Chapter 12 LOGICAL DYNAMICS IN PHILOSOPHY

Logical dynamics is a way of doing logic, but it is definitely also a philosophical stance. Making actions, events, and games first-class citizens of logical theory enriches the ways in which logic interacts with philosophy in general. For a general survey of the interface of logic and philosophy over the last century, avoiding a priori Whig history, cf. van Benthem 2007, which is a narrative of *themes* rather than sub-disciplines, all running in between philosophy, logic, linguistics, computer science, and other disciplines. In this chapter, we will look at a few such themes that emerge once one takes the 'dynamic stance'.

12.1 A first sample: verificationism and the paradox of the knower

Thinking with the mind-set of this book will cast 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 issues in the next sections.

The issue: verificationism incurs the Fitch paradox Verificationism is an account of meaning which says that truth can only be assigned to propositions for which we have evidence. This view is found with logical proof theorists like Dummett and Martin-Löf, but it is also quite influential in philosophy. Stated as a sweeping claim, this take on truth suggests the general verificationist thesis that *what is true can be known*:

$$\phi \to \Diamond K\phi \qquad \textbf{\textit{VT}}$$

Here the K can be taken as a knowledge modality, while the $\Diamond$ is a modality "can" of feasibility in some sense. Now, a surprising argument by Fitch trivializes this principle:

Fact The Verificationist Thesis is inconsistent.

Proof Fitch uses just a weak modal logic to show that ***VT*** collapses the notions of truth and knowledge, by taking the following clever 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:

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

Thus, a contradiction follows from the assumption $q \wedge \neg Kq$, and we have shown over-all that q implies Kq, making truth and knowledge equivalent. ■

Not every paradox is a deep problem. Some are just spats on the Apple of Knowledge, which can be removed with a damp cloth. But others are tell-tale signs of worm rot inside, and surgery is needed to restore consistency – the Apple may not even remain in one piece. Proposed remedies for the Paradox fall into two kinds (Brogaard and Salerno 2002, van Benthem 2004). Some solutions weaken the logic in the 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 strategies is inevitable. But one really wants a *new systematic viewpoint* going beyond plugging holes, and opening up a new line of thinking with benefits elsewhere. In our view, the locus for this is not Fitch' proof as such, but rather our understanding of the two key modalities involved, either the K or the $\Diamond$, or both.

A first analysis: epistemic logic and evidence Fitch's substitution instance uses an old conundrum called *Moore's Paradox*: the statement "*P*, but I don't believe it" can be true, but cannot be consistently believed. Transposed to knowledge, Hintikka 1962 observed the inconsistency of $K(q \ \& \ \neg Kq)$ in epistemic logic. Some truths are fragile, while knowledge is robust: and so truth need not always support knowledge. Thus, one sensible approach to the paradox weakens the scope of applicability of *VT* as follows (Tennant 2002):

Claim *VT* only for propositions ϕ such that $K\phi$ *is consistent* ***CK***

CK provides no exciting account of knowledge K or feasibility $\Diamond$. We have put our finger in the dike, and that is all. Indeed, there is a missing link. We have ϕ true in some epistemic model $\boldsymbol{M}$ with actual world s, representing our current information. But consistency of $K\phi$ gives only truth of $K\phi$ in some possibly *quite different* epistemic model $\boldsymbol{N}, t$. The issue is:

What natural step of 'coming to know' would take us from $(\boldsymbol{M}, s)$ to $(\boldsymbol{N}, t)$?

Indeed ◇ has been unpacked in the proof-theoretic origins of *VT*. Type-theoretic proofs display the *evidence* for a conclusion in assertions $p: \phi$, where the p is a proof for ϕ, or some piece of evidence in a general sense. This is the most sophisticated underpinning of Verificationism to date. [201] But right here, we strike out in a semantic direction.

From Moore sentences to PAL-style dynamics We have seen Moore sentences many times in Chapters 3 and after, as 'self-refuting assertions', and so the Fitch Paradox at once recalls our findings in Chapters 3, 4 and following. In this dynamic setting, *VT* becomes:

What is true may come to be *known* ***VT-dyn***

In terms of our public announcement logic *PAL*, this says the following. Some true public statement, or observation, can be made that changes the current epistemic model $(\boldsymbol{M}, s)$ to a sub-model $(\boldsymbol{N}, s)$ where the formula ϕ is known. But we already know that announcing ϕ *itself,* though an obvious candidate, need not work. For the moment, we just observe a connection with the earlier proposal. Clearly, ***VT-dyn*** implies ***CK***, but the converse fails:

Fact ***CK*** does not imply ***VT-dyn*** for all propositions ϕ.

Proof Here is a counter-example. The formula $\phi = (q \,\&\, \neg Kq) \vee K\neg q$ is knowable in the sense of ***CK***, since $K((q \,\&\, \neg Kq) \vee K\neg q)$ is consistent. The latter formula holds in a model consisting of just one world with $\neg q$. (In S5, the statement $K\phi$ is equivalent to $K\neg q$.) But here is a two-world epistemic *S5*-model $\boldsymbol{M}$ where ϕ holds in the actual world, even though there is no truthful announcement that would ever make us learn that ϕ:

q, the actual world ············ *some other world,* $\neg q$

[201] Van Benthem 1993 took this evidence idea to epistemic logic, and proposed an explicit calculus of evidence for *K*-assertions. One striking modern view of this kind is the 'logic of proofs' of Artemov 1994, 2005, which replaces the box $\Box\phi$ of modal provability logic by operators $[p]\phi$: '*p* is a proof for ϕ'. Labels *p* of many sorts appear in the 'labeled deductive systems' of Gabbay 1996. I consider this new evidence parameter for logical investigation as a deep response to any paradox – even though I am not aware of an inspiring solution to Fitch-style problems in this setting.

In the actual world, $(q \ \& \ \neg Kq) \vee K\neg q$ holds, but it fails in the other one. Hence, $K((q \ \& \ \neg Kq) \vee K\neg q)$ fails in the actual world. Now, there is only one truthful update of this epistemic model $\boldsymbol{M}$ by public announcement which just retains its actual world with q:

q, the actual world

But in this one-world model, the formula $K((q \ \& \ \neg Kq) \vee K\neg q)$ evidently fails. ■

From paradox to normality What is the general import of this analogy? Our dynamic epistemic logics provided some definite answers. First, the law $[!\varphi]K\varphi$ or $[!\varphi]C_G\varphi$ is not valid, but *factual* statements do satisfy it. But with epistemic operators present, we knew that self-refutation may occur – and even be useful. Consider the ignorance statement of the Muddy Children: in the last round of the puzzle, its true announcement makes it false, since then, children learnt their status. And this puzzle again suggests general applications to methods for solving games (Chapter 15). Here is one more illustration:

Example The Surprise Examination.

Gerbrandy 2005 gives a new analysis of the Paradox of the *Surprise Exam*, revolving around a teacher's assertion that the exam will take place on a day where the student does not expect it. He dissolves the usual perplexity by showing how the teacher's assertion can be true but self-refuting. With two days, here is an example (E_i for 'the exam is on day i'):

$$(E_1 \ \& \ \neg K_{you} E_1) \vee (E_2 \ \& \ [!\neg E_1]\neg K_{you}E_2)$$

Simple epistemic models of the *PAL* type clarify various surprise exam scenarios. [202] ■

Dynamic typology These observations do not support a ban on self-refuting assertions – as in most remedies to the Fitch Paradox. They rather call for a *dynamic typology* of epistemic assertions. We can investigate which precise forms of assertion are ‘self-fulfilling’, in that they do become common knowledge upon announcement, or ‘self-refuting’. This was the ‘Learning Problem’ in Chapters 3, and in Chapters 9, 15, where we also study statements

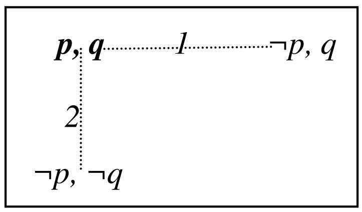
[202] Chapters 3 and 11 did suggest temporal versions of *PAL*, with a past operator Y referring to the previous stage which validates $\varphi \rightarrow [!\varphi]C_GY\varphi$. This is a sense in which the Verificationist thesis *VT* does hold generally: “Every truth *right now* can come to be known *as such* at some later stage”.

that induce common knowledge of themselves or their negations in the long run. The Fitch Paradox is not a problem, but a gateway to an exciting area of study. [203]

Many agents and communication Finally, the Fitch Paradox takes on interesting new aspects with more than one agent, as in Muddy Children. In Chapters 2, 10 we studied scenarios where agents convey truths to each other so as to produce common knowledge.

Example Fitch with communication.

Consider the following model ***M*** with actual world ***p, q***, with a group of agents *{1, 2}:*



Saying q makes *2* know that $p \,\&\, \neg K_1 p$, which cannot be common knowledge. But $p \,\&\, q$ can become common knowledge, when *1* announces that q, and *2* then says that p. ■

Van Benthem 2008 discusses ways in which *VT* may be true or false when more agents are informed – including issues like when what is true about *you* can become known to *me*:

If ϕ is true, then *someone* could come to know it. $VT_{multi\text{-}agent}$

This principle is true in some construals, though things get complex with assertions about the whole group. In this setting the Fitch Paradox meets game theory and learning theory. After all, learning usually involves two roles: a *Student* and a *Teacher*. Verificationism would then also need a take on what we ask of others. The point that *seeking* and *finding* are intertwined in inquiry was made long ago in Hintikka 1973. We leave it to the verificationists to amend their Thesis to that attractive social setting.

---

[203] Van Benthem 2004 has three types of learnability: $\models \varphi \rightarrow \exists A\, \langle A!\rangle K\varphi$ *(Local Learnability),* $\exists A$: $\models \varphi \rightarrow \langle A!\rangle K\varphi$ *(Uniform Learnability),* $\models \varphi \rightarrow \langle !\varphi\rangle K\varphi$ *(Autodidactic Learning)*. He shows that each successive type is stronger than the preceding. In *S5*, all three notions are decidable. One can generalize all these notions to reachability of models via *DEL*-style product update (Chapter 4).

This concludes out first case of a dynamic perspective on an existing philosophical issue. The paradox does not go away, verificationism has not been confirmed or refuted – but all the issues are no seen in a broader light, and perhaps with a new thrust attached to them.

12.2 Knowledge and epistemology

Behind the specific case study of Section 12.1, there is the issue of how epistemic logic in its dynamic variants relates to epistemology. We now look at this in more generality. [204]

Worlds apart? At first sight, modern epistemology has little to do with logic (Klein 1993, Kim and Sosa 2000), Still, epistemic logic started with a view to epistemology, in the book *Knowledge and Belief* (Hintikka 1962, 2005). Formulas like $K_i\phi$ for "the agent *i* knows that ϕ" and $B_i\phi$ for "*i* believes that ϕ" provided logical forms for philosophical arguments. And their semantics (cf. Chapter 2) were an appealing way of thinking about what agents know or believe in a given situation. In particular, an agent knows those propositions which are true in all situations compatible with what she knows about the actual world; i.e., her current range of uncertainty. These models for epistemic logic correspond to a widespread notion of *information* as a range of alternatives that are still open. The laws validated in this way are familiar from modal logic. A typical example is the implication

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

Read as a principle of 'epistemic omniscience' saying that knowledge is closed under known entailments (and in particular, logical consequences), this has sparked controversy until today. Thus, whether positively or negatively, epistemic logic still serves to set patterns of debate. Another example of the same role is the implication

$$K_i \phi \rightarrow K_i K_i\phi \qquad \textit{Introspection Axiom}$$

which highlights the issue whether it is plausible to assume immediate introspection into epistemic states. Still, these notations might just be the last vestiges of a passion long gone – and with a few exceptions, the philosophical role of epistemic logic has diminished: Dretske 1981 ignores it, Barwise & Perry 1983 fights it, and Willamson 2001 seems at best

[204] This section follows the main lines of van Benthem 2006.

neutral. But in line with the intellectual migration history of van Benthem 2007, themes from epistemic logic have moved to other areas, such as economics and computer science. And this, in logical dynamics, the agenda has shifted to the study of information update, communication, and interaction among arbitrary agents, whether humans or machines – resulting in the atmosphere of the present book. Can there be a return to epistemology?

Definitions of knowledge As our running theme, let us take the issue what knowledge really *is*. A good starting point is still Plato's Formula

knowledge = justified true belief.

We step back, and look at this with the eyes of a logician. First of all, the formula intertwines knowledge with *other attitudes*, viz. belief, while it also highlights *evidence*: sources of knowledge and their certification. Both are major issues in their own right, to which we return below. But the 20th century has produced many new views of knowledge. Hintikka's take was *truth throughout the logical space of possibilities*, the modern 'forcing view' (Hendricks 2005). By contrast, a post-Gettier proposal like Dretske 1981 favoured information theory, defining knowledge as *belief based on reliable correlations* supporting information flow. And yet another major idea is the 'truth tracking' of Nozick 1981, who says that knowledge of *P* involves a counterfactual aspect – in one simplified rendering: *true belief in P, while, if P had not been the case, I would have believed ¬P*. On the latter account, intriguingly, knowledge becomes interwined, not only with static beliefs, but also with dynamic actions of belief *revision* underlying the counterfactual.

Clearly, these accounts are richer than that of 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. [205]

---

[205] For instance, Nozick's Formula $K_i\phi \leftrightarrow \phi \,\&\, B_i\phi \,\&\, (\neg\phi \Rightarrow B_i\neg\phi)$ is a logical challenge. Its adoption blocks standard laws of epistemic logic, such as Distribution or Introspection. Are there any valid inference patterns left? Given some plausible logic of belief and counterfactuals, *what is the complete set ofvalidities* of Nozick's *K*? Arlo Costa 2005 has a modal logic formulation in terms

The right repertoire: epistemic attitudes and epistemic actions One key theme in the logical, computational and psychological literature on agency is which notions belong together in successful cognitive functioning. These include knowledge, belief, conditionals, and even intentions and desires (Wooldridge 2002). The philosophical point is that we may not be able to explain knowledge per se without tackling the proper cluster of propositional attitudes at the same time, a point also made in the neglected gem Lenzen 1980. [206]

But which notions? One basic insight in computer science which has also guided this book throughout is a *'Tandem Principle'*: never study static notions without also studying the dynamic processes which give rise to these. Thus, the right repertoire of cognitive attitudes will unfold only when we simultaneously study the repertoire of *epistemic actions*. Indeed, we would only say that someone knows *P* if that person displays further expert behavior having to do with *P*. She should have learnt *P* on the basis of reliable procedures, but she should also be able to repeat the trick: learn other things related to *P*, use *P* in new settings, and very importantly, be able to *communicate* her knowledge to others.

Calculus of evidence Plato's formula also highlights the existence of a justification. This has been explained in many ways: proof, observation, informational correlation, etcetera. No matter how this is taken, note the logical shift. Hintikka-style knowledge revolves around a *universal* quantifier: $K_i\varphi$ says that φ is true in all situations agent *i* considers as candidates for the current situation *s*. But the evidence quantifier is *existential*: it says that there exists *a justification*. Now, co-existence of $\forall$ and $\exists$ views is not unheard of in logic. The semantic notion of logical validity says that a proposition is universally valid: i.e., true on all domains under all interpretations. The syntactic notion says that there exists a proof for the proposition. And Gödel's completeness theorem established a harmony, at least for first-order logic: a formula satisfies the first condition if and only if it satisfies the second.

---

of neighborhood topology, while Kelly 2002 proposes a more recursion-theoretic account in terms of learning theory over a branching temporal universe.

[206] One relevant insight from Chapters 2, 11 was that the *complexity* of combined logics for modal notions can go up dramatically from the components (Spaan 1993), say, when describing epistemic temporal agents with Perfect Recall. Thus, combination is not just a simple matter of 'adding up'.

But Plato's formula does not state an equivalence, but an additional requirement with bite. That is why van Benthem 1993 proposed a merge of epistemic logic with a calculus of evidence, to do its job more properly. Indeed, logical proof theories provide co-existence of knowledge and justification, by manipulating binary type-theoretic assertions of the form

x is a proof for φ. [207]

Another calculus of this sort extends the 'provability interpretation of modal logic', where we a necessity operator $\Box\varphi$ says that there is a proof for φ in some relevant calculus. This existential quantifier is unpacked in the 'logic of proofs' or 'justification logic' of Artemov 1994, 2005, which includes operations of combination (#), choice *(+)* and checking *(!)* on proofs. Then, earlier epistemic axioms get indexed for the evidence supporting them:

$[x]\ K_i(\phi \rightarrow \psi)\ \&\ [y]\ K_i\phi \ \rightarrow\ [x \# y]\ K_i\psi$	Explicit Omniscience
$[x]\ K_i\,\phi \ \rightarrow [!x]\ K_i K_i\phi$	Explicit Introspection

This is an interesting way to go, but it is not the logical dynamics of this book. [208]

Dynamics: bring in the actions! But the main line of this book unpacks 'evidence' in another way, as dynamic events of observation or communication that produce knowledge. Indeed, knowledge and actions producing and transforming it seem on a par. This fits with the common sense observation that the 'quality' of knowledge does not reside in some static relationship between a proposition, an agent and the world, but in sustained behavior of being able to learn and communicate. The quality of what we *have* epistemically resides largely in what we *do* – individually, or socially with others. When I say "I see that φ", I really refer to an act of observation or comprehension; when I ask a question, I tap into the knowledge of others, and so on with learning, grasping, questioning, inferring, and so on.

---

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

[208] Similar forms of indexing may 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 know that I am not a brain in a vat."* This is again modal distribution, and it might be analyzed as requiring 'context management': $[c]\ K_i(\phi \rightarrow \psi)\ \&\ [c']\ K_i\phi \rightarrow [c \# c']\ K_i\psi$. Contexts are a powerful device in linguistics, computer science, and AI (van Benthem & ter Meulen 1997, McCarthy 1993).

Here is one pattern behind all this diversity. Many philosophical views of knowledge try to get at its *robustness* or *stability*. But as in science, robustness can only be explained well if you also include an explicit account of the transformations that potentially disturb a system.

Concretely, we have already seen how this works in a specific case, our analysis of the Fitch Paradox and verificationism. But many more themes in the preceding chapters apply at once to current epistemological issues. In particular, the Tandem View that process analysis works together with design of the right static notions was exemplified in Chapter 6, where we needed three natural notions in scenarios with events of hard information: belief, knowledge, and a new attitude intermediate between these two: *stable belief* under announcements of true facts. Many more examples of logical patterns in epistemology are found in the forthcoming study Baltag, van Benthem & Smets 2008, which classifies most proposed notions of knowledge in terms of dynamic actions behind them: from observation to contrary-to-fact variation, plus the sort of stability required. In addition, the temporal perspective in Chapters 9, 11 fits well with formal epistemology in the guise of *learning theory* (Kelly 1996, Hendricks 2002), as well as the evolution of cognitive practices (Skyrms 1990). Finally, much of what we have said about interaction and group knowledge in preceding chapters fits well with similar trends in modern epistemology.

In all, dynamic epistemic logic captures many notions beyond Hintikka's original one in terms of interactive events, making both *dynamics* and *information* major epistemological categories. Still, this is largely about semantic information – and we will raise the issue of other processes, and other sorts of information later.

12.3 But what is rational agency? [209]

Behind all these special links and topics, I see a broader question, also for this book. What ís a rational agent really, and what task have we set as theorists of intelligent interaction? One cannot consult some standard text for this purpose, because there are none.

Classical foundations To see the point, compare the foundations of mathematics, which started with the formal systems of Frege, Russell, and others. *Hilbert's Program* provided

[209] This section follows the main lines of van Benthem 2009 (Beijing DLMPS).

the first 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! But the foundational discoveries of the 1930s demonstrated its infeasibility, by Gödel's Incompleteness Theorems, and Turing and Church's undecidability results for natural computational and logical problems. Foundational research made exciting refutable claims, and its eventual refutation had positive spin-off. Like Vergilius' Romans after the fall of Troy, logicians founded an empire of recursion theory, proof theory, and model theory, all traceable to that insights from that turbulent period. In particular, the Universal Turing Machine from the same foundational era is still our general model of computation, being a lucid analysis of the key features of a human doing sums with pencil and paper. If our counterpart is the Generic Rational Agent, what are its defining skills and properties, which go far beyond pencil and paper sums? I have asked many colleagues which features they consider constitutive of rationality. Answers were lively, but diverse. I have no conclusive answer, but I will list some issues that I myself find most central.

Idealized or bounded processing powers? The dynamic logics in this book idealize agents, endowing them with unlimited inferential and observational powers, and memory to store their fruits. But I am also attracted by the opposite tendency in the literature, stressing the limitations on all these powers that human cognition operates under. In that case, the heart of rationality would be optimal performance given heart-breaking constraints. Gigerenzer 1999 gives surprising examples of optimal behaviour even then. This provides a 'tension': we need to explain how our logical systems can function in such a setting.

Which core tasks? And then, 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 indeed important, especially when taken in a broad sense. [210] But other crucial abilities such as acumen in perception and observation, and talents for successful interaction, do not reduce to 'reasoning' in any illuminating way.

---

[210] For instance, decision-theoretic views look 'forward' at how agents predict the future, and plan their actions. But colleagues responding to my request for a 'core list' also emphasized a dual 'backward-looking' talent, viz. explaining and *rationalizing* what has already happened.

Revision and learning I do not think that informational 'soundness': being right all the time, is a 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 general *learning* are the true tests of rationality in my view, rather than flawless update.

Communication and interaction But reasoning and learning are still too restricted. They apply to a single agent. But 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 paradoxically, I state this desideratum:

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

Here are some crucial aspects of this social perspective.

Diversity 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 non-trivial task for logics for agents, as we have seen at several places in this book. Successful behaviour means functioning well in a wide-range environment of agents with different capacities and habits.

Switching Here is another social skill which glues us together: *the ability to put yourself in someone else's place*. In its bleakest form, this is the logician's 'role switch' in a game. But in a concrete form, it is the ability to see social scenarios through other people's eyes, as in Kant's Categorical Imperative: "Treat others as you would wish to be treated by them."

Intelligent groups Humans typically form new entities, viz. *groups*, with lives of their own. Our identity is made up of layers of 'belonging' to groups or coalitions, which showed in our logics of common knowledge and group structure (Chapter 9). The formation of *'rational we's'* and intelligent organizations, too, seems crucial to rational agency.

All this does not add up to one universal model of rational agency yet. The field of intelligent interaction is still waiting for its modern Turing. But I do think that these foundational defining questions should be asked, much more than they have been so far.

Are there refutable claims and goals? But suppose we find one model of agency, what is the agenda providing focus and thrill? Could there be an analogue to Hilbert's Program beyond just 'learning more', setting a worthy goal for interactive logicians to march toward? I will not attack this either – but it seems another crucial question well-worth asking. Instead of providing an answer, I end with another perspective on the field: it might find its unity, not through a priori analysis, but through evolutionary convergence of ideas.

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 9), epistemic temporal logic (Chapter 11), and probability theory (Chapter 7). An interesting analogy is again with the foundational era. The 1930s saw many competing paradigms for defining computation. But eventually, it became clear that, at a well-chosen level of input-output behaviour, these all described the same computable functions. *Church's Thesis* then proclaimed the unity of the field, saying all approaches described the same notion of computability – despite 'intensional differences' making one or the other more suitable for particular applications. This led to a common field of Recursion Theory, everyone got a place in the joint history, and internal sniping was replaced by external vigour. Something similar might happen in the study of intelligent interaction. If we do not have a Hilbert or Turing, we might at least have a Church. [211]

12.4 Philosophy of information [212]

Information is a notion of wide use and intuitive appeal, which has been used throughout this book. Different formal paradigms claim part of it, from Shannon channel theory to Kolmogorov complexity, witness the Handbook Adriaans & van Benthem, eds., 2008.

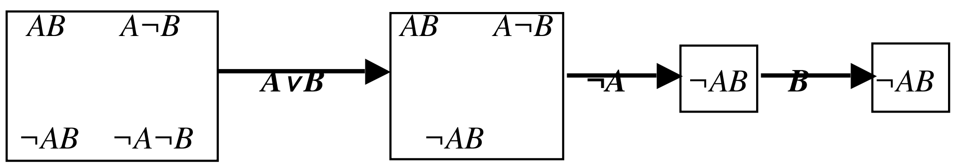
---

[211] 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.

[212] This section is based on parts of van Benthem & Martinez 2008, to which we refer for details.

Information is also a widely used term in logic, even though it has largely remained implicit in the background – but a similar diversity reigns: there are several respectable, but competing logical accounts of this notion, ranging from semantic to syntactic. [213]

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. For a concrete illustration, think of 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:



Information as correlation A second major strand high-lights another semantic feature, viz. that information is *about something* relevant to us, and so it turns on connections between different situations: my own, and others. This notion of *information as correlation* has been developed in situation theory, starting from a theory of meaning in information-rich physical environments (Barwise & Perry 1983), and moving to a view of distributed systems whose parts show dependencies via 'channels' (Barwise & Seligman 1995).

Correlation also ties in with inference. One recurrent 'syllogism' 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 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.*" [214] Compare this with the Aristotelean syllogism, which is about one situation – while now, inference crosses over. Given suitable channels, observations

---

[213] Indeed, 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 Napoléon, who inquired into the absence of God in his *Mécanique Céleste*: "Sire, je n'avais pas besoin de cette hypothèse".

[214] This is almost the running example in Barwise & Seligman 1997 on seeing a flash-light on the mountain suggesting a person in distress there to some observer safely in the valley.

about one situation give reliable information concerning another. [215] Incidentally, on the Indian view, reflected in parts of Western logic, inference is a sort of last resort, when other informational processes have failed. If I can see for myself what is happening in the room, that suffices. If I can ask some reliable person who knows, then that suffices as well. But if no direct or indirect observation is possible, we must resort to reasoning. [216] Again, we see the entanglement of different informational processes driving this book.

Information as code Finally, there is a third major logical sense of information, oriented toward syntax, inference, and computation. It is the sense in which valid conclusions 'add no information' to the premises. Thinking of information as encoded in sentences at some abstraction level, we come to *information as code*. In this setting, the major paradigm is 'inference' in some general sense, involving *proof theory* and theory of computation. Again dynamic processes are of the essence here, as both deduction and computation are stepwise activities of 'elucidation' that manipulate syntactic representations. For a concrete illustration, think of successive stages in the solution of a 3x3 'Sudokoid':

1	•	•
•	•	*2*
•	•	•

,

1	•	*3*
•	•	*2*
•	•	•

,

1	*2*	*3*
•	•	*2*
•	•	•

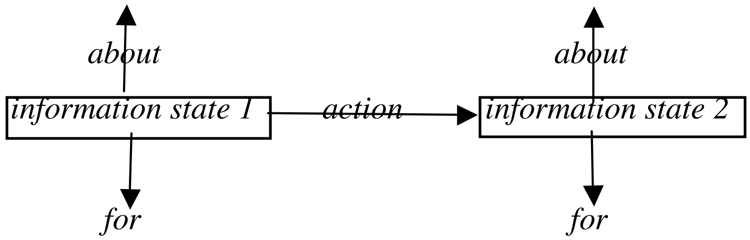
...

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

Co-existence and unification In all, then, we see several notions of information in logic, and dynamic processes transforming them. In all these, we find 'aboutness': information is *about something*, and 'agency': information is *for someone*. In an unpretentious diagram:

[215] Barwise & van Benthem 1999 develop the model theory of 'entalment across a relation' between models, including generalized interpolation theorems that allow for transfer of information.

[216] Other *très* Indian examples include observing 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.



This picture suggests that information as range and correlation are compatible. Likewise, the co-existence of semantics and syntax invites comparison. Even so, no grand unification of all logical notions of information is known. We even doubt whether it is desirable.

Merging range and correlation It is often thought that epistemic logic and situation theory are hostile paradigms, but range and correlation views mix well, especially in modal logics.

Modal constraint logic In a world of one-shot events, no significant information can flow. Genuine constraints arise in situations with different states that can be correlated. To make this more precise, 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:

s_1: p, s_2:q	s_1: p, s_2: $\neg q$
s_1: $\neg p$, s_2: q	s_1: $\neg p$, s_2: $\neg q$

With all these present, no situation 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 two states:

s_1: p, s_2: q,	s_1: $\neg p$, s_2: $\neg q$

Now, the truth value of p in s_1 will determine that of q in s_2, and vice versa:

the constraint s_1: $p \leftrightarrow s_2$: q holds.

Correlation between situations are restrictions on the state space of possible behaviours.

Definition Constraint models.

Constraint models $\boldsymbol{M} = (Sit, State, \boldsymbol{C}, Pred)$ have a set *Sit* of situations, a set *State* of valuations, a predicate *Pred* recording which atomic predicates hold where, and a 'constraint relation' *C* stating which assignments of states to situations are possible. ■ [217]

Definition Modal constraint logic.

Take a language with names x for situations (a tuple $\boldsymbol{x}$ names a tuple of situations), and atomic assertions $P\boldsymbol{x}$ for properties of or relations between situations. We also take Boolean operations, plus a universal modality $U\phi$ ('ϕ is true everywhere'):

$$P\boldsymbol{x} \mid \neg \mid \vee \mid U.$$

The semantic interpretation has obvious clauses for the notion:

$\boldsymbol{M}, \boldsymbol{s} \models \phi$ $\quad$ ϕ is true in global state $\boldsymbol{s}$ of model $\boldsymbol{M}$

In particular, $P\boldsymbol{x}$ holds at $\boldsymbol{s}$ if the tuple of local states assigned by $\boldsymbol{s}$ to the tuple $\boldsymbol{x}$ satisfies the predicate denoted by P. The resulting logic is classical propositional logic plus the modal logic *S5* for the universal modality U. Next, consider the shift relation:

$$\boldsymbol{s} \sim_x \boldsymbol{t} \quad \text{iff} \quad \boldsymbol{s}(x) = \boldsymbol{t}(x) \text{ for all } x \in \boldsymbol{x},$$

which lifts to tuples of situations $\boldsymbol{x}$ by requiring equality of $\boldsymbol{s}$ and $\boldsymbol{t}$ for all coordinates in $\boldsymbol{x}$. Thus, there are modalities $[]_x\phi$ for each such tuple, which say intuitively that the situations in $\boldsymbol{x}$ settle the truth of ϕ in the current system:

$$\boldsymbol{M}, \boldsymbol{s} \models []_x\phi \quad \text{iff} \quad \boldsymbol{M}, \boldsymbol{t} \models \phi \text{ for each global state } \boldsymbol{t} \sim_x \boldsymbol{s}$$ [218]

Constraint models satisfy the following two persistence properties for atomic facts:

$$P\boldsymbol{x} \rightarrow []_x P\boldsymbol{x}, \qquad \neg P\boldsymbol{x} \rightarrow []_x \neg P\boldsymbol{x}$$

The extended modal constraint language has a decidable complete logic with modal *S5* for each tuple modality, plus all axioms $U\phi \rightarrow []_x\phi$, and $[]_x\phi \rightarrow []_y\phi$ whenever $\boldsymbol{y} \subseteq \boldsymbol{x}$. ■

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

[218] There is a formal analogy here with *distributed knowledge* for groups of agents, Chapter 2.

Digression: modal constraint logic and first-order logic of dependence Van Benthem & Martinez 2008 point out how modal constraint logic equals the decidable first-order logic of dependent variables in van Benthem 1996, including single and polyadic quantifiers as well as single and simultaneous substitution operators. Van Benthem 2005 shows how modal constraint logic can be faithfully embedded into the latter, and also vice versa. Thus, constraints and dependence are the same topic in two different guises. Dependence is a major theme in the foundations of logic these days (Abramsky 2006, Väänänen 2007).

Combining epistemic logic and constraint logic Adding epistemic structure is natural in this setting. A blinking dot on my radar screen is correlated with an airplane approaching. But it does so whether or not I observe it. I may 'have' the information about the airplane, when I am in a situation at the screen, but unless I *know* that there is a blinking dot, it will not do me much good. That knowledge arises from an event: my observing the screen. To model this, we can use a combined *epistemic constraint language* interpreted in bi-modal structures of the form $\boldsymbol{M} = (Sit, State, \boldsymbol{C}, Pred, \sim_i)$, combining correlation and range talk. E.g., suppose that our model $\boldsymbol{M}$ satisfies the constraint $s_1\!: p \rightarrow s_2\!: q$. Then the agent knows this, as the implication is true in all worlds in $\boldsymbol{M}$. Now suppose the agent knows that $s_1\!: p$. In that case, the agent also knows that $s_2\!: q$, by the Distribution Law of epistemic logic:

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

The converse requires more thought. The point is that, if the agent were to *learn* that $s_1\!: p$, she would also know that $s_2 :q$. In dynamic-epistemic terms:

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

This formula is equivalent to the constraint – by the axioms f *PAL* (Chapter 3). Next, what do agents know about specific situations x? If $[]_x\phi$ holds at world $\boldsymbol{s}$, must the agent know this: $[]_x\phi \rightarrow \boldsymbol{K}\,[]_x\phi$? Not so: $[]_x\phi$ can be true at a world, and false at epistemically accessible ones. What a situation x 'knows' in the impersonal sense of correlation need not be known to an external agent, unless she makes an observation about x. Thus, a combined modal-epistemic logic brings out the interaction between our two senses of information – and it shows that range and correlation views of information can co-exist in obvious ways.

Explicit dynamics The co-existence extends to dynamic aspects. Like dynamic epistemic logic, situation theory involves event scenarios – in particular, for making use of ('harnessing') information. A typical example is the 'Mousetrap' of Israel & Perry 1991. [219] These scenarios suggest *dynamic constraint models* ***M*** *= (Sit, State,* ***C****, Pred, Event)* whose modal languages and logics are again familiar and perspicuous. More sophisticated models with epistemic information, correlations, and informational events are discussed in Baltag & Smets 2007 on the structure of quantum information states and measurement actions.

Inferential information and realization Our next step should draw inferential information into the circle of ideas developed here. But that is precisely what was done in Chapter 5 of this book, both at a static and a dynamic level. The upshot of that analysis was two-fold. First, models can be found for inferential information flow that resemble those of dynamic-epistemic logic, endowing worlds with syntactic 'access'. But the resulting mechanism works for a much wider variety of actions than inference, including acts of introspection – all under the common denominator of converting *implicit* into *explicit* knowledge. [220]

Grand frameworks: resource logics and their ilk Co-existence is not yet unification. There are several abstract perspectives claiming to merge all three notions of logical information. The 'Gaggle Theory' of Dunn 1991, inspired by algebraic semantics for relevant logic, is an abstract framework that can be specialized to combinatory logic, lambda calculus and proof theory, but on the other hand to relational algebra and dynamic logic, i.e., the modal approach to informational events. Van Benthem 1991 has a similar duality in categorial grammars for natural language, which sit at the interface of parsing-as-deduction and dynamic semantics. He points out how the basic laws of the categorial 'Lambek Calculus' for product and directed implications have both dynamic and informational interpretations:

$$A \bullet B \Rightarrow C \quad \text{iff} \quad B \Rightarrow A \rightarrow C$$

$$A \bullet B \Rightarrow C \quad \text{iff} \quad A \Rightarrow C \leftarrow B$$

[219] Van Benthem, Israel & Perry 2008 explicitize the dynamics in situation-theoretic scenarios.

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

Here, the product can be read dynamically as *composition* of binary transition relations of some process, and the implications as the right- and left-inverses. But these laws also describe a universe of information pieces that can be *merged* by the product operation. $A \rightarrow B$ is then a directed implication denoting $\{x \mid \forall y \in A: y \bullet x \in B\}$, with $B \leftarrow A$ read in the corresponding left-adjoining manner. On both interpretations, the principles of the Lambek Calculus hold. Beyond that, the usual structural rules of classical inference fail [221], and thus, there is a strong connection between *sub-structural logics* and abstract information theory (Mares 1996, Restall 2000). Sequoiah-Grayson 2007 is a defense of the Lambek calculus as a core system of information structure and information flow. While this is appealing, the above axioms merely encode the minimal properties of mathematical adjunctions, and these are so ubiquitous that they can hardly be seen as a substantial theory of information.

Conclusion Our three logical notions of information are compatible, but we remain somewhat agnostic on the issue whether they all reduce to a significant common source.

12.5 Philosophy of mathematics: intuitionistic logic as information theory

As a concrete test of our combination of observational and access dynamics in Chapter 5, we bring it to bear on a much older system that has long been connected with proof and information, viz. *intuitionistic logic*, a famous alternative to epistemic logic.

Semantics After its proof-theoretic origins, intuitionistic logic picked up algebraic and topological 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. The current version of this are intuitionistic Kripke models, which we will take here as *partial orders* $\boldsymbol{M} = (W, \leq, V)$ 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* $\boldsymbol{M}, t \models \psi$

[221] In particular, the rules of Contraction and Permutation would express highly questionable assumptions about procedural or informational resources, which have no appeal in general.

$\boldsymbol{M}, s \models \neg\varphi$ iff *for no* $t \geq s$, $\boldsymbol{M}, t \models \varphi$

Here, 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 this behaviour to all formulas. E.g., a negation says the formula itself will never become true at any further stage of the process. 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 two pictures below, which stand for the start of informational processes unfolding as downward trees:

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 end-points 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. Even so, the technical perspectives of Chapter 2 for epistemic logic fully apply in this setting. For instance, modal bisimulation is also an equivalence between informational processes.

Procedural information The main point is that intuitionistic models register *two* notions:

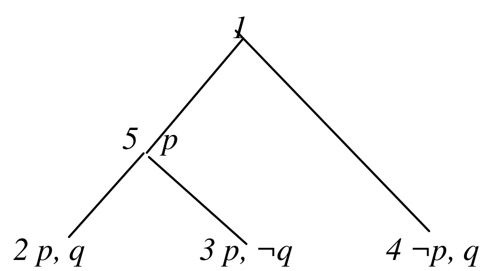
(a) *factual information* about how the world is; but on a par with this:

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

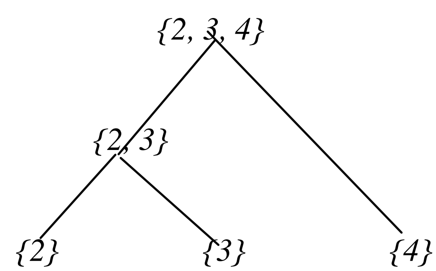
How we get our 'knowledge that' matters deeply, and while the leaves record factual information, the branching structure of our tree models, and in particular, available and missing intermediate points, encodes agents' knowledge of the latter kind. In fact, the distinction between factual and procedural information makes sense much more widely. [222]

Procedural information already appeared in Section 3.5, which suggested extending dynamic epistemic logic with *protocols* stating which histories are admissible in the investigative process. Protocols return in the *epistemic temporal logics* of Chapter 11, which describe branching histories using temporal operators staying on one branch, and modal operators quantifying over branches. Now we can draw an interesting 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 end points below, or the *complete histories* of the process. Thus, we can assign epistemic models as follows:



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

Warm-up: trading future for current uncertainty in games This is reminiscent of the epistemic analysis of *games*. In a game of 'perfect information', players know where they

[222] All points here also apply to *modal* generalizations of intuitionistic logic, whose richer language over pre-orders allows non-persistent statements in the investigative process (van Benthem 1989).

are at each node in the extensive game tree, but they do not know what the future will be. But there is a folklore observation that such 'global' uncertainty about the future can be converted into 'local' uncertainty about the present (cf. van Benthem 2004). Given any game tree *G*, assign epistemic models $\boldsymbol{M}_s$ to each node *s* whose domain is the set of complete histories passing through *s* (which all share the same past up to *s*), letting the agent be uncertain about all of them. 'Worlds' in these models may be seen as pairs *(h, s)* with *h* any history passing through *s*. It is natural to view the resulting structure ***M****(G)* as a *TPAL* protocol model, where the actions are announcements which move is taking place. This will cut down the current set of histories in just the right manner. [223] Van Benthem 2008 also discusses appropriate logical languages matching this reduction.

The same construction converts intuitionistic trees into *PAL* protocol models. But there are differences with the above scenario. First, we lack unique labels for 'moves': there are just anonymous upward inclusion links. Also, we have no unique description of each history, since we need not (and cannot) assume that different end-points in the tree carry different valuations. So, in the spirit of 'dynamification', what are the underlying actions?

Announcement actions The first type of action are *public announcements*. Assuming each end-point is uniquely definable in the language, each stepwise shrinking of the set of reachable endpoints is defined by the announcement of the negations of all definitions for endpoints that drop out. And in case there is no reduction in reachable endpoints (see below), we can still put the announcement *!T*. Thus, to a first approximation,

> *Intuitionistic logic describes effects of observations of facts,*
> *but without making the nature of these observations explicit.*

Actions of explicit 'seeing' Much more intriguing from the perspective of this chapter, however, is the need for a second type of dynamic action, which comes to light when we consider our initial examples in Section 2 of failure for the Double Negation law:

---

[223] A game of perfect information becomes a game of *imperfect information* in this way. But it is natural to distinguish two kinds of imperfect information now, *observation uncertainty* about how the game has developed so far, and *expectation uncertainty* about how it is going to continue.

The second model M_2 poses no problems. We put different singletons at the end-points, and their union at the root. A protocol with just announcements *!p* and *!¬p* will 'split' these as required. But now consider the first model M_1. One natural *PAL* version would put the same singleton set *{s}* at both nodes, as nothing is ruled out going from one to the other. But then no information flows, and, if we think of *p* as a property of the end state, we already *knew that p* at the start. Knowledge in this sense matches the intuitionistic operator ¬¬, or its modal counterpart $\Box\Diamond$, referring to eventual truth in all reachable end-points. But in intuitionistic models, actually putting *p* at a stage means more than just its 'inevitability' in the latter sense. What stronger event is taking place? Our proposal is that it is just a notion developed in this chapter, viz. an dynamic update of *explicit seeing*.

Conclusion In our perspective, intuitionistic logic as an account of information flow is akin to both dynamic-epistemic logic of observation and syntactic logics of explicit seeing and 'awareness', merging events of public observation with private acts of realization.

Of course, many other aspects of intuitionistic logic are not explained in this way. These include its elaborate accounts of constructive *proof* and of constructive *definition*. But our analysis does show that intuitionistic logic has a plausible interpretation as a theory of information-driven agency, something that even gets reinforced through its multi-agent game-theoretic interpretation via dialogues (Lorenzen 1955). I am somewhat undecided on what to make of this right now – but in the final analysis, the Logical Dynamics of this book might sit well with constructive logic, and perhaps even constructive mathematics. What this would mean to me, in particular, is that there is no principled borderline between 'science' and 'common sense' in this area, and thus, that intuitionistic logic should also be a natural fit with other aspects of agency in this book, such as belief revision.

12.6 Philosophy of logic

Our final topic is the view of logic emerging from this book. As we have stated before, discussions in the philosophy of logic tend to address issues reflecting practice of 50 years ago, and sometimes, they use definitions of 'logic' that were out of date even then. So, where is logic heading? There is a feeling that the field is 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?

Two views There seem to be two broad answers in circulation today. One is *logical pluralism,* locating the new scope of logic in charting a wide variety of *reasoning styles*, often marked by non-classical structural rules of inference. The latter view still sees the heart of logic in describing consequence relations, a view which I myself subscribed to in my work on sub-structural logics around 1990. But gradually, I have changed my mind about the crux of what logic should become. I would now say, with the *logical dynamics* of this book, that the main issue is not variety of reasoning styles and notions of consequence, but the variety of *informational tasks* performed by intelligent interacting agents, of which inference is only one among many, involving observation, memory, questions and answers, dialogue, or general communication. And logical systems should deal with a wide variety of these, making information-carrying events first-class citizens in their set-up. In this final section, I will contrast the two views, always with a view toward reconciling them. In particular, I argue that logical dynamics sets itself the more ambitious diagnostic goal of explaining *why* sub-structural phenomena occur, by 'deconstructing' them into classical logic plus an explicit account of the relevant informational events. I see this as a more challenging departure from traditional logic, and richer fare for philosophers to chew on.

Styles of reasoning 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 interesting principles high-lighted in Scott 1971 and later publications. These are the following *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 80s saw a wave of notions of consequence for different reasoning styles. Relevant logic dropped monotonicity, default logics dropped monotonicity and transitivity, resource logics in categorical grammar and linear logic dropped contraction, and so on. A general theory developed in the work of Gabbay 1996, Dunn 1991, Restall 2000, and others, while Dosen & Schroeder-Heister 1993 coined the term 'sub-structural logics'. Van Benthem 1989 noted the analogy with the agenda for logic in Bernard Bolzano's *Wissenschaftslehre* (1837) which did not focus on 'logical constants', but on charting formal properties of different reasoning styles: deductive or probabilistic, in the common sense or according to philosophical standards. [224]

Representation theorems An analysis of consequence in terms of structural rules has an attractive Spartan austerity. Still, one wants to tie this abstract level to some richer picture. This is provided by the usual *representation theorems* (van Benthem 1991, 1996A). Here is a folklore example, showing the semantic bare bones of classical consequence:

Theorem An inference 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. ■

---

224 The term I proposed for this enterprise: *'Bolzano's Program'*, has never caught on, though this German-Italian pioneer continues to exert an appeal to logicians (van Benthem 1985, 2003A).

More sophisticated representation theorems work for further notions of consequence.

Two worries Even so, in Benthem 1989, I voiced two concerns. First, it seemed that many facts in terms of structural rules address *mere symptoms* of some underlying phenomenon. For instance, non-monotonicity is like 'fever': it does not tell you which disease causes it. Thus, I was missing a deeper analysis of the underlying phenomena as a matter of logic. Matching this was a second worry. Sub-structural logics often arise from 'giving up' some properties of classical consequence, while retaining the old formal language. But why not be radical *with respect to the language* as well, and reconsider what we want to say? Admittedly, this happened with linear logic and its splitting' of classical connectives, and the same is true to some extent for relevant logic as well. But, for instance, it has not happened with circumscription and default logics, and we will return to that issue below.

So, given this picture, and these concerns, can broad-minded logicians 'dig deeper'?

Logical dynamics, rational agency, and intelligent interaction Well, this is precisely what the logical dynamics program of this book is all about. Instead of consequence, we take logic to be a much broader study of all sorts of informational processes: computation, observation, communication, and the like. [225] This offers a coherent vision of what logic could become, and we will now contrast and compare this with 'logical pluralism'.

Information update and structural rules for dynamic inference A first way of comparing already occurred in Chapters 1, 3. The Restaurant example of Chapter 1 had a natural matching notion of consequence. One first processes the information in the premises, and then checks the conclusion ('update-to-test'; Veltman 1997, van Benthem 1996):

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 a model where announcement of φ effects no further change: i.e., in the model $(\ldots(\boldsymbol{M}|P_1)\ldots)|P_k, s)$ the formula φ was already true everywhere, even before it was announced. ■

[225] 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.

Modulo a few details, dynamic validity amounts to *PAL* validity of a dynamic-epistemic formula which says the conclusion becomes *common knowledge*: $[!P_1]\ldots[!P_k]\ C_G\varphi$ (#) [226]
In the case of a single S5-agent, we can replace the common knowledge modality by just K.

Fact For purely factual (non-epistemic) formulas $P_1, \ldots, P_k, \varphi$, dynamic consequence holds if and only if 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 a model $\boldsymbol{M}$ to an updated model $\boldsymbol{M}|P$. This is not true in general (Chapter 3), and the following was easy to show using Moore-type statements:

Fact All classical structural rules fail for dynamic validity.

Now, precisely the same phenomenon emerges that we already know from general sub-structural logic. There are *modified structural rules* that do remain valid in this setting:

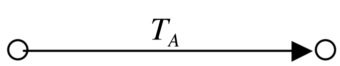
Fact Dynamic consequence satisfies the following structural rules:

$\text{if } \boldsymbol{P} \Rightarrow C \text{, then } A, \boldsymbol{P} \Rightarrow C$ *Left-Monotonicity*

$\text{if } \boldsymbol{P} \Rightarrow A \text{ and } \boldsymbol{P}, A, \boldsymbol{Q} \Rightarrow C \text{, then } \boldsymbol{P}, \boldsymbol{Q} \Rightarrow C$ *Left-Cut*

$\text{if } \boldsymbol{P} \Rightarrow A \text{ and } \boldsymbol{P}, \boldsymbol{Q} \Rightarrow C \text{, then } \boldsymbol{P}, A, \boldsymbol{Q} \Rightarrow C$ *Cautious Monotonicity*

Now we are at the abstraction level of structural rules, and indeed, these rules are valid in a much more general setting. We can view propositions A dynamically as *partial functions* T_A taking input states meeting the preconditions of update with A to output states:



Definition Abstract transition models $\boldsymbol{M} = (S, \{T_A\}_{A\ Prop})$ consist of states S with a family of transition relations T_A for each abstract 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

[226] Here, validity refers to the *Supermodel* of all epistemic models with arbitrary announcement steps. But when modeling realistic scenarios of conversation or enquiry, we can also relativize this to smaller *restricted families* $\boldsymbol{M}$ of epistemic models, with *protocols* of admissible announcements.

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 say *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, Chapter 7, proves the following representation result:

Theorem A sequent σ is derivable from a set of sequents Σ by these three rules iff σ is true in all models where all sequents in Σ are true.

This abstract analysis of a sub-structural consequence relation extracts precisely the 'gist' of dynamic inference in dynamic-epistemic logic (van Benthem 2003B has the details):

Definition A *meta-sequent* $\Sigma \Rightarrow \sigma$ from a set of sequents Σ to a sequent σ is *update-valid* if all its substitution instances with epistemic formulas, reading sequents again as type (#) dynamic-epistemic formulas, gives a valid implication between *PAL*-formulas. ■

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.

Natural levels of abstraction: modal logic of dynamic consequence Actually, the leap from *PAL* to a sequent-style analysis is drastic. Other natural abstraction levels behind our *PAL*-style dynamic-epistemic logic of agency include slightly richer modal languages. Indeed, the above transition models are really just models for a standard *poly-modal logic*. The above notion of dynamic validity needs *two* basic kinds of modality, viz. (a) universal modal boxes for the premise transitions, and (b) a 'loop modality' for the fixed-points:

$$\boldsymbol{M}, s \models (a)\varphi \text{ iff } R_a ss \ \& \ \boldsymbol{M}, s \models \varphi$$ [227]

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. But the modal language can also express properties of consequence beyond mere structural rules. Thus, modal logic seems a natural stage for a richer abstract theory of dynamic inference, while the above representation result can also be extended.

---

[227] Added to *PAL*, such fixed-point operators add expressive power: cf. Baltag & Smets 2007.

Summary The relation between abstract analysis of consequence relations and dynamic logics of agency gives rise to non-trivial questions. One has the following schema:

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

'From dynamics to consequence', the issue is to find good *abstraction levels* capturing significant properties of consequence relations generated by the concrete activity modeled in the dynamic logic. 'From consequence to dynamics', one reconstructs (or just brings out of the closet) the dynamic practice generating the given consequence relation, and this is what representation theorems do. The two directions obviously 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.

Non-monotonic reasoning and dynamic logic of belief change A good test case for this view is its extension to logics for belief change, rather than knowledge update. Classical consequence from ***P*** to *C* says all models of ***P*** are models for *C*. The famous insight in McCarthy 1980 was that problem solving and planning go beyond this, by zooming in on the most 'congenial' models. A *circumscriptive* consequence from ***P*** to *C* says that

C is true in all the *minimal* models for ***P***

Abstract structural description levels here include conditional logic (Shoham 1988, Gaerdenfors & Rott 1995, Gabbay 1996) and belief revision theory in terms of the *AGM*-postulates (Gaerdenfors 1987). Again there are many representation theorems.

Dynamification: non-monotonic logic as monotonic logic of belief revision Now let us extract the underlying activities gain. This is precisely what was done in Chapter 6: the relevant process is belief formation and *belief revision*. The puzzles that motivated non-monotonic logic are naturally recast in these terms (van Benthem 2009): in problem solving and planning, we form beliefs and drop them when needed. So recall our basic scenarios:

Belief change under hard information The complete logic of conditional belief under public announcements was axiomatized by recursion axioms in Chapter 6. And again, it supports dynamic consequence – in fact even two. Both notions state what happens once the premises are processed: either *knowledge* results just as before [228] – or we go to *belief*:

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

Given our semantics of belief, and working with factual propositions, the latter is precisely the dynamic counterpart to minimizing consequence relations like circumscription:

$$P_1, \dots, P_k \Rightarrow \varphi$$

Here, our claim is simply this. Circumscription leads to *beliefs* rather than knowledge, since its conclusions may be retracted on the basis of further evidence. But then, what has traditionally been cast as a new 'non-standard' consequence relation may also be seen differently through 'dynamification'. Making the dynamic setting more explicit, we have a *dynamic logic of belief formation under incoming factual propositions*. Technically, the hallmark failure of monotonicity then occurs because of the minimization in the definition of belief – not because of some special feature of the notion of consequence as such.

But our setting is richer. A dynamic epistemic and doxastic language allows complex *non-factual* propositions for premises and conclusions. The usual accounts of structural rules and the 'intuitions' associated with them do not take this more sophisticated information into account. But the above calculus will keep all this absolutely straight: we do have the complete logic for the total language! Also, as we saw in Chapter 6, the semantic framework allows for new operators beyond knowledge and belief, like 'safe belief'. And so it raises new questions. What would be the *new consequence relation* associated with using safe belief rather than plain belief? A final dynamic degree of freedom in setting up consequence relations is *the way in which we add* the information from the premises:

[228] This should be *common knowledge* in the multi-agent case, and likewise later *common belief*.

Belief change under soft information Soft triggers merely rearranged plausibility patterns among worlds. A typical example was *lexicographic upgrade* $\Uparrow P$ which radically changes the current ordering relation ≤ between worlds as follows:

> all P-worlds in the current model become better than all $\neg P$-worlds,
> while, within those two zones, the old plausibility ordering remains.

Again there was a complete dynamic doxastic logic for this way of revising beliefs. But more to the point here, we can now define new notions of dynamic consequence beyond circumscription – by processing the successive premises 'softly'. Here is one question then:

> *What are complete sets of structural rules for the consequence relations:*
> $P_1, \dots, P_k \Rightarrow_{circ\text{-}hard} \varphi$ iff $[!P_1] \dots [!P_k]\, B\varphi$
> $P_1, \dots, P_k \Rightarrow_{circ\text{-}soft} \varphi$ iff $[\Uparrow P_1] \dots [\Uparrow P_k]\, B\varphi$*?*

Here is at least a structural difference between the two notions, even for factual assertions:

Fact For factual assertions P, Q, (i) $P, Q \Rightarrow_{circ\text{-}hard} P$, (ii) not $P, Q \Rightarrow_{circ\text{-}soft} P$.

Proof (i) Successive hard updates yield subsets of the P-worlds. (ii) The last upgrade with Q may have demoted all P-worlds from their former top positions. ■

How to choose between such alternative notions? It all depends on the scenario of problem solving or game playing that we are engaged in. Indeed, our logic provides many more, once we look at other plausibility-changing events. And the shifts in that plausibility order are really the primary issue in understanding how we navigate through the task at hand.

Summary The intuitions behind circumscriptive inference styles involve knowledge and belief. They are also dynamic, involving agents' responses to incoming information. Thus, in a dynamic epistemic perspective, circumscription and other styles of non-monotonic reasoning are at heart about cognitive attitudes and responses to information. Moreover, these responses can be quite different, from hard information update to soft plausibility change. Merging things in this way fits with the general conception of agency stated in

Section 2: processes of *inference* and *self-correction* go hand in hand! This perspective also generates new consequence relations, and new employment for sub-structural analysis. [229]

Conclusion and outlook 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. Its 'strong arm' is mathematical analysis of possible consequence relations. The other program is 'Logical Dynamics', emphasizing events of information flow from inference to observation, and the many processes by which rational agents harness this to interact. Its paradigm is dynamic logic in some broad sense. We have shown how to move back and forth between the two perspectives by abstraction and 'dynamification', and sometimes get precise connecting results. But, I am not completely neutral. Even though I started out on the consequence side, I now prefer the dynamic view – partly because it is more ambitious, and gives logicians many more things to do!

Also, we saw how a non-monotonic consequence relation reduces to a classical dynamic logic of the process that *causes* the non-monotonicity. In a slogan, *monotonic dynamic logic can model non-monotonic consequence!* The general scope of this slogan remains to be understood, but Baltag & Smets 2007 show similar things for 'quantum logic'.

Challenges to dynamics We have done only two case studies. One would also want a dynamic take on *para-consistent* logics in terms of underlying processes that handle inconsistencies, and on *resource-conscious logics* such as linear logic of the Lambek Calculus, maybe putting explicit operations on information pieces into the logic. Also, one should find a way to handle *language change* (reflecting conceptual changes) in response to incoming events. The latter interest would go back to Bolzano after all, who did include language as an explicit and crucial parameter in his account of logical consequence.

Once again: what is logic? The background to this section is the question what logic really is. Some seek this in mathematical notions of semantic invariance (Bonnay 2006), others in some proof-theoretic essence (Martin–Löf 1996). I myself see the point of logical dynamics

---

[229] Of course, it would add support to my general position in this paper if we could do such a dynamified reconstruction of other non-monotonic reasoning styles, too: say, for *abduction*.

as loosening the magnetic spell of mind grooves from the grand foundational period of the 1930s. This includes a turn away from proof and computation as the only paradigms toward a wide spectrum of rational agency and intelligent interaction. I have no view on specific 'axioms' that need to be preserved in this endeavour. I see the essence of the discipline of logic as a dynamic activity, not as any static product of that activity: proofs, formal systems, or languages. And that is Logical Dynamics again, now applied to the whole field.

12.6 Conclusion

This concludes our tour of contacts between logical dynamics and philosophy. Many more are waiting to be created – and all it takes is acquiring a sensibility to the 'dynamic stance'.