Chapter 13 LOGICAL DYNAMICS IN PHILOSOPHY

Logical dynamics is a way of doing logic, but it is also a general stance. Making actions, events, and procedures first-class citizens enriches the ways in which logic interacts with philosophy, and it provides a fresh look at many traditional themes. Looking at logic and philosophy over the last century, key topics like quantification, knowledge or conditionality have had a natural evolution that went back and forth across disciplines. They often moved from philosophy to linguistics, mathematics, computer science, or economics – and sometimes back to philosophy (cf. van Benthem 2007E). In the same free spirit, this book has studied the logical dynamics of intelligent agency drawing on several disciplines, and the final chapter sequence explores some interfaces in more detail. We start with philosophy, again disregarding boundaries, moving freely between epistemology, philosophy of information, philosophy of action, or philosophy of science.

This chapter discusses a number of philosophical themes that emerge on a dynamic stance, and tries to convey how fresh, but also how radical logical dynamics can be in its way of dealing with existing issues. [1] We start with a small pilot study on the Fitch Paradox of knowability. Then we move to general epistemology, articulating a view of knowledge as dynamic robustness under various informational processes. This raises a further issue, now in the philosophy of information, namely, what information is, even within the restricted compass of logic. We find several legitimate and complementary notions whose interplay is crucial to rational agency. This leads to the issue of natural mixtures of different notions of information and actions that manipulate them, and we cross over to the philosophy of mathematics, studying intuitionistic logic with dynamic-epistemic tools as a mixture of factual and 'procedural' information. Still, intuitionistic logic and indeed most information and process-oriented logics leave the agents doing all these things out of the picture. We go on to a basic issue in the philosophy of action, and ask ourselves what a rational agent really is, proposing a list of core features on an analogy with the analysis of computation in Turing machines. Having done that, we consider whether the resulting logical dynamics of common sense agency also applies to the more exalted domain of science, discussing some

[1] The reader may want to follow the main line in a first pass, skipping the digressions.

issues in the philosophy of science in this light. Finally, we raise the question what all this means for logic itself. We compare logical dynamics with logical pluralism, a recent position in the philosophy of logic that reflects on what modern logic has become. We emphasize the characteristic features of the dynamics in this book, and what these entail for our understanding of what the field of logic is, or might become.

These sections are based on the various papers listed in the literature section at the end. The subject of logical dynamics and philosophy clearly deserves a larger book of its own – but hopefully, this chapter will whet the reader's appetite.

13.1 Verificationism and the paradox of the knower

Thinking with the mind-set of this book casts many philosophical issues in a new light. It will not necessarily solve them, but it may shift them in interesting ways. We start with a small case study in epistemology, tackling broader themes in the next sections.

Verificationism and the Fitch paradox What is known is evidently true. Now verification-oriented theories of meaning assign truth only to propositions for which we have evidence. But doing that suggests the claim of Verificationism that *what is true can be known*:

$$\varphi \to \Diamond K\varphi \qquad VT$$

Here the K is a knowledge modality of some kind, while the $\Diamond$ is a modality "can" of feasibility. Now, a surprising argument by Fitch trivializes *VT* on quite general grounds:

Fact The Verificationist Thesis is inconsistent.

Proof Take the following Moore-style substitution instance for the formula φ:

$$q \wedge \neg Kq \to \Diamond K(q \wedge \neg Kq)$$

Then we have the following chain of three conditionals:

$$\Diamond K(q \wedge \neg Kq) \to \Diamond(Kq \wedge K\neg Kq) \to \Diamond(Kq \wedge \neg Kq) \to \Diamond\bot \to \bot$$

Thus, even in a very weak logic, a contradiction follows from the assumption $q \wedge \neg Kq$, and we have shown that q implies Kq, making truth and knowledge equivalent. ■

How disturbing is this? Some paradoxes are just spats on the Apple of Knowledge. But others are signs of deep rot, and cuts are needed to restore consistency. Remedies for the

Fitch Paradox fall into two kinds (Brogaard and Salerno 2002, van Benthem 2004B). Some solutions weaken the logic in the above proof. This is like tuning down the volume on your radio so as not to hear the bad news. You will not hear much good news either. Other remedies leave the logic untouched, but weaken the verificationist principle. This is like censoring the news: you hear things loud and clear, but they may not be so interesting. Some choice between these options is inevitable. But one really wants a non-ad-hoc *new viewpoint* with benefits elsewhere. In our view, its locus is not the Fitch proof as such, but our understanding of the two key modalities involved, either the K or the $\Diamond$, or both.

Acts of coming to know Fitch's substitution instance uses an epistemic Moore-formula $q \wedge \neg Kq$ ("q, but you don't know it") that can be true, but not known, since $K(q \wedge \neg Kq)$ is inconsistent in epistemic logic (cf. Chapter 2). Some truths are fragile, while knowledge is robust. One approach to the paradox weakens *VT* as follows (Tennant 2002):

Truth is knowable only for propositions φ *with* $K\varphi$ *consistent* *CK*

But *CK* provides no exciting account of feasibility. We have put our finger in the dike, and that is all. Indeed, the principle suggests a missing link: we need an informational scenario. We have φ true in an epistemic model $\boldsymbol{M}$ with actual world s, standing for our current information. Consistency of $K\varphi$ gives truth of $K\varphi$ in some possibly quite different model $\boldsymbol{N}, t$. The real issue is then the dynamic sense of the feasibility modality $\Diamond$:

What natural step of *coming to know* would take us from $\boldsymbol{M}, s$ to $\boldsymbol{N}, t$?

$\Diamond$ is sometimes unpacked in terms of *evidence* for φ. We will briefly discuss this view in Section 13.3. But for now, we strike out in an alternative semantic direction:

Announcement dynamics: from paradox to normality We have encountered Moore sentences in this book as true but self-refuting assertions, and so the Fitch Paradox recalls our findings in Chapters 3 and following. In this dynamic setting, *VT* becomes:

A formula φ *that is true may* come to be *known* VT^{dyn}

In terms of public announcement logic *PAL*, this says that there is a true statement or correct observation that changes the current epistemic model $\boldsymbol{M}, s$ to a submodel $\boldsymbol{N}, s$ where φ is known. But we already saw that announcing φ *itself,* though an obvious candidate, need not always work for this purpose. [2] What is the import of this analogy? Dynamic epistemic logic provides definite answers. First, the simple knowledge-generating law $[!\varphi]K\varphi$ or $[!\varphi]C_G\varphi$ is only valid *factual* statements φ. With epistemic operators present, we found that self-refutation may occur. Still, this was not paradoxical, but quite useful. The Muddy Children puzzle (Chapter 3) got solved precisely because in the last round, uttering a true ignorance statement makes it false, and children learn their status. In Chapter 15, a similar scenario inspires game solution procedures. [3] [4]

Dynamic typology These observations do not suggest a ban on self-refuting epistemic assertions – the usual remedy to the Fitch Paradox. They rather call for a typology of *which assertions* are self-fulfilling, becoming common knowledge upon announcement – or self-refuting, making their negation common knowledge. This was the Learning Problem in Chapter 3, and it was just a start. In Chapters 9, 15, we also studied statements that induce common knowledge of themselves or their negations when repeated to a first fixed-point.

---

[2] Going back to the earlier proposal, VT^{dyn} implies *CK,* but here is a counter-example to the converse. The formula $\varphi = (q \wedge \neg Kq) \vee K\neg q$ has $K((q \wedge \neg Kq) \vee K\neg q)$ consistent. (The latter holds in a one-world model with $\neg q$.) Now take an epistemic *S5*-model with just a q- and a $\neg q$-world. Here φ holds in the actual world, but no truthful announcement would ever make us learn that φ. The only relevant update is to a one-world model with q: but there, $K((q \wedge \neg Kq) \vee K\neg q)$ fails.

[3] As a further illustration in a similar spirit, Gerbrandy 2005 gives a *PAL* analysis of the *Surprise Exam*. A teacher says the exam will take place on a day next week where the student will not expect it – which flings in the face of a simple Backward Induction argument over weekdays that there can be no such surprise. Gerbrandy dissolves the perplexity by showing how the teacher's assertion can be true but self-refuting. For instance, with two days, it says (with E_i for 'the exam is on day i'): $(E_1 \wedge \neg K_{you} E_1) \vee (E_2 \wedge [!\neg E_1]\neg K_{you}E_2)$. Simple models for *PAL* then clarify variations on the surprise exam. This seems an original track in a well-trodden area, though still ignoring the intuitive self-reference in the teacher's assertion. Baltag & Smets 2010 refine the analysis in terms of iterated soft announcements that change agents' plausibility orderings on worlds (cf. Chapter 7). The iteration limit is a fixed-point model that gets closer to a self-referential effect, creating a final plausibility pattern for the students where beliefs match the intended surprise.

[4] As another example, Bonnay & Égré 2007 use dynamic epistemic techniques coupled with their epistemic 'centering semantics' to analyze Williamson's Margin of Error Paradox.

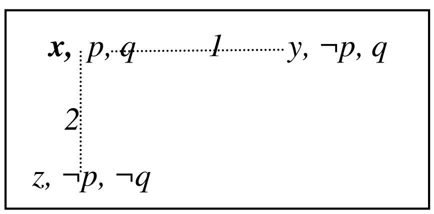
The Liar Paradox in Kripke's Theory of Truth changed a problem into a gateway to a wonderland of truth predicates reached by approximation procedures. Likewise (though for different reasons), in dynamic epistemic logic, the Fitch Paradox is not a problem, but an entry point for an exciting area of epistemic effects of a wide array of statements. [5]

More generally, dynamic epistemic techniques can help provide a typology of paradoxes. For instance, the above features of true but self-refuting assertions, and iteration of announcements to reach stable fixed-points seem common patterns.

Many agents and communication The Fitch Paradox takes on new aspects in social settings with more than one agent. In Chapters 3, 12 we studied communication scenarios that produce common knowledge. These also loosen up the situation here:

Example Fitch with communication.

Consider this model ***M*** with actual world *z* and a group of agents *{1, 2}:*



1's saying q makes *2* know that $p \wedge \neg K_1 p$, which cannot be common knowledge. But the important fact $p \wedge q$ does become common knowledge, even if *2* says that $p \wedge \neg K_1 p$. ■

We can now have social variants of *VT*, distinguishing between roles of agents, the more common situation in knowledge gathering, where agents can learn facts from, and about others. "If φ is true, then someone could come to know it" changes the game, witness the consistency of $K_1(q \wedge \neg K_2 q)$. Verificationism might need a take on what we ask of *others*.

---

[5] Van Benthem 2004B defines three notions of $\varphi \to \exists A <A!>K\varphi$ *(Local Learnability),* $\exists A: \varphi \to <A!>K\varphi$ *(Uniform Learnability),* and $\varphi \to <!\varphi>K\varphi$ *(Autodidactic Learning)*. He shows that each successive type is stronger than the preceding. In *S5*, all three notions are decidable, but their general theory is open, especially, when generalized to *DEL*-style product update (Chapter 4).

This concludes our first dynamic take on an existing philosophical issue. Admittedly, the Fitch Paradox does not go away, Verificationism has not been confirmed or refuted – but all the usual issues are seen in a new light, and with a new thrust. [6]

13.2 Knowledge and dynamic robustness

The preceding case study leads to the general issue how our logics relate to epistemology. We will now look at this in more generality, focusing on the notion of knowledge.

Worlds apart? In modern surveys, the agenda of epistemology has little to do with logic (cf. Kim and Sosa, eds., 2000). Still, epistemic logic started with a view to philosophy, in the book *Knowledge and Belief* (Hintikka 1962). Formulas like $K_i\varphi$ for "the agent *i* knows that φ" and $B_i\varphi$ for "*i* believes that φ" provided logical forms for philosophical arguments. And the semantics for this language (cf. Chapter 2) was an appealing way of representing what agents know in a world, namely, those propositions that are true in all situations that the agents consider compatible with that world. This was no formal trick, but a philosophical use of the legitimate notion of *information* as a range of alternatives that are still open (Carnap & Bar-Hillel 1953; cf. Chapters 1, 2). Moreover, the modal properties of knowledge validated in this way have been widely discussed. One is

$$K_i(\varphi \rightarrow \psi) \rightarrow (K_i\varphi \rightarrow K_i\psi) \qquad \textit{Distribution Axiom}$$

Read as a law of omniscience saying that knowledge is closed under known entailments (in particular, logical consequences), this has sparked controversy. Thus, right or wrong, epistemic logic has served at least to focus philosophical debate on important issues. Of course, in this book, we have rejected the usual terms of this debate. Distribution obviously holds for implicit knowledge $K_i\varphi$, and it equally obviously fails for explicit knowledge $K_i^+\varphi$ that includes syntactic awareness. There is little more to be said at this static level. But Chapter 5 then went further by noting how, in a dynamic epistemic perspective, the real issue is rather *which epistemic actions a* produce *explicit knowledge*. Thus, the debate should not be about 'omniscience', but about ways of validating the implication

[6] This book also suggests alternative takes. Chapters 3 and 11 had temporal versions of *PAL*, with a past operator *Y* for the previous stage. Then $\varphi \rightarrow [!\varphi]C_GY\varphi$ is valid unrestrictedly, yielding a true Verificationist thesis "Every truth can come to be known *as having been true*".

$$K_i\,(\varphi \rightarrow \psi) \rightarrow (K_i\varphi \rightarrow [a]\,K_i\psi)$$

Another example of debate fueled by epistemic axioms is the implication

$$K_i\,\varphi \rightarrow K_i\,K_i\varphi \qquad \textit{Introspection Axiom}$$

that raises the issue whether introspection into epistemic states is plausible. Analogous points apply here, and Chapter 5 gave introspection a similar dynamic twist.

Still, in philosophy, formal notations might just be the last vestiges of a passion long gone – and the role of epistemic logic has clearly diminished: Dretske 1981 ignores it, Barwise & Perry 1983 opposes it, and Willamson 2001 seems neutral at best. But in a remarkable intellectual migration in the 1970s, epistemic logic moved to other disciplines such as economics and computer science. Inspired by this, in the logical dynamics of this book, its agenda has then shifted to the study of information update, communication, and interaction among agents, from humans to machines. Could there now be a return to philosophy? We will illustrate how this might happen in the heartland of epistemology.

Definitions of knowledge Consider the fundamental epistemological issue of what knowledge really *is*. A good starting point is still Plato's Formula

knowledge = justified true belief.

From a logical viewpoint, this intertwines knowledge with *other epistemic attitudes*, viz. belief, while it also highlights *evidence*: sources of knowledge and their certification. Both are major issues in their own right, that have played throughout this book.

But the 20th century has produced further innovative views of knowledge. Hintikka's take, mentioned above, was *truth throughout a space of possibilities* (called the 'forcing view' in Hendricks 2005), grounding knowledge in semantic information that occurs widely in the study of natural language, but also in much of science. By contrast, Dretske 1981 favoured information theory, defining knowledge as *belief based on reliable correlations* supporting information flow. And another major idea is the 'truth tracking' of Nozick 1981, which gives knowledge of P a counterfactual aspect. In a simplified version, knowledge is

true belief in P, while, if P had not been the case, I would have believed $\neg P$.

On the latter account, knowledge becomes intertwined, not only with static beliefs, but also with dynamic actions of belief *revision* underlying the counterfactual.

Clearly, these accounts are richer than standard epistemic logic: the philosophers are ahead of the logicians in terms of imagination. But also, these accounts are still formal, involving connections to belief, evidence, information, or counterfactuals, the very topics modern logicians are interested in. Thus, the distance seems accidental rather than essential. [7]

The right repertoire: epistemic attitudes and actions Our book adds several strands to these developments. Looking at knowledge from our dynamic perspective involves two major methodological points. A first key theme in the logical, computational, and also the psychological literature on agency is this:

> *Step back and consider which notions belong together in cognitive functioning*.

These include knowledge, belief, conditionals, and perhaps even intentions and desires (Cohen & Levesque 1990, Wooldridge 2002). The philosophical point is this. It cannot be taken for granted that we can explain knowledge per se without tackling the proper cluster of attitudes at the same time – a point also made in Lenzen 1980. [8]

A second basic insight guiding this book is a Tandem Principle from computer science:

> *Never study static notions without studying the dynamic processes that use them.*

And the two methodological recommendations are connected. The right repertoire of cognitive attitudes will unfold only when we study the right repertoire of epistemic actions. This viewpoint also makes sense in our ordinary use of the term knowledge. Someone only

---

[7] For instance, Nozick's Formula $K_i\varphi \leftrightarrow \varphi \wedge B_i\varphi \wedge (\neg\varphi \Rightarrow B_i\neg\varphi)$ is a logical challenge. Its adoption blocks standard laws of epistemic logic, such as Distribution or Introspection. Are there any inference patterns left? Given some plausible background logic of belief and counterfactuals, what is the complete set of validities of Nozick's *K*? Arlo Costa & Pacuit 2006 has a modal formulation with neighborhood models, while Kelly 2002 proposes a learning-theoretic account.

[8] An additional insight from Chapters 2, 11 was that this natural task may be hard. The *complexity* of combined modal logics can go up dramatically from their components, e.g., with epistemic temporal agents having Perfect Recall. Combination is not a simple matter of adding up.

knows φ if she displays dynamic expert behaviour. She should have learnt φ via reliable procedures, but she should also be able to repeat the trick: learn other things related to φ, use φ in new settings, and be able to communicate her knowledge to others.

Knowledge, true belief, and dynamic robustness Rephrasing the above, here is our view on the third ingredient that makes true belief into knowledge. The striking characteristic of knowledge is not some static relationship between a proposition, an agent and the world, but sustained acts of learning and communication. The quality of what we *have* resides largely in what we *do* – individually, and socially with others. When saying "I see that φ", I refer to an act of observation or comprehension; when asking a question, I tap into the knowledge of others, and so on with learning, grasping, questioning, or inferring.

The pattern behind all this diversity, also in the above-mentioned philosophical literature, is the *robustness* or *stability of knowledge as informational events occur*. [9] In science, robustness can only be explained if you also give an explicit account of the transformations that can change and disturb a system. In this spirit, the logical dynamics of this book provides a rich account of the informational events that affect knowledge and its relatives. We have seen this at work with the Fitch Paradox and Verificationism. But many more themes in the preceding chapters apply to epistemological issues. In particular, the Tandem View as pursued in Chapter 7 identified three epistemic notions in events of information: belief, knowledge, and a new epistemic attitude intermediate between the two: *stable belief* under learning true facts. Also, the epistemic temporal logic in Chapters 9, 11 of actions over time turned out to fit well with learning theory (Kelly 1996, Hendricks 2002). Finally, much of our work on multi-agent interaction and group knowledge (Chapters 10, 12) fits well with social trends in modern epistemology (cf. Goldman 1999). [10]

Thus, dynamic epistemic logic presents a much richer picture than older epistemic logics, in terms of interactive events making dynamics and information major themes. We can see

---

[9] Related ideas may be found in Hendricks 2005, as well as Roush 2006 on 'truth tracking'.

[10] Many further logical patterns in epistemology are found in Baltag, van Benthem & Smets 2010, that classifies most proposed notions of knowledge in terms of dynamic actions behind them: from observation to contrary-to-fact variation, identifying the sort of stability required.

a lot more in terms of the new techniques, giving earlier issues new and deeper twists. Things have changed, and contacts between logic and philosophy deserve a new try.

Coda: logic and evidence Before discussing further repercussions of our dynamic stance, we briefly return to Plato's Formula. The latter also contains the notion of justification, or *evidence* that underlies our knowledge. Justification has been explained in many ways, such as proof or information correlation. [11] Evidence seems a natural epistemic notion. That is why van Benthem 1993 proposed a merge of semantics-based epistemic logic with a calculus of evidence. And nowadays there is a candidate for such a merge. Logical proof systems often provide explicit binary type-theoretic assertions of the form

x is a proof for φ. [12]

One such calculus extends the provability reading for modal $\Box\varphi$ as saying that φ has a proof. This existential quantifier is unpacked in the Justification Logic of Artemov 1994, 2005 that includes explicit operations of combination (#), choice *(+)* and checking *(!)* on proofs. Then, epistemic axioms can be indexed for the evidence supporting them:

$[x]\, K(\varphi \to \psi) \wedge [y]\, K\varphi \;\; \to \;\; [x\#y]\, K\psi$ Explicit Omniscience

$[x]\, K\varphi \;\to [!x]\, KK\varphi$ Explicit Introspection

[13] This is an interesting way to go, but the relation to the logical dynamics of this book is not obvious. Renne 2008 is a first attempt at combining the frameworks.

---

[11] Note a logical shift. Epistemic logic revolves around a *universal* quantifier: $K_i\varphi$ says that φ is true in all situations agent *i* considers as candidates for the current world. But the justification quantifier is *existential*: it says that there exists evidence. Now, co-existence of $\forall$ and $\exists$ views is not unheard of in logic. The semantic notion of validity says that a formula φ is true in all models. The syntactic notion says there exists a proof for φ. And Gödel's *completeness theorem* established a harmony: a first-order formula φ is valid if and only if it has a proof. Still, Plato's Formula does not suggest an equivalence, but an additional requirement with bite.

[12] In the labeled deductive systems of Gabbay 1996, the *x* can even be any sort of evidence.

[13] Similar forms of indexing work for epistemological issues such as the Skeptical Argument: *"I know that I have two hands. I know that, if I have two hands, I am not a brain in a vat. So (?): I*

More immediate from the viewpoint of this book are two other approaches. The first would just identify explicit evidence with the information-producing events of *DEL* (Chapter 4). Another way of combining semantic information with evidence is found in Chapter 5. This was the merge of public announcements with dynamic logic of syntactic awareness-raising steps (acts of inference, introspection, 'paying attention'). Even so, a systematic logical treatment of evidence remains a challenge outside the scope of this chapter.

13.3 Varieties of logical information

We have explained knowledge in terms of dynamic robustness under informational actions. But then the issue comes up *what notions of information* fuel this logical dynamics. There are clear differences, say, between the semantic models of epistemic logic and the syntactic information flow found in proofs, as we have seen in Chapters 1, 2, 5. Thus, information dynamics for knowledge may come in different kinds, more semantic and more syntactic. This diversity is an important topic by itself, and we will now discuss it in more detail. Of course, at the same time, this is a basic issue in the philosophy of information.

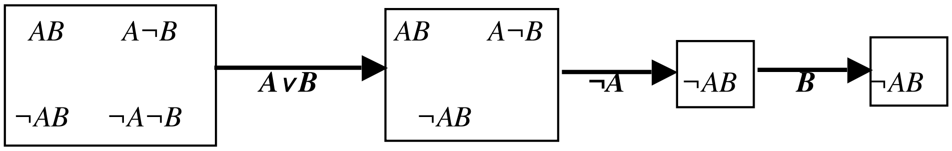
Information is a notion of wide use and intuitive appeal. Different paradigms claim it, from Shannon channel theory to Kolmogorov complexity, witness the Handbook Adriaans & van Benthem, eds., 2008. Information is also a widely used term in logic – but a similar diversity reigns: there are competing accounts, ranging from semantic to syntactic. [14] We will look at the earlier ones as well as others, to get a more complete picture.

Information as range The first logical notion of information is semantic, associated with possible worlds, and we call it *information as range*. This is the main notion in this book, studied in epistemic logic in tandem with the dynamic processes transforming ranges. A concrete illustration were the successive updates for learning first $A \vee B$ and then $\neg A$, starting from an initial situation where all 4 propositional valuations are still possible:

---

know that I am not a brain in a vat." This is again modal distribution, and it might be analyzed as requiring context management: $K_i(\varphi \rightarrow \psi) \wedge [c'] K_i\varphi \rightarrow [c \# c'] K_i\psi$.

[14] Many logicians feel that this diversity is significant. We do not need this notion in the mechanics or the foundations of the formal theory. As Laplace once said to Napoleon, who inquired into the absence of God in his *Mécanique Céleste*: "Sire, je n'avais pas besoin de cette hypothèse".



This is the sense in which valid conclusions ‘add no information’ to the premises.

Information as correlation But there are other semantic features of information. A second major aspect is that information is *about* something relevant to us, and so it turns on connections between different situations: my own, and others. *Information as correlation* has been developed in situation theory, starting from a theory of meaning in information-rich physical environments (Barwise & Perry 1983) and evolving to a theory of distributed systems whose parts show dependencies via channels (Barwise & Seligman 1995).

Correlation ties in with inference. One recurrent example in Indian logic runs as follows (Staal 1988). I am standing at the foot of the mountain, and cannot see what is going on there. But I can observe in my current situation. Then, one useful inference is this:

> "*I see smoke right here. Seeing smoke here indicates fire on the mountain.*
> *So, there is a fire on the mountain top.*" [15]

The Aristotelean syllogism is about one situation – while now, inference crosses over. Given the right channels, observations about one situation give reliable information about another. [16] Incidentally, on some of these views, inference seems a last resort when other processes have failed. If I can see for myself what is happening, that suffices. But if no direct observation is possible, we resort to reasoning. [17] Again, we see the entanglement of different informational processes that guides this book.

---

[15] This is almost the running example in Barwise & Seligman 1995 on seeing a flash-light on the mountain showing that there is a person in distress to some observer safely in the valley.

[16] Barwise & van Benthem 1999 develop the theory of *entailment across a relation* between models, including generalized interpolation theorems that allow for transfer of information.

[17] A very Indian traditional example is that of a coiled object in a dark room – using logic, rather than touch, to find out if it is a piece of rope or a cobra.

Information as code But inference also involves a third major logical sense of information, oriented toward syntax. This is the sense of *information as code* in which valid conclusions do add information to the premises. Its paradigmatic dynamic process is proof or computation, with acts of what may be called elucidation that manipulate representations. Chapter 5 developed systems for this with proofs adding new formulas to our stock of explicit knowledge. For instance, the earlier inference step from $A \vee B$ and $\neg A$ to B, though adding no semantic information to the final semantic information state $\{w\}$, did increase explicit knowledge about the single final world w by adding an explicit property B:

from $w, \{A \vee B, \neg A\}$, we went to $w, \{A \vee B, \neg A, B\}$

But insightful representations can also be more graphical. A concrete illustration in Chapter 1 were the stages in the solution of a 3x3 'Sudokoid':

$$\begin{array}{|ccc|} \hline 1 & \bullet & \bullet \\ \bullet & \bullet & 2 \\ \bullet & \bullet & \bullet \\ \hline \end{array} \quad , \quad \begin{array}{|ccc|} \hline 1 & \bullet & 3 \\ \bullet & \bullet & 2 \\ \bullet & \bullet & \bullet \\ \hline \end{array} \quad , \quad \begin{array}{|ccc|} \hline 1 & 2 & 3 \\ \bullet & \bullet & 2 \\ \bullet & \bullet & \bullet \\ \hline \end{array} \quad \ldots$$

Each successive diagram displays a bit more information about the eventual solution.

Diversity and co-existence Summing up, our first claim is that there are three different major notions of information in logic, each with its own dynamic process. [18]

Naturally, this diversity invites comparison as to compatibility:

Case study: co-existence of range and constraint logics We will not look at all possible merges between the three notions, but here is a brief case study on the relation between information as range and information as correlation. We first formulate the latter view in a congenial logical framework, for greater ease of comparison:

Example Modal constraint models.

Consider two situations s_1, s_2, where s_1 can have some proposition letter p either true or false, and s_2 a proposition letter q. There are four possible configurations: (a) s_1: p, s_2:q, (b) s_1: p, s_2: $\neg q$, (c) s_1: $\neg p$, s_2: q, and (d) s_1: $\neg p$, s_2: $\neg q$. With all these present, no situation

---

[18] Later on, we will propose a fourth notion of *procedural information* arising from the total process agents are engaged in, over and above single informational steps (cf. Chapters 3, 11).

carries information about another, as p and q do not correlate in any way. A significant constraint on the system arises only when we *leave out* some possible configurations. For instance, let the system have just the two states (a) $s_1: p, s_2: q$, (d) $s_1: \neg p, s_2: \neg q$. Now, the truth value of p in s_1 determines that of q in s_2, and vice versa: the constraint $s_1: p \leftrightarrow s_2: q$ holds. This motivates general constraint models $\boldsymbol{M} = (Sit, State, \boldsymbol{C}, Pred)$ with a set *Sit* of situations, a set *State* of valuations, a predicate *Pred* recording which atoms hold where, and a constraint C stating which assignments of states to situations can occur. [19] ■

These models support an obvious modal language for reasoning with constraints, with names x for situations ($\boldsymbol{x}$ denotes a tuple), and atoms $P\boldsymbol{x}$ for predicates of situations. Next, there is a shift relation $\boldsymbol{s} \sim_x \boldsymbol{t}$ *iff* $\boldsymbol{s}(x) = \boldsymbol{t}(x)$ *for all* $x \in \boldsymbol{x}$, that lifts to tuples of situations $\boldsymbol{x}$ by having equality of $\boldsymbol{s}$ and $\boldsymbol{t}$ for all coordinates in $\boldsymbol{x}$. Thus, there are modalities $[]_{\boldsymbol{x}}\varphi$ for each such tuple, which say that the situations in $\boldsymbol{x}$ settle the truth of φ in the current system:

$$\boldsymbol{M}, \boldsymbol{s} \models []_{\boldsymbol{x}}\varphi \quad \text{iff} \quad \boldsymbol{M}, \boldsymbol{t} \models \varphi \text{ for each global state } \boldsymbol{t} \sim_{\boldsymbol{x}} \boldsymbol{s}$$ [20]

This extended modal constraint language has a decidable complete logic for reasoning about constraints and informational correlation.

Correlation and dependence There is more here than meets the eye. Modal constraint logic resembles the decidable first-order logic of *dependent variables* in van Benthem 1996, with modal-style single and polyadic quantifiers as well as substitution operators. Van Benthem 2005C proves a precise equivalence, making the above correlation view of information and the idea of logical dependence the same topic in different guises.

Adding epistemic structure Adding epistemic structure is natural here. A blinking dot on my radar screen is correlated with an airplane approaching. But unless I *know* that there is a blinking dot, the correlation will not help me. That knowledge arises from an event: my observing the screen. This suggests an *epistemic constraint language* for models $\boldsymbol{M} = (Sit, State, \boldsymbol{C}, Pred, \sim_i)$ that combine correlation and range. E.g., suppose that $\boldsymbol{M}$ satisfies the

[19] Constraint models are like the ‘context models’ in Ghidini & Giunchiglia 2001, and they also resemble the ‘local state models’ of interpreted systems in Fagin et al. 1995.

[20] There is an analogy here with *distributed knowledge* for groups of agents, cf. Chapter 2.

constraint $s_1{:}p \rightarrow s_2{:}q$. Then the agent knows this, as the implication is true in all worlds in ***M***. Now suppose the agent knows that $s_1{:}p$. In that case, she also knows that $s_2{:}q$:

$$(K\, s_1{:}p \wedge K\,(s_1{:}p \rightarrow s_2{:}q)) \rightarrow K\, s_2{:}q$$

Dynamically in *PAL*, if the agent *learns* that $s_1{:}\, p$, she would also know that $s_2{:}q$:

$$[!\, s_1{:}p]\, K\, s_2{:}q.$$

Thus, range and correlation views of information co-exist in obvious ways. This merge is natural in many fields, from philosophy to physics and economics. [21] [22] [23]

Inferential information and realization On the path of comparison, our next task is to draw inferential information into the same circle of ideas. But that is precisely what was done in Chapter 5 of this book, at both static and dynamic levels. The upshot of the analysis was two-fold. Models for inferential information flow resemble those of dynamic-epistemic logic, endowing worlds with syntactic access to, or awareness of propositions. And the update mechanism works for a much wider variety of actions than inference, including acts of introspection – all falling under the common denominator of converting *implicit* into *explicit* knowledge. [24] For instance, the axiom $K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$ for semantic information told us that $K\psi$ is implicit knowledge. But as we hinted at in the above, the real issue of interest is another one. To those willing to do some epistemic work, the premises also *licensed* an act of realization producing explicit knowledge:

$$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow [\#\psi]Ex\psi)$$

In particular, an explicit system like this can help disentangle the notions of information that are run together in some philosophical discussions of omniscience.

---

[21] The co-existence extends even further. Like *DEL*, situation theory involves event scenarios that ‘harness’ information, such as the Mousetrap of Israel & Perry 1990. These scenarios suggest *dynamic constraint models* $\boldsymbol{M} = (Sit, State, \boldsymbol{C}, Pred, Event)$ supporting standard modal logics.

[22] Similar models with epistemic information, correlations, and informational events occur in Baltag & Smets 2007 on the structure of quantum information states and measurement actions.

[23] Sadzik 2009 connects epistemic logic to correlated equilibria in game theory.

[24] Van Benthem & Martinez 2008 survey many further approaches to inferential information.

Peaceful co-existence or unification? We have demonstrated how the different notions of information identified here live together fruitfully in precise logical settings. Still, peaceful co-existence is not Grand Unification. Several abstract frameworks claim to have unified all notions of logical information. In particular, algebraic and relevant logics have been put forward for this purpose (cf. Dunn 1991, Mares 1996, Restall 2000), while van Benthem 1991 used a categorial Lambek Calculus with dynamic and informational interpretations. Sequoiah-Grayson 2007 is a philosophical defense of the non-associative Lambek calculus as a core system of information structure and information flow. [25]

While all this is appealing, axioms in such systems tend to encode minimal properties of mathematical adjunctions, and these are so ubiquitous that they can hardly be seen as a substantial theory of information (cf. van Benthem 2010B). No Grand Unification has occurred yet that we are aware of. And maybe we do not need one:

Conclusion We have identified three major logical notions of information. We have also shown that they are compatible, though we doubt that a grand unification between the three views is possible, or even desirable. They are probably better viewed as *complementary stances* on a rich and complex phenomenon. In particular, we see these stances as a way of sharpening up epistemological definitions of knowledge in terms of three different kinds of information, though we will not pursue this reanalysis here.

Finally, we have not closed the book yet on information diversity in logic. In the following section, in line with the global temporal process perspective of Chapters 4, 11, we will even propose a fourth stance of ‘procedural information’.

13.5 Intuitionistic logic as information theory

Given the entanglement of different notions of information, let us see how this functions in practice. As a concrete test for our analysis, including the combination of observational and access dynamics, we go to the philosophy of mathematics. We bring our dynamic logics to bear on a much older system that has long been connected with both proof and semantic information, viz. *intuitionistic logic*, a famous alternative to epistemic logic. Our aim is to see whether our framework applies, and what new things come to light.

---

[25] Abramsky 2008 is relevant, too, proposing a category-theoretic framework for information.

From proof to semantics Intuitionistic logic had its origins in the analysis of constructive mathematical proof, with logical constants acquiring their meanings in natural deduction rules via the Brouwer-Heyting-Kolmogorov interpretation (Dummett 1977, van Dalen 2002). It then picked up algebraic and topological semantic models in the 1930s. In the 1950s, Beth proposed models over trees of finite or infinite sequences, and in line with the proof idea, intuitionistic formulas are true at a node when 'verified' there in some strong intuitive sense. A later simplified version of this semantics, due to Kripke, uses *partially ordered* models $\boldsymbol{M} = (W, \leq, V)$ with a valuation V, setting:

$\boldsymbol{M}, s \models p$	iff	$s \in V(p)$
$\boldsymbol{M}, s \models \varphi \wedge \psi$	iff	$\boldsymbol{M}, s \models \varphi$ *and* $\boldsymbol{M}, s \models \psi$
$\boldsymbol{M}, s \models \varphi \vee \psi$	iff	$\boldsymbol{M}, s \models \varphi$ *or* $\boldsymbol{M}, s \models \psi$
$\boldsymbol{M}, s \models \varphi \rightarrow \psi$	iff	*for all* $t \geq s$, *if* $\boldsymbol{M}, t \models \varphi$, *then* $\boldsymbol{M}, t \models \psi$
$\boldsymbol{M}, s \models \neg\varphi$	iff	*for no* $t \geq s$, $\boldsymbol{M}, t \models \varphi$

In line with the idea of accumulating certainty, the valuation is *persistent*:

if $\boldsymbol{M}, s \models p$, and $s \leq t$, then also $\boldsymbol{M}, t \models p$.

The truth definition lifts the persistence to all formulas. In particular, a negation says the formula itself will never become true at any further stage. This makes Excluded Middle $p \vee \neg p$ invalid, as this fails at states where p is not yet verified, though it will later become so. This may happen in several ways: see the black dots in the pictures below, that stand for the start of informational processes unfolding as downward trees:

Implicit versus explicit epistemics Note how information or knowledge enter here, not in the explicit format of epistemic logic, but through re-interpretation of the classical logical constants. Van Benthem 1993 calls this 'meaning loading': new phenomena are absorbed implicitly into an old semantics, rather than getting separate explicit treatment. This is both intriguing, and a source of deep difficulty in comparing with other approaches.

Interpreting the models What notion of information is represented by these models? Intuitively, each branching tree describes an *informational process* where an agent learns progressively about the state of the actual world, encoded in a propositional valuation. At endpoints of the tree, all information is in, and the agent knows the actual world. Thus, these models seem a sort of alternative to the epistemic models we have had so far. [26]

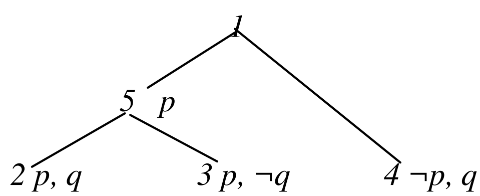
Procedural information One first striking point is that intuitionistic models immediately extend our analysis of information so far, by registering *two* notions: [27]

(a) *factual information* about how the world is; but on a par with this:

(b) *procedural information* about our current investigative process.

How we get our knowledge matters deeply, and while leaves record factual information, the branching structure of our trees itself, with available and missing intermediate points, encodes agents' knowledge of the latter kind. The distinction between factual and procedural information makes sense much more widely, and we saw it already in Chapters 3, 11 that extended dynamic epistemic logic with protocols stating which histories are admissible in the investigative process. Now we can draw a comparison:

From intuitionistic to epistemic information How can we model intuitionistic scenarios in dynamic-epistemic logic? For an illustration, consider the following tree:

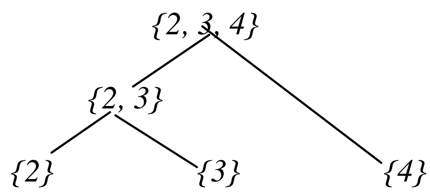


Epistemic logic casts knowledge in terms of worlds representing ways the actual situation might be. At stages of the tree, the obvious candidates are the endpoints below, or the *complete histories* of the process. Thus, we can assign epistemic models as follows:

---

[26] Even so, the technical perspectives of Chapter 2 for epistemic logic fully apply in this setting. For instance, modal bisimulation induces a similar invariance between intuitionistic models.

[27] The following point, and others to be made here also apply to modal extensions of intuitionistic logic that allow non-persistent statements in the investigative process (van Benthem 1989).



One way of seeing this is as a family of epistemic models that decrease over time.

An analogy: protocols and games This picture is reminiscent of the epistemic analysis of games in Chapter 10. Indeed, with some give and take, [28] intuitionistic trees convert into the *PAL* protocol models of Chapter 11, relating them to the perspective of this book.

But here is a difference with the game scenario: moves are now just anonymous upward inclusion links. So, in the spirit of dynamification, *what are the underlying actions?*

Acts of announcement The first type of action that we see in intuitionistic tree models are *public announcements*. If each endpoint is uniquely definable, each step shrinking the set of reachable endpoints announces the disjunction of all definitions for endpoints that remain. Thus, intuitionistic logic describes effects of observing facts – though without making these acts of observation explicit.

Acts of explicit seeing Equally intriguing from our perspective, however, is a second type of action, that comes to light in our initial failures of the Double Negation law:

The second model $\boldsymbol{M_2}$ poses no problems. An epistemic-temporal protocol with just the announcement actions *!p* and *!¬p* splits the two branches as required.

---

[28] In a game of perfect information, players knew where they were at a node in the game tree, but they did not know what the future history will be. In Section 10.8, we converted global uncertainty about the future into local uncertainty about the present, by taking worlds to be pairs *(h, s)* with *h* any history passing through the tree node *s*. This yields an epistemic protocol model.

But now consider the first model M_1. A natural *PAL* version would put the set *{s}* at both nodes, as nothing is ruled out. But then no information flows, and, if we think of *p* as a property of the endpoint, we already knew that *p* at the start. [29] But in intuitionistic models, actually putting *p* at a stage means more than just its inevitability in the latter sense. What event is taking place then? It is tempting to interpret this as an event of the sort that we studied in Chapter 5, an act *#p* of awareness raising, or *explicit seeing*.

Thus, intuitionistic logic combines all notions that we have studied, but in new ways.

Conclusion In our perspective, intuitionistic logic as an account of information flow is akin to both semantic dynamic-epistemic logic of observation and syntactic logics of explicit seeing and awareness, merging events of public observation with private acts of realization. [30] Moreover, its semantics in terms of processes of investigation emphasizes the importance of procedural information as a further basic category. Of course, many aspects of intuitionistic logic are not explained in this way, such as its accounts of constructive proof and definition. But our analysis does show that intuitionistic logic has a plausible interpretation as a theory of information-driven agency, something that fits its multi-agent game-theoretic interpretation via dialogues (Lorenzen 1955). Thus the logical dynamics of this book might sit well with constructive logic and mathematics. [31]

Open problem: implicit versus explicit approaches once more Still there remains a major conceptual issue that we have not been able to resolve. There is a clear methodological contrast between implicit logical approaches like intuitionism that 'load meanings' of classical constants and explicit approaches like epistemic logic with new vocabulary for knowledge on a classical base. The contrast occurs far beyond this single case study. We also found it with our explicit dynamic logics of questions in Chapter 6 versus 'inquisitive semantics' that provides propositional logic with a new semantics without explicit syntax

---

[29] Knowledge in this sense matches the intuitionistic operator $\neg\neg$ or its modal counterpart $\Box\Diamond$, referring to eventual truth in all reachable end-points.

[30] This view has many repercussions. Van Benthem 2009E shows how the usual reduction of intuitionistic logic to *S4* is misunderstood as a move to a weaker epistemic logic for *K*, whereas in fact, it mixes factual and procedural information in a new *epistemic-temporal* modality.

[31] Intuitionistic logic should then also be a natural fit with other aspects of agency, such as belief.

for question operators. And this is again just an instance of a general contrast between our dynamic logics and *dynamic semantics* for natural language, where meanings of the basic propositional operators and quantifiers change from static to dynamic ones.

How the two approaches compare in general is something we leave for another study.

13.5 What are rational agents?

Our topics so far were mainly concerned with knowledge, information, and action. But one major topic has been left out of consideration, namely, *who performs* all these activities. What is a rational agent, and what tasks have we set for her as theorists of intelligent interaction? Perhaps surprisingly, this question seldom gets asked in the burgeoning literature on agency. Chapter 1 has stated the main features of rational agency that motivated the logical research program in this book. But I really think of the underlying issues as deeper, requiring philosophical analysis far beyond what I present in this book. I will discuss a few major features, broadly following the program outlined in Chapter 1:

Classical foundations To see the point, compare the foundations of mathematics, starting with the logics of Frege, Russell, and others. Hilbert's Program provided appealing *goals*: establish the consistency of formalized mathematics, and where possible completeness, with a logic that is simple, perhaps decidable. This was a program with panache! The 1930s then discovered its infeasibility with Gödel's incompleteness theorems and the undecidability results of Church and Turing for natural computational and logical problems. Foundational research made exciting refutable claims, and their refutation had positive spin-off. Like Vergilius' Romans after the fall of Troy, the logicians retreating from Paradise founded an empire of recursion theory, proof theory, and model theory, thanks to insights from that turbulent period. In particular, the Universal Turing Machine from the foundational era is still our general model of computation, being a lucid analysis of key features of a human doing sums with pencil and paper.

Now, if the hero of our time is the Generic Rational Agent, what are its key skills and properties, clearly far beyond doing sums? I have asked many colleagues: answers were lively, but diverse. I list some issues that I myself find important:

Prelude: idealized or bounded processing powers? But first, we need to set an appropriate level of idealization. The dynamic logics in this book mostly endow agents with unlimited inferential and observational powers, and memory to store their fruits. But I am also attracted by the opposite tendency in cognitive science, stressing the limitations on these powers that human cognition operates under. In that case, the heart of rationality would be optimal performance given heart-breaking constraints. [32] Now, this book does provide tools for this: *DEL* describes limited observational access, its inferential versions describe limited inferential power, and we also had possible variations on memory. Moreover, we should not confuse (as is sometimes done) the idealization in a logical system with the nature of the agents that it describes: one can have a perfect logic of memory-less agents (cf. Chapter 11) or under-informed players of a game (Chapter 10). Still, we have no systematic account of bounds on agents and their variation.

Which core tasks? But next, powers to what end? A Turing Machine must just compute. Do rational agents have a core business? Is it *reasoning* – as a normal form for all other intelligent activities? Reasoning is important, especially when taken in a broad sense. [33] But other crucial abilities such as acumen in perception or other information sources do not reduce to reasoning in any illuminating way. [34] My first key feature is this:

Rational agents intelligently take and combine information from many sources.

Revision and learning But I do not think that informational soundness: being right all the time, is the main hall-mark of rational agents. Their peak performances are rather in spotting problems, and trying to solve them. Rationality is constant *self-correction*. This reflects my general take on the foundational collapse in the 1930s. The most interesting issue in science is not guarantees for consistency and safe foundations, but the dynamic ability of repairing theories, and coming up with creative responses to challenges. Thus belief revision and learning are the true tests of rationality, rather than flawless update. Chapter 7 showed how our logics can do justice to this second major feature of agency:

Rational agents crucially correct themselves and respond intelligently to errors.

---

[32] Gigerenzer & Todd 1999 gives surprising examples of optimal behaviour even then.

[33] Colleagues responding to my request for a core list mentioned making predictions, but even more, a backward-looking talent of explaining and *rationalizing* what has already happened.

[34] Again, the emphasis on reasoning may confuse object-level agency and meta-level theory.

Evaluation, goals and commitments We have also seen in Chapter 1, 9 how rational agency is not just blind information processing. It crucially involves a balance between information and *evaluation*. This led to our study of a second stream of preference dynamics behind the flow of information, but this was just a start. There is much more to goals, intentions, and commitments that truly 'make sense' of behaviour. [35]

Rational agents act in balance between their information, preferences, and goals.

Communication, interaction, and groups But even rich single-agent views are still too restricted. The core phenomenon we are after is intelligent interaction. A truly intelligent agent can perform tasks directed toward others: ask the right questions, explain things, convince, persuade, understand strategic behaviour, synchronize beliefs and preferences with other agents, and so on. Almost circularly, I state my final desideratum:

A rational agent is someone who interacts rationally with other agents!

This social perspective showed throughout this book, in logics of mutual knowledge, communication and games. It has many features that transcend specific logical systems. One is a crucial skill that glues us together: the ability to put yourself in someone else's place. In its bleakest form, this is a role switch in a game. More concretely, it is the ability to see social scenarios through other people's eyes, as in the Categorical Imperative: "Treat others as you would wish to be treated by them."

Another social talent is the typical ability of humans to form new entities, viz. intelligent groups with lives of their own. Our identity is made up of layers of belonging to groups, as acknowledged in our logics of common knowledge and group structure. And finally,

---

[35] Frankly, the computational literature leads the way in logical studies of action and commitment. Cohen & Levesque 1990, Rao & Georgeff 1991 are major sources for the *BDI* framework ('beliefs, desires, intentions') that still needs to be connected in a serious way to the logical dynamics in this book. Meyer, van der Hoek & van Linder 1999 study commitment with tools from dynamic logic. Dunin-Keplicz and Verbrugge 2002 add connections with cooperative problem solving. Roy 2008 uses ideas from *DEL* to show (partly in cooperation with van Hees) how intentions can facilitate making decisions and even solving games. Closer to the *BDI* framework, Icard, Pacuit & Shoham 2009 develop a dynamic logic of commitment. Shoham 2009 analyzes logical theories of intention from a more concrete computational perspective, linking many strands.

agents are not all the same, and they form groups whose members have diverse abilities, strategies, and so on. Making room for this diversity is a key task in logic and philosophy, as we have seen before. Successful behaviour means functioning well in an environment of agents with different capacities and habits.

Summary I have now identified the four basic features that I consider as constitutive of rational agency. The systems in this book show that these features are not just slogans, since they can be analyzed precisely and brought together in one logical framework.

Even so, all this does not add up to one crisp universal model of rational agency. The field of intelligent interaction is still waiting for its Turing. [36] Still, I do think that these foundational questions should be asked, much more than they have been so far.

But even then, a difference remains with the foundations of mathematics:

'So what': are there refutable claims and goals? Suppose we find one model of agency, what is the agenda providing focus and thrill? Could there be an analogue to Hilbert's Program, setting a worthy heroic goal for interactive logicians to march toward – and fail? Instead of even suggesting an answer, I end with a more modest perspective:

Integrating trends A field may also form around a shared modus operandi. In particular, this book has shown trends toward framework integration between dynamic epistemic logic and game theory (Chapter 10), epistemic temporal logic (Chapter 11), probability theory (Chapter 8) and other areas. An interesting analogy is again with the foundational era. The 1930s saw competing paradigms for defining computation. But gradually it became clear that, at a well-chosen level of input-output behaviour, these all captured the same computable functions. *Church's Thesis* then proclaimed the unity of the field, saying that all approaches described the same notion of computability – despite intensional differences making one or the other better for particular applications. This led to a common field of Recursion Theory, everyone got a place in the joint history, and internal sniping

[36] Interestingly, the *Turing Test* in AI is already closer to what I am after. An ability to recognize other agents as machines or humans is a typical more delicate social rational skill.

was replaced by external vigour. Something similar might happen in the study of intelligent interaction. Failing a Hilbert or Turing, we might at least have a Church. [37]

13.6 Agency in science and mathematics

Even in the absence of a daring refutable program for rational agency, we can start with what we have. In particular, the dynamic stance of this book may have something to say about the foundations of science, not just the daily world of common sense activities where most of our motivating examples have come from. Let us look briefly at the philosophy of science to confirm this. In fact, logical dynamics forms a very nice fit:

Common sense and scientific agency First, perhaps, one should remove a barrier. Classics like Nagel 1961 assiduously enumerate all the differences between the two worlds of science and common sense, the habitat of rational agency as studied in this book. I cannot discuss these claims in detail (cf. van Benthem 2010D for details, also in what follows), but let me just state my view. Science is the exercise of certain qualities of common sense agency, but taken further in isolation, and also simplified, since some strategic features of intelligent interaction are put out of play. [38] Still, science is clearly a social process – indeed, about the oldest successful one, predating all established empires and religions. And one finds respectable social aspects in science wherever one looks: even traditional logical formalizations emphasized its 'intersubjectivity', a social feature. Thus, there is no barrier, and much in this book immediately applies to our understanding of science.

Diversity of information sources: observation on a par with inference As we have seen from the start, agents manipulate many sources of information, observation, inference, and communication. And this moves us with one great stride from traditional mathematics-centered approaches to the reality of the experimental sciences where observation is a

---

[37] Of course, one further unifying force across the area is the *empirical reality* of intelligent interaction, and hence an independent sanity check for whatever theory we come up with.

[38] The very term 'research' sounds dynamic. The heart of science seems to lie in its modus operandi: the processes that generate its products, not a museum of certified theories.

fundamental source of information. [39] Indeed, the hard part may be rather explaining the rationale of mathematical proof. Even though this, too, plays an essential role in science, it is less easy to say just how. In the philosophy of science, this has been a persistent problem: cf. Fitelson 2006 on explaining the informativeness of Einstein's deductions from the General Theory of Relativity. This problem of information gain through inference was our major concern in Chapter 5, and in the above discussion of information.

Theories, attitudes, and informational actions over time Very typical for our logics of agency was the wide spectrum of attitudes that agents can have toward information, ranging from knowledge and belief to neutral 'entertainment', or even doubt. I feel that all of these are not alien to scientific research, but at the very heart of it. Accordingly, we also need a much richer view of scientific theories in terms of epistemic attitudes. This would start with beliefs, since, as Popper pointed out, *belief revision* as an engine of learning is as essential to science as peaceful accumulation and regurgitation of knowledge. Continuing with this additional structure, there is also the *research agenda* as an object in its own right. What philosophers have said about that fits well with our interest in questions and issues (Chapter 6). A theory in science, and even in mathematics, is not just the set of its propositions (answers to past questions), but also its current agenda of open problems. Finally, science is a *long-term process* with features that only emerge in the long run. Likewise, our dynamic logics led to epistemic temporal logics of agency that interface with formal learning theory, a branch of the philosophy of science (cf. Kelly 1996).

The others: social aspects revisited Our third source of information was communication, making conversation and argumentation key topics in logic. But science, too, essentially involves many agents: as we said, it is one of the oldest social activities in history. Various authors have cast experiments in terms of games played by Man and Nature (Giles 1974, Hintikka 1973, Lorenz & Lorenzen 1978, Mittelstaedt 1978). But also, interaction between human agents seems essential to science: for instance, its styles of debate are a major engine of progress. In this book, social agency led to contacts with game theory (Chapter 10). In the philosophy of science, interfaces are often rather with evolutionary games

[39] Newton's *Principia Mathematica* has mainly mathematical axioms, but read his *Optics*, and you will see that experiments, i.e., the voice of Nature, are treated with equal respect.

(Skyrms 1990), but the two can meet. And there is more to the social aspect of science than mere interaction of individuals. Scientific theories are usually community creations, and their development is a collective group activity. This matches our step toward logics of how groups form and evolve, along with their beliefs and actions (Chapter 12). [40]

Science and values Finally, alongside with the dynamics of information flow, there was a second major cognitive system permeating rational action, viz. the dynamics of evaluation. Rational action strives for a balance between information and desire. Now, science is often said to be value-neutral, and only about information. Is that so? Or are we missing crucial aspects of the scientific activity by ignoring its goals, perhaps changing over time?

Conclusion The analogies pointed out here amply justify taking a fresh look at the formal philosophy of science in terms of logical dynamics. I submit that much more of what really makes science tick will come to light that way. [41]

13.7 What is logic?

From the philosophy of science, it is only one small self-referential turn to the philosophy of logic. Our final topic is the nature of logic itself. As we have said before, textbooks in the philosophy of logic tend to address issues reflecting the state of the art of 50 years ago – and sometimes, they use definitions of the field that were out of date even then, ignoring definability and computability in favour of proof theory. But, where is logic heading?

---

[40] Logical dynamics might also overcome the divide between formal approaches in the philosophy of science and historical or sociological views that are supposed to be anti-logical. In our view, the latter rightly stress the importance of revision next to accumulation, and the pervasive role of agency. But as we have seen, these are not a threat to logic, but rather an enhancement.

[41] Our discussion may have suggested a certain antagonism between logical dynamics and the role of *mathematics* in thinking about science. This is by no means the case. As we have emphasized in Chapter 1, logical dynamics uses mathematics as its methodology. But also, it can analyze the mathematical activity *itself*. Van Benthem 2010C is a case study, looking at mathematical theories from all the angles discussed here. In particular, it finds new challenges to our logics beyond the dynamics of inference addressed in Chapter 5. One is the treatment of beliefs in a deductive setting, where the models of Chapter 7 seem inadequate. The other is the crucial role of *definitions* in mathematics, and hence the logical dynamics of language selection and language change.

There is a common feeling that it has been broadening its scope and agenda beyond classical foundational issues, and maybe a concern that, like Stephen Leacock's famous horseman, it is 'riding off madly in all directions'. What is the resultant vector?

Dynamics versus pluralism Two broad answers are in circulation today that try to do justice to the real state of the field. One is Logical Pluralism, locating the new scope of the discipline in charting a wide variety of reasoning styles, often marked by non-classical substructural rules of inference. This sees the essence of logic as describing consequence relations, a view to which I subscribed in my work on substructural logics around 1990. But I have changed my mind since then. The Logical Dynamics of this book says that the main issue is not reasoning styles, but the *variety of informational tasks* performed by intelligent agents, of which inference is only one among many, including observation, memory, questions and answers, dialogue, and strategic interaction. Logical systems should deal with all of these, making information-carrying events first-class citizens.

In this final section, I will contrast the two views. I argue that logical dynamics sets itself the more ambitious goal, including an explanation of *why* substructural phenomena occur, by deconstructing them into classical logic plus an explicit account of the underlying informational events. I see this as a more challenging agenda for logic, and much richer fare for philosophers of logic. However, if I am right in my depiction, this also contains the seeds for a creative co-existence of the two views, and I will show how.

Styles of reasoning Let us start with the basics of the pluralist view, which starts from an appealing abstraction level that uncovers patterns across many different logical systems. Classical consequence $\boldsymbol{P} \Rightarrow C$ from a finite sequence of premises $\boldsymbol{P}$ to a conclusion C says that C is true in every situation where all the propositions in $\boldsymbol{P}$ are true. This relation between premises and conclusions satisfies a number of *structural rules*:

if $\boldsymbol{P}, Q, R, \boldsymbol{S} \Rightarrow C$, *then* $\boldsymbol{P}, R, Q, \boldsymbol{S} \Rightarrow C$	*Permutation*
if $\boldsymbol{P}, Q, Q \Rightarrow C$, *then* $\boldsymbol{P}, Q \Rightarrow C$	*Contraction*
$C \Rightarrow C$	*Reflexivity*
if $\boldsymbol{P} \Rightarrow Q$ *and* $\boldsymbol{P}, Q \Rightarrow C$, *then* $\boldsymbol{P} \Rightarrow C$	*Cut*
if $\boldsymbol{P} \Rightarrow C$, *then* $\boldsymbol{P}, Q \Rightarrow C$	*Monotonicity*

Together, these laws encode the basic style of reasoning behind classical consequence. It treats the data that feed into a conclusion as sets (order and multiplicity do not matter), the inferential relation is a pre-order allowing for chaining of conclusions, and overkill does not matter: accumulating more data is not going to endanger earlier conclusions.

The Bolzano Program The 1970s and 1980s saw a wave of different reasoning styles. Relevant logic dropped monotonicity, default logics dropped monotonicity and transitivity, resource logics in categorial grammar and linear logic dropped contraction, and so on. A general theory developed in the work of Gabbay 1996, Dunn 1991, Restall 2000, while Dosen & Schroeder-Heister 1993 coined the term 'sub-structural logics'. Van Benthem 1989 noted an analogy with Bolzano's *Wissenschaftslehre* (1837) that did not focus on logical constants, but on charting properties of different reasoning styles: deductive or probabilistic, in the common sense, or according to philosophical standards. [42]

An analysis of consequence in terms of structural rules has an attractive Spartan austerity. Ties to a richer semantic picture are provided by *representation theorems* (van Benthem 1991, 1996) that are close to techniques in algebraic logic (Venema 2006). Here is a simple example, showing the bare bones of classical consequence:

Theorem A consequence relation $\boldsymbol{P} \Rightarrow C$ satisfies the above five structural rules iff

it can be represented by a map sending propositions P to sets $Set(P)$ with

$P_1, \ldots, P_n \Rightarrow C$ iff $\bigcap_{1 \leq i \leq n} Set(P_i) \subseteq Set(C)$.

Proof The proof is simply by setting $Set(B) =_{def} \{\boldsymbol{A} \mid \boldsymbol{A} \Rightarrow B$ in the given relation$\}$, and then checking that the given equivalence holds by an appeal to all given structural rules. ■

More sophisticated representation theorems work for further notions of consequence. Hence, the structural analysis of consequence captures essences of reasoning, while still being able to reintroduce the original practice when needed.

Two worries, and the move to logical dynamics Even so, in Benthem 1989, I voiced two concerns. First, it seemed that structural rules address mere *symptoms* of some underlying

[42] The term I proposed for this enterprise: the *Bolzano Program*, has never caught on, though this Austrian-Italian pioneer continues to exert an appeal to logicians (van Benthem 2003A).

phenomenon. Non-monotonicity is like fever: it does not tell you which disease causes it. Thus, I was missing a deeper analysis of underlying phenomena. Matching this was a second worry. Substructural logics give up properties of classical consequence, but they retain the old formal language. Why not be radical *with respect to the language* as well, and reconsider what we want to say? [43] All this is what the logical dynamics of this book is about. We take logic to be a study of all sorts of informational processes in new languages, not just pluralist consequence relations, and design appropriate languages. [44]

The difference in approach may be seen with the logic *PAL* of public announcements in Chapter 3. We did not just look at new 'epistemic consequence relations', but made it the job of logic to explicitly describe any events *!P* that change agents' information states.

Here is another illustration of the difference: *nonmonotonic logics*, a showpiece of pluralism, since they form a stable practice beyond classical consequence.

Non-monotonic consequence or dynamic logic of belief change Classical consequence from ***P*** to *C* says that all models of ***P*** are models for *C*. But McCarthy 1980 showed that problem solving goes beyond this, by zooming in on the 'most congenial' models. A *circumscriptive* consequence from ***P*** to *C* says that *C* is true in all *minimal models* for ***P***. Abstract description levels here include conditional logic (Shoham 1988, Gabbay 1996) and belief revision theory (Gärdenfors 1988). Again there are representation theorems.

In contrast with this, Chapter 7 represented circumscription in terms of acts of *belief revision*, providing explicit logical systems for agents' belief change under events of hard information *!P*, and also under various sorts of soft information $\Uparrow P$. Again, the underlying informational acts have become first-class citizens of the system.

---

[43] Admittedly, this happened with linear logic, and to some extent with relevant logic. But, it has not happened with circumscription and default logics, and we will return to that issue below.

[44] An exclusive concentration on consequence does not even fit the classical realities in the 1950s, when logic already had model theory and recursion theory in addition to proof theory as its main branches, placing computation and expressive power on a par with proof and consequence.

A compromise: back-and-forth between dynamics and consequence But we cannot break free from our personalities for long. Mine is analogy- and consensus-seeking, and let us see how this can be done for the case of Pluralism and Dynamics.

Dynamic consequence for knowledge. A first comparison occurred in Chapter 3. The Restaurant of Chapter 1 had a notion of consequence that fits with public announcement. One first processes the information in the premises, and then checks the conclusion:

Definition A sequent $P_1, \ldots, P_k \Rightarrow \varphi$ is *dynamically valid* if, starting with any epistemic model $\boldsymbol{M}, s$ whatsoever, successive announcements of the premises result in an updated model where announcement of φ has no further effect: in the model $(\ldots(\boldsymbol{M}|P_1)\ldots)|P_k, s,$ the formula φ was already true everywhere, even before it was announced. ■

Dynamic consequence is expressed by a *PAL* validity making the conclusion *common knowledge*: $[!P_1]\ldots[!P_k]\, C_G\varphi$ (#). For factual formulas $P_1, \ldots, P_k, \varphi$, this holds iff φ follows classically from $P_1, \ldots, P_k$. The reason is that factual formulas do not change their truth values at worlds when passing from $\boldsymbol{M}$ to $\boldsymbol{M}|P$. This is not true in general, however – and using Moore-type statements, all classical structural rules failed for dynamic validity. But still, *modified structural rules* remain valid, and we do get a new substructural logic:

Fact Dynamic consequence satisfies the following structural rules:

if $\boldsymbol{P} \Rightarrow C$, *then* $A, \boldsymbol{P} \Rightarrow C$	*Left-Monotonicity*
if $\boldsymbol{P} \Rightarrow A$ *and* $\boldsymbol{P}, A, \boldsymbol{Q} \Rightarrow C$, *then* $\boldsymbol{P}, \boldsymbol{Q} \Rightarrow C$	*Left-Cut*
if $\boldsymbol{P} \Rightarrow A$ *and* $\boldsymbol{P}, \boldsymbol{Q} \Rightarrow C$, *then* $\boldsymbol{P}, A, \boldsymbol{Q} \Rightarrow C$	*Cautious Monotonicity* [45]

[45] At the abstraction level of structural rules, we can view propositions A as *partial functions* T_A taking input states to output states. *Transition models* $\boldsymbol{M} = (S, \{T_A\}_{A\in Prop})$ consist of states in S with a family of transition relations T_A for each proposition A. Here, a sequence of propositions $\boldsymbol{P} = P_1, \ldots, P_k$ *dynamically implies* conclusion C in $\boldsymbol{M}$, if any sequence of premise updates starting anywhere in $\boldsymbol{M}$ ends in a fixed point for the conclusion: *if* $s_1\, T_{P1}\, s_2 \ldots T_{Pk}\, s_{k+1}$, *then* $s_{k+1}\, C\, s_{k+1}$. We then say that *sequent* $P_1, \ldots, P_k \Rightarrow C$ *is true* in the model: $\boldsymbol{M} \models P_1, \ldots, P_k \Rightarrow C$. Van Benthem 1996 proves this representation theorem: A sequent σ is derivable from a set of sequents Σ by these three rules iff σ is true in all abstract transition models where all sequents in Σ are true.

This abstract substructural analysis extracts the gist of dynamic inference. And this is not just a way of reaching out to consequence-oriented Pluralism, but a provable fact. Call a *super-sequent* $\Sigma \Rightarrow \sigma$ from a set of sequents Σ to sequent σ *update-valid* if all substitution instances with epistemic formulas (reading sequents as the above type (#) formulas) yield valid *PAL*-implications. Then we have (cf. van Benthem 2003A for details):

Theorem The update-valid structural inferences $\Sigma \Rightarrow \sigma$ are precisely those whose conclusions σ are derivable from their premise sets Σ by the rules of Left-Monotonicity, Left-Cut, and Cautious Monotonicity.

Thus, we have shown how dynamic logics lead to an interesting abstraction level of dynamic consequence relations and its structural rules. [46]

Dynamic consequence for belief. The same ideas returned in Chapter 7. The logic of belief under hard information, too, supports a matching consequence relation, saying that processing the premises results in belief (or common belief when relevant):

$$[!P_1] \ldots [!P_k] B\varphi$$

The latter is the dynamic counterpart to consequence relations like circumscription, though determining its precise substructural rules is an open problem. [47]

But the dynamic setting also suggests *new consequence relations*, whose trigger is *soft* rather than hard information. The lexicographic upgrade $\Uparrow P$ changed the current ordering $\leq$ of worlds as follows: all P-worlds in the current model become better than all $\neg P$-worlds,

---

[46] The leap from *PAL* to a sequent-style analysis is drastic. Another natural abstraction level is a slightly richer *poly-modal language* over transition models. Dynamic validity needs two basic modalities, universal modal boxes for the premise transitions, and a loop modality for fixed-points: $\boldsymbol{M}, s \models (a)\varphi$ iff $R_a ss$ and $\boldsymbol{M}, s \models \varphi$. The modal loop language is decidable and axiomatizable (van Benthem 1996). Reading dynamic sequents $P_1, \ldots, P_k \Rightarrow C$ as modal formulas $[P_1]\ldots[P_k](C)T$, all earlier structural rules become derivable, as well as other properties, and the representation result can be extended. Thus, modal logic itself seems a natural abstract theory of dynamic inference.

[47] This is because intuitions about circumscription often concern exclusively factual statements, whereas dynamic logics think of consequence between all formulas in the language.

while, within those two zones, the old plausibility ordering remains. Now we get a new notion of dynamic consequence, more attuned to generic belief revision:

$$[\Uparrow P_1] \dots [\Uparrow P_k]\, B\varphi$$

This should be grist to the mill of pluralist logicians, since structural rules for this notion are unknown. Further plausibility changes may even suggest more consequence relations.

Thus, in a dynamic setting, both classical reasoning and non-monotonic default reasoning are at heart about epistemic attitudes and responses to information, from hard update to soft plausibility change. Merging things in this way fits with the earlier general conception of agency: processes of inference and self-correction go hand in hand. [48]

Two directions Abstract consequence relations and dynamic logics relate as follows:

Dynamics	➔ *abstraction* ➔	Consequence
Consequence	➔ *representation* ➔	Dynamics

From dynamics to consequence, the issue is to find abstraction levels capturing significant properties of consequence relations generated by the dynamic activity in the logic. From consequence to dynamics, one looks for representation theorems exhibiting the dynamic practice behind a given consequence relation. The two directions live in harmony, and we can perform a Gestalt Switch one way or the other. The two directions might also be used to describe historical periods. The avant-garde tendency in applied areas in the 1980s was toward abstraction, and maybe that of the current decade more towards concretization.

Conclusion so far We have contrasted two programs for legitimizing the diversity of modern logic. Logical Pluralism emphasizes consequence as the locus of research, and finding natural ways to parametrize it. Logical Dynamics emphasizes informational events, from inference to observation, and the processes by which rational agents harness these in order to act. We have shown how to move back and forth between the two perspectives by abstraction and representation, and sometimes get connecting results.

But I am not neutral, and prefer the Logical Dynamics program for the following reasons:

---

[48] It would be nice to do dynamic reconstructions of other reasoning styles, such as *abduction*.

Dynamics is more classical There is more to be learnt from the preceding case studies. We saw how a non-monotonic consequence relation relates to a classical dynamic logic of the process that *causes* the nonmonotonicity. Thus, when we take the logical analysis one step further, and are willing to choose a language matching this,

monotonic dynamic logic can model non-monotonic consequence.

This observation seems quite general [49], though I have no general proof of its scope.

Dynamics is more challenging Even though I started out on the consequence side, I now prefer dynamics because it is more ambitious, and gives logicians much more to do. [50] The logics of agency in this book provide ample evidence for this point.

Once more, what is logic? In this debate among the avant-garde, we may have lost sight of the larger issue that started our discussion:

What is logic?

Many answers are in circulation. Beyond what we have discussed, some people seek the essence in mathematical notions of semantic invariance (Bonnay 2006) or in proof theory (Martin–Löf 1996). I myself have no magic definition to offer. I see no point to essentialist statics of what logic *is*, but rather to the dynamics of what it could *become*. And for that, this book offers a way of broadening the classical agenda of proof and computation toward a wider one of rational agency and intelligent interaction. Of course, there must be some continuity here, or we have just started doing another subject. But I have no specific axioms or invariances that need to be preserved at all cost. Once more in line with the thrust of this book, I see a discipline as a dynamic activity, not as any of its static products:

---

[49] More evidence are the classical nature of *PAL* versus non-classical intuitionistic logic, as well as the analysis of quantum logic in terms of classical *PDL* in Baltag & Smets 2007.

[50] As for challenges, one would want a dynamic take on *paraconsistent* logics in terms of processes that handle inconsistencies, and on *resource-conscious logics* (linear logic, categorial logics) putting management of information pieces explicitly into the logic. Finally, returning to Bolzano after all, who included language as an explicit and crucial parameter in his account of logical consequence, one should find a way to include *language change* into the dynamics.

proofs, formal systems, or languages. Logic is a stance, a modus operandi, and perhaps a way of life. That is wonderful enough.

13.8 Conclusion

This concludes our tour of contacts with philosophy. We have seen how Logical Dynamics connects naturally with issues in epistemology, philosophy of information, philosophy of science, and other areas, throwing fresh light on traditional discussions, and suggesting new ways of viewing and using logic. These bridgeheads need strengthening, and many more things will happen – and all it takes is a sensibility to the dynamic stance.

13.9 Literature

The treatment of the Fitch Paradox is largely based on van Benthem 2004B, 2009A. The view of knowledge as dynamic robustness is in van Benthem 2006A. The discussion of logical information comes from van Benthem 2005C and van Benthem & Martinez 2008. The dynamic-epistemic analysis of intuitionistic logic is in van Benthem 2009E. The general discussion of rational agency follows van Benthem 2008C. Connections with philosophy of mathematics and philosophy of science are taken from van Benthem 2010C, 2010D. The discussion of dynamics and logical pluralism is in van Benthem 2008D. See http://staff.science.uva.nl/~johan/Research for a collection of relevant documents.