces.

Interpret p_i as the open interval $(-1/i, +1/i)$, all $i\in N$. The $\Box p_i$ denote the same interval, and their intersections are the intersection $\{0\}$ of all these intervals. But the expression $\Box \bigwedge_{i\in I} p_i$ denotes the topological interior of the singleton set $\{0\}$, which is the empty set $\varnothing$. ■

A general completeness result says a modal formula is topologically valid iff it is provable in *S4*. [23] There are also much deeper special completeness results, such as the fact that *S4* is complete for all validities on metric spaces without isolated points, like the reals. All earlier modal techniques generalize, including a notion of 'topological bisimulation' related to homeomorphisms and continuous maps, plus a vivid matching model comparison game. [24]

---

[23] The minimal modal logic arises over the generalized class of *neighbourhood models* where the topological structure is replaced by an abstract point–to-set neighbourhood relation *RxY*.

[24] The Handbook Aiello, Pratt & van Benthem, eds. 2007 has several relevant chapters.

Van Benthem & Sarenac 2005 use this semantics for epistemic logic in this way, showing how the multi-agent setting leads to families of topologies closed under certain operations that correspond with group-forming operations. In particular, they exploit the failure of infinitary distribution to show the following separation of iterative and fixed-point views of common knowledge – a distinction argued for intuitively by Barwise in the 1980s:

Theorem Common knowledge defined through countable iteration as truth in all finitely reachable worlds is strictly weaker than common knowledge defined as a greatest fixed-point for the above equation $C_G\, \phi \leftrightarrow \phi\, \&\, E_G\, C_G\, \phi$.

Moreover, the standard 'product topology' on topological spaces provides a form of group knowledge stronger than both, in line with what has been called 'having a shared situation'.

There is a general issue of extending the theory of this book from relational models to topological semantics. This makes sense especially since some semantic models proposed for belief revision have a topological flavour, often in the form of slightly more abstract 'neighbourhood models'. Frankly, we are not even sure of the best definitions for general update (Chapters 4, 6) in this generalized setting, so we must leave this open here.