LOGIC IN PHILOSOPHY

Johan van Benthem, Amsterdam & Stanford University

To appear in Dale Jacquette, editor, *Handbook of the Philosophy of Logic*

1 Logic in philosophy

The century that was Logic has played an important role in modern philosophy, especially, in alliances with philosophical schools such as the Vienna Circle, neo-positivism, or formal language variants of analytical philosophy. The original impact was via the work of Frege, Russell, and other pioneers, backed up by the prestige of research into the foundations of mathematics, which was fast bringing to light those amazing insights that still impress us to-day. The Golden Age of the 1930s deeply affected philosophy, and heartened the minority of philosophers with a formal-analytical bent. As Brand Blanshard writes in *Reason and Analysis* (1964) – I quote from memory here, to avoid the usual disappointment when re-reading an original text:

> *"It was as if a little band of stragglers, in weary disarray after a lost battle, suddenly found Napoleon's legions marching right alongside of them..."*

In the 1940s and 1950s, people like Carnap, Reichenbach, Quine, and their students made logic-based methodology a highly visible modus operandi in the field. Then, in the 1960s, what came to be called 'philosophical logic' began to flourish, and logicians like Hintikka, Geach, Dummett, Kripke, Lewis, and Stalnaker came to prominence, not just in an auxiliary role as logical lawyers (Prior's phrase), but even setting the agenda for philosophical discussions that were now suddenly considered mainstream.

It seems fair to say that this age of logical influence has waned. Many philosophical circles have no live contact with logic any more, and even basic logical skills seem sometimes absent from the current canon of 'what every philosopher should know'. Even technically inclined philosophers often feel that probability or game theory or dynamical systems are more versatile tools than logic for making formal points and relevant distinctions. And for their part, many logicians have gone elsewhere, to areas with more lively interactions, in departments of computer science, linguistics, or even further astray in the social sciences. Indeed, the bulk of logic research to-day is done in computer science, as can easily be checked in terms of numbers of publications, journals, organizations, and sheer variety of research topics.

Reactions to this situation vary. One is a turn to writing the history of the glory days of the foundational era, often with the concomitant fundamentalist view that everything would be better nowadays if we just returned to the purity – largely mythical – of that lost golden age. A response with which I feel much more sympathy is the activist stance of reintroducing logical ideas into philosophy – contemporary ones, to be sure. Indeed, the gap between current philosophical practice and modern logic may be more apparent than real. Many themes still run on parallel lines, and some differences just arise for practical reasons, viz. the absence of textbooks explaining modern logic to a philosophical audience, or of influential publications in philosophy showing modern logic at work. But take any informal gathering, and it soon turns out that philosophers and logicians are still on speaking terms, and sometimes even more than that!

In this chapter, we do not take a particular stand, but just try to approach the whole issue in a way that is somewhat different from the usual approaches.

Four dead-end streets In setting out this role of logic in philosophy, however, there are some situations to be avoided that obstruct an impartial view.

One are broad *philosophized histories* of what took place. A good example is the recent popularity of schematic views of the history of logic which claim that the field has been heading to a 'meta-level' self-understanding in the 1960s, occupied mostly with the study of logical calculi and their meta-properties. While this may be true for some areas of mathematical logic, this inward turn has been counter-balanced, throughout the second half of the 20th century, by new contacts with other areas, and 'object-level' logical analyses of new notions. The philosophized eschatology of some of my good friends just does not match up with what really took place.

A second barrier are divisions of philosophy into *fixed subdisciplines* such as epistemology, metaphysics, ethics, philosophy of language – and why not? – logic. There is some sense to such clustering. But, even in a formal discipline like mathematics, a geography in terms of different countries 'algebra', 'geometry', 'analysis' hides as much as it reveals – and creative mathematicians think in terms of *themes* and *methods* rather than rigid subfields. It is themes and their metamorphoses across subdisciplines that provide the coherence of a field. Here is the worst that can happen. Some atlases of philosophical logic even copy philosophical geography (epistemic logic, deontic logic, alethic modal logic), leading to a bad copy of a bad map of reality.

The third trap to be avoided is *system imprisonment* (van Benthem 1999B). Many logicians see the structure of logic as a family of formal systems: propositional logic, first-order predicate logic, higher-order logic, modal logics, etc. The formalists may not have won with their philosophy, but their world-view has insidiously inserted itself into the subconscious of the field. This systems view represents a dramatic change from thinking in terms of broad logical themes, such as Negation, Implication, or Validity, which may be studied using particular 'logics', but are not exhausted by them. The formal systems way of perceiving the field has led to a host of 'system-generated issues', which are then touted (often even successfully) as philosophically relevant. Examples include Quine's view of ontological commitment as an issue of ranges for first-order variables, the never-ending discussion about compositionality, or the elusive search for the first-order/higher-order boundary. It is only with great effort that some modern logicians are trying to break out of this mold, realizing that many notions that we study transcend particular formal systems, even though we lack the vocabulary to say what a general 'monotonicity inference', or a general 'recursive definition' is.

A fourth and final barrier to seeing things for what they are is the term *'application'*. Logic is not applied to philosophical problems the way an engineer might apply some technique for computing stress in a bridge. Its role seems much more indirect. Logic offers a technical language with relatively precise meanings as an enhancement of philosophical discourse, and an aid to precise *communication*. This is as useful as having mathematical language around in other disciplines: perhaps not just the Book of Nature, but also the Book of Ideas is written in mathematical language. Also, logical tools can sometimes be used to *analyze arguments* from the philosophical tradition, and throw new light on them. But again, there is usually no counterpart to 'solution' in the engineering sense. The problem gets illuminated: it may go away, but it may also acquire new nuances, and it may even become more urgent than ever! Analysis in logical form also helps bring out *analogies* between different notions or problems that were not observed before. Finally, on the more activist side, logic can sometimes help make philosophical argumentation more precise, and it can help the philosopher *construct new conceptual frameworks*. And yes, sometimes, *meta-theorems* about logical systems in the foundational mode have some bearing on philosophical issues – even though this particular use seems greatly over-rated in practice.

Tracking themes through time This chapter will not state a priori what the role of logic in philosophy is, or should be. Our plan is to just look at some themes as they developed through the 20th century. The richness, and surprising plot twists, of the resulting stories speak for themselves. This way of writing history was tried out in a Stanford course 'Logic in Philosophy' (2003D), and it will be the basis for a new textbook in philosophical logic. Our first two themes show how some of the core ideas of pre-modern logic survived the Fregean revolution, returning in modern forms:

i Logical form and monotonicity: from syllogistics to generalized quantifiers
ii Reasoning styles: from Bolzano to conditional logic and AI

The next theme shows how formal systems, viewed with some disrespect for authority, are indeed useful 'laboratories' for experimenting with new philosophical ideas:

iii Mechanisms of semantic interpretation: from Tarski to 'dynamics'

Our final two themes address the modern topic of placing actions at logical center stage, relating them to much earlier traditions:

iv Theories and belief revision: from neo-positivism to computer science
v Dynamic logic, communication, action, and games

In each case, we briefly trace a historical development of ideas up to the present time, without going into details. The references provide further depth, while Section 9 lists a few further themes. All our story lines involve other fields, mainly computer science. Indeed, confining these things to just philosophy and logic would distort the intellectual history. It would also do a disservice to philosophy, whose themes often cross over, and then return in new forms – a win-win scenario for everyone involved.

Finally, the references in this paper are meant to be illustrative, but they are not an official historical record. Also, documents mentioned need not be the original sources for a theme described here: they are sometimes later surveys. Indeed, the *Handbook of Philosophical Logic* (Gabbay & Guenthner, eds., 1983 and later years) as well as the *Companion to Philosophical Logic* (Jacquette, ed., 2002) are good general sources.

2 A joy-ride through history

For a start, we provide some further historical perspective. To dispel dogmatic slumbers, it is good to realize that there is no stable definition of the discipline of logic! Throughout the centuries, there have been many changes in the agenda of logic, its interdisciplinary environment, and the guiding interests of its leading practitioners. The entrenched fundamentalists of to-day are often the invaders of earlier periods.

Classical pre-19th century logic goes back to the dialectical tradition in Antiquity, which realized that success in debate involves fixed patterns of valid, or invalid, reasoning that can be studied as such. For instance, Modus Tollens is a wonderful 'attacking move' against one's opponent, by pointing out some false consequence of what she says. Here is a sobering point about human imagination. It took until the mid-20th century before the logician Paul Lorenzen (Lorenz & Lorenzen 1978) gave the first game-theoretic account of argumentation, doing justice to the fact that the dialectical tradition is about dialogue between different agents instead of lonesome arm-chair deduction, and winning debates in *interaction* with other participants. Debate was still very much a feature of logical disputations in the Middle Ages (Dutilh-Novaes 2003), but medieval logic as a whole took logic in the direction of what we would now call ontology and philosophy of language. Logic was then reinterpreted in a more epistemological manner in Kant's famous Table of Categories for the basic forms of judgment. The same author also proclaimed the 'end of history' for logic – always a dangerous type of prediction...

In the 19th century, leading logicians like Bolzano and Mill saw the subject on a continuum with general methodology, and what we now call philosophy of science. Gradually, influences from mathematics – already present with Leibniz – became prominent, witness the work of Boole and de Morgan on algebraic mathematical methods improving on earlier systems in the syllogistic paradigm. Toward the end of the 19th century, it was no longer a case of mathematics helping logic, but rather the other way around. Frege developed his logical systems for the purpose of securing the foundations of mathematics. Even so, Frege's contemporary Peirce still had a much broader view of what logic is about, including not just deduction, but also inductive and abductive reasoning. It is interesting to see that the 'fundamentalists' won the day with their narrower agenda, making mathematical proof the shining example of logical reasoning – instead of an extreme case of normal reasoning, found only under very rare circumstances. About a century had to elapse before Peirce became a source of inspiration again, now in more computational-logical circles.

Around 1900, modern logic also came to have philosophical impact. Russell's *misleading form thesis* posited that linguistic form, the philosophers' analytical tool until then, can be systematically misleading. One must find the underlying logical forms to express the real meanings and keep argumentation lucid. On this view, logic becomes the 'calculus' of philosophy. Thus, until the 1940s, several logic-inspired programs existed for philosophy. There were the foundations of mathematics with its broad philosophical positions of logicism, formalism, and intuitionism. This research had the Popperian advantage of working with a refutable claim, viz. *Hilbert's Program* for finding conclusive proofs of consistency, completeness, and decidability were around the corner. Gödel's Theorems refuted this, but in the interesting way that mathematical impossibility results normally do. They leave us sadder, but above all: wiser, and full of new directions to pursue. The same was true for Turing's analysis of computation: it showed that many natural computational questions are undecidable, but the fall-out gave us the foundations of computer science. The second, broader line of logic-inspired philosophy is connected with Russell, the early Wittgenstein, and the *Vienna Circle*, a program for a logic-based renewal of methodology, philosophy of science, and philosophy in general. This program, too, had its refutable components, and these were indeed refuted by Quine in 'Two Dogmas of Empiricism' – though Popper later claimed he had helped put in a dagger from behind. Again, the more interesting point are the enduring insights from this second logical line, such the seminal work by Carnap, Reichenbach, and others on methodology and meaning.

These golden years produced the basic modus operandi in modern logic, of unquestioned importance even to those who would now like to broaden the agenda again. First-order logic, the modern notion of a truth definition and semantic validity, completeness theorems, proof theory, and many other basic notions go back to work in the foundationalist logic of those days. Over time, the broader philosophical repercussions of these notions and concerns started penetrating, witness Tarski's semantic conception of truth, or Turing's analysis of computability and intelligence.

Nevertheless, some originally 'logical' topics did drop by the wayside, as communities started forming with their own subcultures. Carnap is now considered a 'philosopher of science' – and during the 1950s philosophers of science took care of topics that had no place on the foundationalist agenda, such as inductive reasoning, explanation, or scientific theory structure. At the same time, the 1950s also saw the first challenges to the logical way of doing philosophy. Of course, mainstream philosophers perhaps ignored logic, and only a few traditionalist logicians were still bewailing the Fregean Coup – sometimes for very legitimate reasons (Wundt 1880/3, Blanshard 1964). But now, criticism came from within the analytical tradition, with the Wittgenstein-Austin 'natural language turn'. Informal linguistic analysis could serve just as well for the purposes of philosophy, or, as Austin said about common usage: "There is gold in them there hills". New appealing paradigms such as 'language games' arose, and there were even attempts to replace logic in philosophical curricula by courses in *game theory*, the upcoming formal paradigm of the day. Just try to imagine another possible world, where Wittgenstein, Nash, and Aumann are the heroes of formal philosophy!

But there was enough vigor in the logical tradition to survive all this. Indeed the 1960s/1970s became the heighday of an emergent milieu called *philosophical logic*, providing new formal tools for looking at long-standing philosophical problems. This was often triggered by taking a much closer look at natural language and our natural reasoning practices than what had been usual in the foundationalist tradition. It suffices to mention a brief list of eminent authors emerging in that period, whose work has sparked a follow-up which continues until to-day: Hintikka, Prior, Geach, Kripke, Rescher, Lewis, Montague, and Stalnaker. Their work was concerned with what may be called *modal logic* in a broad sense: including time, knowledge, duty, action, and counterfactuals. Many of these logicians wrote influential publications which started philosophical research lines, witness famous book titles like *Knowledge and Belief*, *Naming and Necessity*, *Counterfactuals*, or *Formal Philosophy*.

Incidentally, around 1970, logic was heavily attacked inside *psychology* in Wason & Johnson-Laird 1972, a publication which convinced many psychologists that logic was spectacularly unsuited for describing the human reasoning with which we really solve our problems and get by in life. Not many logicians will have cared about this threat in the distance, especially since they may have felt that psychologists were not quite people one would want to associate with anyway, given Frege's famous 'anti-psychologism'. But there may be more to this in a modern light: see below.

At the risk of offending some, I would say that the Golden Age of philosophical logic was over by 1980, and this new stream entered into steady state, just as mathematical logic had done, with technical depth taking over from the original intellectual passion. But at the same time, the flame of logic carried over to other fields. Logic gained importance in disciplines beyond philosophy and mathematics, such as linguistics, computer science, AI, and economics – and these are still flourishing interfaces to-day, often with their own communities, publications, and gatherings. Many topics from the mathematical and philosophical heartland acquired new lives in these settings, witness new theories of meaning, information update, belief revision, communication, and general interaction. One of the most surprising developments has been the design of logic systems of programming, performing various intelligent tasks (cf. the program of 'Logical AI' (McCarthy 2005), or Reiter's 2001 book *Knowledge in Action*. Even if logics do not completely capture all of our natural cognitive habits, they may at least be consistent with them, and then – gently inserted – enrich our repertoire of behavior! Thus, contemporary logic finds itself in a much more diverse interdisciplinary environment, perhaps closer to the richer agenda of logicians from earlier centuries.

At the same time, challenges persist. In both computer science and game theory, and indeed philosophy itself, *statistical* and probabilistic methods have gained importance, dealing with phenomena where logic has little to offer. Indeed, modern evolutionary approaches to games and methodology rely on dynamical systems rather than logic. Combining logic and probability seems a major challenge today. An even aggressively anti-logical stance surfaced in the computational paradigm of massive parallellism, replacing the Turing machine view of cognitive behavior. At the central AI conference on Knowledge Representation in Toronto around 1990, a keynote speaker felt we were like the French Ancien Regime in our ballroom, while teeming masses of neural netters were shouting "A la laterne, les aristocrates". But again, history has its own ways. By now, it is becoming clear that neural nets and logics can live in harmony (d'Avila Garcez, Broda & Gabbay 2002, Leitgeb 2004, Smolensky & Legendre 2005), and that dynamic logics and dynamic systems are a match. Thus, the current interface between logic and cognitive science is in flux again (Hodges 2005), and doom scenarios have not materialized – so far.

Role models Another way of high-lighting an intellectual development looks at typical representatives. A Dutch professor of linguistics once said, when challenged about a logical turn in his work: "Linguistics is that work which is being done by prominent linguists. I am one of this country's most prominent linguists. Therefore, what I am doing is linguistics". Role models for a person-oriented line would include philosophers/logicians like Peirce, Ramsey, Carnap, Geach, Dummett, Hintikka, Grice, Kripke, Lewis, Stalnaker, Belnap, Kamp, and Barwise – and a very interesting chronicle might result. What did these leaders consider the most pressing themes, and why? Nevertheless, in the rest of this paper, we consider themes rather than people.

3 Logical form and natural logic: from syllogistics to generalized quantifiers

Natural Logic Is natural language unsuitable as a vehicle for logical inference? Frege thought so, and cited its 'prolixity' as a reason for designing the formal language of

Begriffsschrift. In opposition to this, a tiny band of traditionalist critics defended classical syllogistics as staying much closer to natural inferences that people actually make (Sommers 1982, Englebretsen 1981). This debate was reopened in the 1980s under a foreign influence, viz. the penetration of logic into *linguistic semantics*. Richard Montague's earlier work (Montague 1974) – influenced to some extent by Donald Davidson's critical stance toward the Tarskian formal logic of his day – was seminal in this development, in being two-edged. It used formal logic to analyze natural language expressions – but in doing so, it also showed that natural language did have a respectable structure that supports impeccable meanings and inferences. The modern theory of *generalized quantifiers* spawned by Montague Grammar has replaced Misleading Form antagonism by concrete questions about the actual interactions of reasoning encoded in natural language and formal logic.

Traditional logic A brief account of logical form runs as follows (GAMUT 1991). Traditional logic used the subject predicate format *S is P*, reflecting the structure of natural language sentences. Even so, this was not really a logic of sentence forms, but of intensional relations between concepts denoted by the terms *S, P* – as is still clear, in Kant's famous Table of Categories. The inference engine for this reasoning about concepts were Aristotle's syllogisms for the basic quantifier expressions *all, some, no, not all,* forming the classic Square of Opposition. In the Middle Ages, the coverage of syllogisms was extended by the ingenious doctrine of *distribution*. This uses the fact that quantified premises allow for inference in deep sentence position, providing we use universal relationships to replace only occurrences of predicates that are 'about all objects' (in modern terms: downward monotonic ones), or their positive monotonic counterparts. This allows inferences that go beyond the traditional syllogism, such as

> If all men are mortal, and no rich mortal's son escapes Judgment Day,
> then no rich man's son escapes Judgment Day.
> If all men are mortal, and Judgment Day is hard on many men,
> then Judgment Day is hard on many mortals.

In this way, traditional logic could and did deal with many forms of judgment and reasoning – and even Leibniz felt it needed perfecting, rather than demolishing.

The critics win the day In the 19th century, progressive waves of critics stormed the bastion. Boole's attack was implicit, in showing how at least propositional reasoning could be dealt with in a perspicuous algebraic manner owing nothing to Aristotle.

Later historians found analogues in the work of the Stoic logicians – but it was Boolean insights that rehabilitated Stoic logic (Mates 1953), rather than Stoicism supporting Boole. Famous examples of the 'inadequacy' of traditional logic still taught to-day include De Morgan's non-syllogistic forms:

All horses are animals. *Therefore*, horse tails are animal tails

This purported knock-down objection is actually well within the scope of the medieval theory of Distribution. The fact that a binary relation occurs ("animal *x* possesses tail *y*") is completely immaterial, as the modern equation of the Syllogistic with a fragment of *monadic* predicate logic is an uncharitable later construal. Frege's later examples of the inadequacy of traditional logic concern more essential uses of relations:

from "The Greeks defeated the Persians at Plataeae,
to "The Persians were defeated by the Greeks at Plataeae".

Like most conservatives, Frege preferred high-lighting the aristocratic Spartan victory at Plataeae over the prior democratic Athenian one at Salamis. But his logical point is a strong one. Moreover, his anti-intensional compositional treatment of quantifiers (Dummett 1973) was a startling innovation. A sentence like "Everybody walks" now stood for a multitude of individual instances, such as "Mary walks", "John walks",... This extensional view derived complex sentence meanings recursively from properties and relations over some given domain of individual objects. And like all creative simplifications in science, it paved the way for new insights and concerns.

Why the quick collapse? It is an intriguing question why the traditional bastion crumbled so fast. Some modern historians (cf. Barth 1974) tried to paint an image of an Old Guard defending itself to the bitter end, and evil forces pitted against the Russelian and Wittgensteinian forces of Light (cf. the 'Kingdom of Darkness' in Geach 1972). But what they mainly managed to drag into the open were some pitiful traditional logicians, looking bewildered as the walls were tumbling around them. Indeed, the complaints of the old guard about unfairness were often on the mark. Van Benthem 1981 discusses two so-called 'crude errors' of traditional logic: its reduction of binary predicates to unary ones, and its use of generic objects, and shows how both still make a lot of sense, when viewed with some mathematical sensitivity. Still, the bastion did crumble, largely because the critics has a more dynamic agenda of research – and also, because bastions crumble, not so much because of the vigor of external onslaughts, but because of the internal dry-rot of stagnation and boredom.

From misleading form to Montague's Thesis As we saw, the Misleading Form Thesis made logic a favored calculus for philosophy. Even so, natural language refused to go away in philosophical practice and theory. Indeed, the very success of Montague's work around 1970 on logical grammar and semantics of natural language cast the issues in a completely new light. Montague retained the subject-verb form of natural language sentences, with its basic pattern $S \Rightarrow NP\ VP$. To do so, he treated noun phrases as *generalized quantifiers*, i.e., second-order properties of predicates. Thus, "Every bird sings" says that the property of singing holds for all birds – and to make "Tweety sings" come out in the same way, an individual bird like Tweety was 'lifted' to become the set of all her properties. This generalized quantifier view can also deal with iterated quantified expressions, as in "Every bird sings a song". The success of this approach motivated what became known as Montague's Thesis:

> 'There is no difference of principle between natural and formal languages' .

This is the Misleading Form Thesis reversed! From then on, the story shifted to linguistics, where Montague semantics became a powerful paradigm, and even computer science (Janssen 1981, de Roever et al. 2001), as programming languages, too, allow for the same sort of compositional logical treatment.

Fine-structure: generalized quantifier theory The *NP VP* format for quantified *NP* expressions has a three-part structure *(Det A) B*, where *Det* stands for a *determiner* expression: "(*Every* bird) sings". Thus, quantifiers themselves may be viewed a binary relations between extensional predicates *A*, *B*. E.g., *every* stands for inclusion between the sets of "birds" and "singers", and *some* for overlap. This account also works for non-first-order quantifiers. E.g., *most AB* says that the size of the set $A \cap B$ is greater than that of $A – B$. With this view, many new issues have come to light over the past two decades (Keenan & Westerståhl 1997, Peters & Westerståhl, to appear). Of relevance to us here is the inference pattern of *monotonicity*, describing stability under shifts in predicates. E.g., a true statement *every AB* remains true after replacement of its *A*-argument by a predicate denoting a subset, and after replacement of its *B*-argument by a superset. In technical terms, *every* is downward monotone in *A*, and upward monotone in *B*. The other classical quantifiers in the Square of Opposition exemplify the other three types: e.g., *some AB* is upward monotone in both its arguments. Modulo some mathematical assumptions, this can even be reversed: making the classical quantifiers the only 'doubly monotone' ones (van Benthem 1986).

Monotonicity reasoning Now, it turns out that this monotonicity behavior can be described completely generally for any occurrence of a linguistic predicate inside any linguistic expression, simply by interleaving it with grammatical sentence construction in some logic-friendly paradigm like *categorial grammar* (van Benthem 1986, Sanchez Valencia 1991, van Benthem 1991). This categorial 'monotonicity calculus' has been rediscovered several times over the past few years for its elegance and naturalness, and with hindsight, it can be discerned in the traditionalist manifesto Sommers 1982. To-day, it even serves as a platform for cognitive experiments (Geurts & van der Slik 2004). This vindicates the medieval doctrine of Distribution, and it accounts for perhaps the majority of the 'fast inferences' that language users make. Indeed, similar systems of light but pervasive quantifier inference have also turned up in computer science, describing fast algorithmic tasks in handling data bases.

Co-existence of natural logic and formal logic Monotonicity is just one form of 'natural reasoning'. Other principles include systematic restriction of ranges for quantifiers (already described by Peirce), or 'context shifting' across time and space (cf. van Benthem 1987). A complete system of all aspects of natural logic still remains to be formulated. Even so, it seems clear that first-order logic is by no means the only inferential standard. It rather seems a sort of formal laboratory for defining and studying certain notions (including monotonicity), which we can then use to describe actual inferential phenomena, if only in the form of a contrast. The general idea behind this view also reflects current computational realities – as computer scientists populate the world with an ever-growing arsenal of semi-formal languages, artificially designed but used in practice. There is a fruitful *co-existence* of natural and formal languages, and their mutual influences are probably the best way of laying earlier all-or-nothing discussions about misleading form, or about the 'impossibility' of a systematic theory of meaning for natural language, to a well-deserved rest.

But what about philosophy? The moral of this story is that the traditional debate about wholesale choices between natural or formal language as the philosopher's tool are misguided. Logical languages are an *enhancement* of philosophical practice – and in that light, philosophy is precisely one of the mixed settings that we advocated.

4 Reasoning styles: from Bolzano to conditional logic and AI

Styles of reasoning Most textbooks define 'validity' of an inference once and for all, as truth of the conclusion in all models where the premises hold. I still remember my flash experience as a student finding Bolzano's *Wissenschaftslehre* (1837) in our library, and seeing how, long before Tarski, someone had looked at what logic is about, and made some crucial decisions differently. Bolzano's best-known notion of logical consequence is indeed like Tarski's (van Benthem 1985), be it that *nothing*, rather than everything, follows from inconsistent premises. But beyond that, Bolzano describes the aim of logic as charting the *varieties of reasoning styles* that we apply to different tasks. These are deductive in mathematics, often inductive in daily life, or – in his view – strictest of all in philosophical argumentation, where drawing *relevant* conclusions is of the essence. Bolzano then lists formal properties distinguishing such reasoning styles. E.g., he points out – in modern terms – that some are *monotonic* (adding new premises does not affect earlier conclusions), while others are not. This idea of a variety of reasoning styles returns with Peirce, the neo-positivists, and again in modern *AI*, when non-monotonic logic arrived on the scene in the 1980s as an account of how people actually solve puzzles and go about practical planning tasks (McCarthy 1980). It was only in the 1990s that *Bolzano's Program* came into its own with the broad spectrum of non-monotonic reasoning styles that we know today, based on a wide variety of motivations, including default reasoning, resource-sensitive reasoning, and neural computation. And we still have not explored its full depth, which also includes a more dynamic view of the role of changing vocabulary in reasoning (Rott 2001B, van Benthem 2003B).

Logic and methodology Bolzano's book combines what we would call logic to-day with methodology of science. This was quite common in earlier days. Tarski's own basic textbook of the 1940s was called *Logic and Methodology of the Exact Sciences*, and when Tarski and Beth started world conferences in what they saw as their field (Van Ulsen 2000), they initiated the still existing tradition of joint meetings in "Logic, Methodology, and Philosophy of Science". In the meantime, the different component communities have drifted apart – but there are also confluences again, partly through the intermediary of Artificial Intelligence.

Different functions of reasoning Valid semantic consequence $\phi|=\psi$ in predicate logic says that all models of ϕ are models for ψ. Or, in terms of proof, $\phi|-\psi$ says that there is a derivation of elementary steps from ϕ to ψ. By Gödel's completeness theorem, the two notions are extensionally equivalent for first-order logic, and the same holds for many other logical systems. Either way, consequence can serve very different reasoning purposes. In a forward direction, one derives new conclusions, like a mathematician proving new theorems. But in a backward direction, one refutes a hypothesis by deriving some false conclusion: a process of hypothetical reasoning and falsification. A first connection with methodology here is Popper's insistence on the backward process of refutation as more fundamental than accumulating truth. Variety of functions even increases when we consider actual reasoning practices. E.g., in juridical procedure, the prosecutor must show that a given assertion, guilt of the accused, follows from the evidence on the table. But in the same setting, the lawyer must show that the accused's innocence is consistent with that evidence. Thus, inference is deeply intertwined with 'consistency management'. Perhaps the richest view of logical tasks is found with Peirce as early as the 19th century. His distinctions between *deduction* (forward-looking derivation), *induction*, and *abduction* (backward-looking search for premises for given conclusions) fit remarkably well with the many processes that humans engage in (Aliseda-Llera 1997, Flach & Kakas, eds., 2000).

Hypothetico-deductive explanation With this broad view of uses, the methodology of science becomes relevant at once, with its further natural varieties of reasoning such as *explanation* or *confirmation.* These notions occur in science, but they also resonate with ordinary practice. A concrete case is the Hempel-Oppenheim view of explanation. Given a theory T, certain facts F explain a given observation O if

$$T + F \models O, \quad \text{not } T \models O, \quad \text{not } F \models O$$

This is the sense in which the lighting of a match in a garage explains the explosion of the car, given the laws of physics. This simple formulation still suppress a third ingredient in the inference, the 'auxiliary hypotheses' which make sure that the situation described is a 'normal one', fit for standard explanations. E.g., we assume that oxygen is present in a normal garage, and that the car does not run on wood. The new notion is no longer classical consequence. In particular, it is *non-monotonic*: stronger premises F may no longer explain O in T, as they imply O by themselves – and stronger theories T may no longer need F for deriving O. What is interesting here from a logical point of view are two things. We work with structured premises, whose parts play different roles. And we mix logical consequence with non-consequence. But is there any logic here? And is the notion of interest? Carnap worried that any fact O with an explanation F in T also has the trivial explanation $T \rightarrow F$. So, is genuine explanation an art beyond logic? At least there is more to the pattern of explanation.

As for other notions studied in methodology, we note that Hempel 1965 explained confirmation of a law as truth of the lawlike proposition in a *minimal model* of the evidence so far. This is the sense in which successive observations Qd_1, Qd_2, ... would confirm the universal regularity that $\forall xQx$, even without logically implying it. Again we see a divergence from classical consequence, as we now look at just the smallest models of the premises, rather than all of them, to check for the conclusion.

Counterfactuals One very fruitful strand in the philosophy of science has been the analysis of counterfactual conditionals. In thinking about what makes a statement a *law*, as opposed to just an arbitrary generalization, Goodman observed in the 1940s that a genuine law supports *counterfactual assertions* about situations that did not actually occur, such as, "if this match had not been lit, no explosion would have occurred". Carnap observed in the same vein in the 1950s that scientific theories with laws often support *dispositional* predicates, such as 'fragile': a fragile object 'would break if hit', even if it stays whole forever. Counterfactuals are clearly unlike ordinary logical conditionals. In particular, again, they are non-monotonic. If this match had not been lit, but a hand-grenade was thrown, an explosion would have occurred after all. By now, the time was ripe for a generalization behind these notions.

Conditional logic An early appealing account of non-classical conditionals $A \Rightarrow B$ comes from Ramsey 1931. It says this: "add A to your stock of beliefs T, make the minimal adjustment to T to keep the addition of A consistent, and then see if B follows." At the time, this seemed far outside of the logical tradition, but things changed in the 1960s when Lewis and Stalnaker initiated conditional logic. Here is their basic idea (Lewis 1973). Assume that situations or models come ordered by some relation $\leq$ of *relative similarity*. Its source can be a variety of considerations: relative plausibility in one's judgment, objective likeness, and so on. Now we say that

$A \Rightarrow B$ is true at s iff B is true in all A-worlds $\leq$-closest as seen from s

This scheme generates a true logic of valid principles for reasoning with this notion. These include Reflexivity, Conjunction of Consequents, Disjunction of Antecedents, and Upward Monotonicity for Consequents. What fails is, of course, Downward Monotonicity for Antecedents. Instead, on a minimal view of the comparison ordering $\leq$, the only thing that does hold is a substitute, called Cautious Monotonicity:

$$A \Rightarrow B,\ A \Rightarrow C \qquad \textit{imply} \qquad A\&B \Rightarrow C$$

This phenomenon of 'modified rules' would become significant in the 1980s. Conditional logics have been used for many purposes in philosophy, including Lewis' own counterfactual account of causality, and the counterfactual account of knowledge in Nozick 1981. In each case, counterfactual statements express the robustness in inference that a notion like 'law', 'cause' or 'knowledge' is supposed to support.

Nonmonotonic logic in AI The idea that reasoning styles depend on the task at hand returned in strength in *AI* around 1980. McCarthy 1980 argued that our ordinary processes of problem solving make all sorts of systematic additional assumptions, in particular, the use of *minimal models* (in some suitable sense) for the given premises, representing the fact that we are assuming a 'closed world' setting for the task at hand, without surprises. Thus, in *circumscription* (cf. Shoham 1988), we say that

> the conclusion must hold in all *minimal models* of the premises.

Note the similarity with Hempel's view – though Circumscription is much more powerful and sophisticated. Other examples of such styles have come up ever since, including abduction in expert systems and logic programming. Thus, at last, Bolzano's pluralistic view was vindicated, be it not by logicians and philosophers (yet), but by computer scientists. Here 'minimality' is one of the constants of research into logics for a variety of reasoning tasks. It returned in accounts of practical *default reasoning* (defeasible rules of thumb; Reiter 2001, Veltman 1996, Shanahan 1997), and again in the theory of *belief revision* (Gärdenfors 1987), which will be the topic of Section 6.

Structural rules Summarizing two decades of research on all the new-fangled logical systems that have been proposed for dealing with circumscription, default reasoning, and belief revision, it seems fair to say that the core logic of minimality *is* conditional logic. Bolzano's observations about formal properties of different reasoning styles then have their correlate in the deviant properties of classical consequence versus that expressed in conditional logic. The locus of these differences is often put in the so-called *structural rules* governing the over-all properties inference without any special logical operators. Complete packages of structural rules have been found for many new notions of inference, including Bolzano's own (van Benthem 1996A, B, 2003B, 2003E). Deviant structural rules also occur for reasons very different from minimality, however, witness their role in relevant logic, and 'substructural logics' of syntactic combination, dynamic update, or interaction (Dosen & Schroeder-Heister, eds., 1994).

Tricky issues Given all this, is there just proliferation, or can we now begin to chart a neat landscape of natural reasoning styles? No simple map has emerged yet, and some people feel the only way of constraining the options is by asking cognitive science which non-monotonic logics match realities in the mind, or the brain. Another problem is this. Notions that deviate from classical consequence are often more complex in their cost of computing validity or satisfiability. On top of predicate logic, they often become non-axiomatizable, or worse. This seems strange, given their motivation in ordinary practice – and the phenomenon is still not totally understood.

Migrations Even so, our story across logic, philosophy of science, and computer science shows that topics like reasoning styles and conditionality migrate across borders without any problems. This continues until to-day. Van Rooy & Schultz 2004 apply circumscription to the semantics of questions in linguistics (cf. van Benthem 1989), suggesting that the natural logic of Section 3 may also include non-monotonic mechanisms in information exchange. More generally, through these migrations the distinction between scientific method and common sense – once thought so iron-clad – is evaporating. There is just one form of rationality, whether displayed in the kitchen or the Halls of Minerva. And often, it is computer science that brings the viewpoints together in one single perspective (McCarthy 2005). Indeed, computer science plays a role in the development of about every logical theme in the last century. And perhaps this is not so strange. When all is said and done, computer science and its more pregnant form of Artificial Intelligence are just – in Clausewitz' happy phrase – 'the continuation of philosophy by other means'. And the relative neglect of computation and information as fundamental categories in mainstream philosophy (cf. Floridi, ed., 2004) seems a case of rejecting that which is closest to us.

5 Mechanisms of semantic interpretation: from Tarski to 'dynamics'

First-order predicate logic may not be a very realistic description of our natural logic, but it is an amazing model for studying a tightly interwoven set of issues at the same time. The formal language has a clear inductive grammatical structure, and the mechanism of semantic interpretation capitalizes on this in a compositional manner. Thus, it explains the most complex forms of assertion by repeating a small number of elementary truth conditions for separate operators. This setting made it possible to study issues of definition and expressive power in logical model theory. This format of interpretation has been very influential in philosophy, influencing theories of truth through Tarski's classic "Der Wahrheitsbegriff in den Formalisierten Sprachen" from the 1930s (cf. Tarski 1956), and innovative later extensions by Davidson 1967, Kripke 1975, and others. Even so, the laboratory is still capable of generating new ideas!

Standard semantics The key notion in first-order semantics is that of truth of a formula ϕ in a model $\boldsymbol{M}$ under some variable assignment *s:*

$$\boldsymbol{M}, s \models \phi$$

The key compositional clause uses the assignment in an essential manner:

$$\boldsymbol{M}, s \models \exists x\, \phi \quad \text{iff} \quad \text{for some object } d, \boldsymbol{M}, s[x:=d] \models \phi$$

In this way, pulling apart the different ingredients in this scheme, truth is a relation between an expression in a formal language and some language-independent structure of objects with their properties and relations. The relation is mediated by means of two links. One is an *interpretation function I* assigning fixed denotations to predicate letters and other expressions whose meaning is considered constant in the current setting. The other is the *variable assignment s*, providing an auxiliary denotation for variable parts of the expression, which may change in the process of evaluation. Logicians look at this scheme from various angles. Given a formula ϕ, one can study its models $MOD(\phi)$. Given a model $\boldsymbol{M}$, one can study its theory $Th(\boldsymbol{M})$: the sentences made true by it. Or, less common but quite possible, given a formula and a structure, one can study interpretation functions making the formula true in the structure.

Basic features of the set-up This scheme embodies several historical steps, and non-trivial conceptual issues. Its decoupling of formal syntax and semantic evaluation is a major abstraction, as our natural language wears its interpretation on its sleeves, leaving only some freedom in interpreting pronouns and other 'lice of thought' – as Italo Calvino once described them. Next, the notion of a model brings its own issues, as first-order models are a *representation* of reality, not reality itself – except perhaps in some parts of abstract mathematics. Indeed, decades before Tarski, de Saussure proposed his famous Semiotic Triangle of *language, representation, reality*. First-order semantics ignores the latter – but issues of 'fit' to reality do not go away. Indeed, a three-part scheme of interpretation may be found in the work of Church and other logicians. With much more impact, this perspective returned in the 1980s, with the work by Kamp and Heim on discourse representation theory (cf. the survey in van Eijck & Kamp 1997). Of the many further features of the scheme, let us just note the compositionality once more. Dummett 1973 called this Frege's major insight – though modern scholarship has shown that Frege actually never formulated this principle,. Instead, he did place a lot of emphasis on a principle of *Contextuality*, stating that meanings of expressions are never to be studied in isolation from their context. Hodges 2001 is a delightful analysis of the connections between the two principles.

Philosophical and linguistic aspects An amazing amount of philosophical literature is tied directly to the logical semantics for which the above scheme is a paradigm. Some classics are Davidson & Harman, eds., 1972, Putnam 1975, Field 1972, while more modern perspectives are in Etchemendy 1990, Zalta 1993. Some of these philosophical issues returned more concretely in the semantics of natural language, such as Montague's pioneering work that was mentioned in Section 3. E.g., concerns with compositionality haunt that literature until to-day, with discussions fed by the study of concrete 'non-Fregean' quantifier combinations that resist easy iterative decomposition (Keenan &Westerståhl 1997). Also, the status of the models becomes more pregnant here, as we need to take ontological decisions. An unresolved case is the ubiquitous co-existence of count and mass quantifiers, and thus, the interplay of discrete and continuous objects in discourse and reasoning.

Proliferation of frameworks In the 1980s and 1990s, the reign of classical semantics was challenged by a host of new semantic frameworks. *Situation semantics* (Barwise & Perry 1983) emphasized that meaning involves many situations, rather than just one model ***M***: including a situation of utterance, a described situation, and a resource situation. Thus, a true account of semantics involves a network of small contexts. This perspective is coming to the fore these days in many disciplines. Modern inter-disciplinary conferences on context (CONTEXT 2005) bring together philosophers, linguists, and computer scientists with overlapping agendas. Next, the above-mentioned *discourse representation theory* is still a major paradigm for meaning in linguistics and computer science, as discourse representation structures are the natural format for computational processing. In particular, the emphasis now shifts to the dynamics of interpretation: we understand a sentence, not by evaluating it in some given model, but rather by constructing some sort of representation of the described situation. Dynamics of interpretation is also crucial to other computational paradigms, such as 'parsing as deduction' (Pereira & Warren 1983), or 'interpretation as abduction' (Hobbs et al. 1990). The *Handbook of Logic and Language* (van Benthem & ter Meulen, eds., 1997) documents this amazing outburst of semantic creativity, little of which has yet penetrated to philosophical discussions of truth and meaning. There are even other new approaches – of which we discuss one in a little more detail. It demonstrates the proper use of the 'first-order laboratory': not as a source of orthodoxy, but as a source of dissent.

Dynamic semantics An idea common to many newer semantic frameworks is that semantics is really about *dynamic procedures* of evaluation, rather than just static relationships between formulas and models. Can we make these actions themselves an explicit part of the logical apparatus? A well-known proposal of this sort is *dynamic predicate logic* (*DPL*; Groenendijk & Stokhof 1991). The motivation is provided by the phenomenon of anaphora, i.e., how we give coherence to what we say. We study this in a simple first-order laboratory setting. In the grand tradition of 'misleading form', students are usually taught some translation folklore to make first-order formulas fit natural language forms. Consider the following statements:

1	*A man came in. He whistled.*	The two underlined phrases can co-refer.
2	*He whistled. A man came in.*	The two underlined phrases cannot co-refer.
3	*If a man came in, he whistled.*	The two underlined phrases can co-refer.

A direct translation $\exists x\, Cx\, \&\, Wx$ for 1 does not give the intended scope, and so we use a bracketing trick to get $\exists x\, (Cx\, \&\, Wx)$. A translation $Wx\, \&\, \neg\exists x\, Cx$ for 2 does give the right scope, and hence no tricks must be used here. But again, a direct translation $\exists x\, Cx \rightarrow Wx$ for 3 does not work, and so students are taught a bracketing trick plus a quantifier-change to $\forall x\, (Cx \rightarrow Wx)$. This seems terribly unprincipled and ad-hoc. By contrast *DPL* changes the mechanism of first-order semantics, reading formulas ϕ as *procedures*, whose meanings are transition relations between variable assignments *s*, viewed as temporary states of the evaluation process. This can be done entirely compositionally, without changing either the first-order language or its models. This illustrates the more complex semantic world since Montague, where ideas can travel between philosophy, linguistics and computer science. The procedural view comes from the semantics of imperative programming languages. In that setting, expressions in programs are instructions for successive changing of states, viewed as assignments of data-objects to variables. Think of registers filled with transient objects. Executing an imperative instruction $x{:=}2$ replaces the current content of register x with a value 2. More precisely, *DPL* records traces of successful executions in the following format:

$\boldsymbol{M}, s_1, s_2 \models \phi$ iff there exists some successful evaluation of ϕ in $\boldsymbol{M}$ starting from input state s_1 and ending in output state s_2

The semantic clauses treat atomic formulas as tests on the current assignment, conjunction becomes composition of actions, negation is a failure test, and the real dynamics is in the clause for the existential quantifier, which looks for a witness:

$\boldsymbol{M}, s_1, s_2 \models \exists x$ iff $s_2 = s_1[x:=d]$ *for any object* d *in the domain*

When read in this dynamic way, the direct translations of all three examples above come out just as they should without re-bracketing tricks. While this may look like just a new way of doing natural language semantics, the scheme really challenges many presuppositions in logic and philosophy. For instance, meaning and inference now reflect an *algebra of procedures*. E.g., the scope widening in $\exists x\ Cx\ \&\ Wx$ is just *associativity* of composition. Or with a *universal quantifier* $\forall x\ \phi = \neg\exists x\neg\phi$, evaluation procedures for $\exists x\ Cx \rightarrow Wx$ and $\forall x\ (Cx \rightarrow Wx)$ turn out algebraically equivalent. Indeed, *DPL* comes with its own notion of dynamic consequence in which processing the successive premises validates the conclusion. This is again a non-standard, and non-monotonic reasoning style (cf. Section 3) – for which we refer to the literature.

More radical dynamics The general idea of dynamic semantics is that meaning involves *information change*, an idea known in discourse representation theory as 'context change potential'. A more radical versions of this is *game-theoretic semantics* (Hintikka & Sandu 1997). Now, evaluation is no longer just a procedure for one 'checker', but a *two-person game* between a Verifier and a Falsifier with opposing roles concerning the given formula. The former tries to show that it is true, the latter that it is false. In that way, *interaction* becomes essential to understanding meaning and inference. In his many publications on the subject, starting from the 1960s, Hintikka has spelled out a wide range of philosophical implications of taking this stance. Even more radically game-theoretic views of interpretation, partly along the lines of Grice 1975, are coming up these days, as semantic research focuses on *interpretative equilibria* sought by speakers and hearers in actual communication (Lewis 1969, Parikh 2001, van Rooy 2004).

Languages again Contemporary semantic frameworks involve ideas from logic, linguistics, and computer science – again demonstrating the unity of perspective across natural, formal, and programming languages. What we have said so far by no means exhausts this potential. For instance, van Benthem 1996, 1999A propose yet another procedural reinterpretation of first-order semantics! It uncovers a decidable modal base logic of compositional evaluation, which includes all the monotonicity reasoning of Section 3. On top of this, further validities express geometrical constraints on the flow patterns of assignment change. The undecidability of first-order logic, often taken as an inevitable hallmark of reasoning with quantifiers, then becomes a side-effect of working with special computation spaces. In the light of these modern developments, many philosophical discussions of first-order logic and the import of its classical meta-theorems seem outdated, and hopelessly conservative.

An interesting issue is what all this says about our understanding of natural language, the vehicle of both common sense and technical philosophy. Changing information states of readers and hearers is definitely a major purpose of any type of discourse, including philosophical dialogue. If we take this dynamic perspective seriously, some people feel it may be the computer science perspective that should have the last word: by viewing natural language as *the programming language of cognition*. Or, if this slogan is too crude for serious philosophy, here is another way of phrasing the point of this section. Meaning is not just *how things are*, it also essentially involves – as Wittgenstein already said – *what we do*.

6 Theories and belief revision: from neo-positivism to computer science

The dynamic turn A computational stance involves a *tandem view*. Representations should always be constructed with some process in mind, while processes should be designed with some data structure in mind. In a sense, this happened even in the early days of logical analysis (cf. Section 4). After all, words like 'explanation' denote a dynamic activity as much as a static proposition or relation between propositions. One conspicuous development over the past few decades has been a Dynamic Turn (van Benthem 1996), moving beyond static relations between propositions on the pattern of logical consequence to activities like proving, testing, learning, or changing one's mind, as independent objects of study. In this perspective, traditional interfaces between logic and philosophy acquire a double aspect: what are the activities involved, and what representations do they work with? The dynamic semantics of Section 5 is one concrete illustration of such a move, but there are many others. In this section, we apply this style of thinking to the scientific theories that briefly surfaced in Section 4.

Theory statics The simplest logical view of a theory is just as a set of sentences. This, of course, will not do as an account of any plausible notion of theory in the empirical sciences – and not even as an account of the 'structured opinions' which people have in the realm of common sense. Richer views of theories developed in the philosophy of science, witness the work of Ramsey 1931, Beth 1949, and Przelecki and Wojcicki (cf. Przelecki 1969). There is a lucid summary in the Introduction to Suppe 1977, which covers both syntactic views in the neo-positivist tradition, and more semantic ones in the style of Suppes and Sneed. Kuipers 2000 is an up-to-date exposition of modern notions of theory, including an account of the major *inter-theory relations* that have been of interest to philosophers, such as various kinds of extension and reduction. Theory structure is an excellent theme by itself as a running interface between logic and philosophy (cf. Pearce & Rantala 1984, van Benthem 1982). In line with all our themes, it has surprising connections to computer science, in particular the area of 'abstract data types' (van Benthem 1989) – but we forego this particular story here. For, theories by themselves are just static representations of concepts and information at the disposal of individuals or communities. In line with our main interest, we want to see how such theories grow, and more generally: *change*.

Theory dynamics Some of the most spectacular philosophy of science is about ways in which theories change. Sometimes, there is cumulative piece-meal growth, but occasionally, there are more drastic paradigm shifts, as described in Kuhn 1962. Processes of theory change were made into serious objects of logical study in the work of Alchourron, Gärdenfors, & Makinson 1985, Gärdenfors 1987, Rott 2001A. The simplest process is *update with consistent new information*, obtaining a new theory $T+\phi$. This adds the new formula ϕ to the current theory T, and then closes the result deductively. Alternatively, viewing the theory as a set of models $MOD(T)$, it forms the intersection $MOD(T)\cap MOD(\phi)$. But when the new information ϕ is in conflict with the current theory T, this no longer works, and we must form a possibly more complex *revision* $T*\phi$. The latter process is quite common, also in ordinary life. When engaged in non-monotonic reasoning, we may have drawn a temporary conclusion on the basis of just reasoning about minimal models representing some 'normal' scenario. Now something abnormal occurs, and we must accommodate this. Then some of the earlier conclusions no longer hold, and we seek a new equilibrium. Intuitively, revision seems a much more complex process than update, and also, it seems less deterministic. What the new theory will look like depends on what we wish to preserve from the old theory.

Belief revision theory *AGM*-style belief revision theory is an abstract postulational account of three basic operations on theories: update, revision, and *contraction* $T - \phi$. The latter removes a proposition from a given theory, and tries to keep the resulting theory as much like the original as possible. The three operations are intertwined:

$T*\phi$	$=$	$(T - \neg\phi) + \phi$	Levi Identity
$T-\phi$	$=$	$T \cap (T*\neg\phi)$	Harper Identity

Whether these are *all* plausible operations on theories remains an interesting question. But even with these three, a large literature has sprung up on plausible postulates and concrete mechanisms for performing these operations. Also, the framework has been re-described in various ways, e.g., in terms of Quinean 'entrenchment relations' encoding an agent's willingness to give up one proposition in *T* rather than another. Similar views of 'structured theories' have come up in computer science when dealing with data bases that may need to be revised in some systematic manner (Ryan 1992). The idea that the three operations $\{+, *, -\}$ should yield a unique output may seem strange, as there are usually different options for accommodating conflicting evidence. Accordingly, non-deterministic relational alternatives have been proposed in Lindström & Rabinowicz 1992. At the least, there should be room for different belief revision *policies*, corresponding to types of behavior: ranging from more 'radical' to more 'conservative' in hanging on to what was there in the old *T*.

Conditional logic again Again, there is a remarkable continuity in ideas. The above Ramsey test for conditionals already contained an element of belief revision, as antecedents conflicting one's current stock of beliefs needed to be 'accommodated'. The applicability of the Ramsey test for belief revision has been questioned, but it has become clear that there is nothing wrong with it – pace the 'impossibility results' of Gärdenfors and others (cf. Leitgeb 2005). Indeed, Ramsey-style axioms are a key ingredient in designing modern logics of belief change (Gerbrandy 1999), without doing any harm in that setting. The connection with conditional logic becomes even clearer in the semantic account of belief revision in Grove 1988. Viewing a theory as its set of models *MOD(T),* forming a revision $T*\phi$ means passing to the set of *T*-closest worlds in $MOD(\phi)$, with closeness given by some Lewis-style ordering of theories – or underneath: a similar ordering of worlds. One easily recognizes the key minimality clause underlying the semantics of conditionals. Indeed, conditional statements may be viewed as parts of our current theory, 'pre-encoding' our tendencies to revise our beliefs when confronted with new evidence.

Computer science and AI Concrete instances of belief revision occur in computer science, with data base maintenance, or belief update in multi-agent systems. Not just conditional logic, but non-monotonic reasoning in general requires belief revision as a necessary anti-dote to its 'jumping to conclusions'. Further topics that have been introduced along these lines are the mixture of belief revision and factual update about real changes that have taken place in the world (Katsuno & Mendelzon 1991).

Learning theory Belief revision, in its guise of belief adaptation, is actually close to a broader recent stream in the philosophy of science. Perhaps the key activity tied up with theory change is *learning*, whether by individuals or whole communities. Modern learning theory (Osherson, Stob & Weinstein 1988, Kelly 1996) describes learning procedures over time, as an account of scientific methods in the face of steadily growing evidence, including surprises contradicting one's current conjecture. In this perspective, update, revision, and contraction are single steps in a larger process, whose temporal structure needs to be brought out explicitly (Kelly 2002, Hendricks 2002). Learning theory is itself a child of recursion theory, and hence it is one more illustration of a computational influence entering philosophy.

Further sources of dynamics There are many other aspects to update and revision. In particular, another fundamental aspect of the Dynamic Turn is the social interactive character of many relevant activities: dialogue, communication, or – for that matter – scientific research. We will look at some of these in Section 7. To conclude, we return to the simple world of common sense and natural language described in Section 3. Both update and revision occur all the time in natural discourse. But even this simple setting shows that more is the matter than just automatic *AGM* operations. If I hear some ϕ which contradicts my beliefs so far, then I will need to decide whether to accept the information ϕ, or reject it, or reinterpret it. And a typical way of doing that is to enter a dialogue with ϕ's source, whether a person or Nature. Belief revision usually involves communication and even negotiation. Processes of the latter sort are, again, the subject of Section 7 below. Another basic feature of revision in common sense settings is the phenomenon of *language adjustment*. The medieval adage was already that "in case of a contradiction, make a distinction". We resolve contradictions either by retracting previous beliefs, or by changing their formulation. New parameters may occur, or new distinctions in the domain of objects. Perhaps you are rich in spirit, while not being rich in material goods. Perhaps Cantor's comprehension axiom creates classes, but not sets. Weinberger 1965 contains an excellent historical and logical survey of basic responses to new information contradicting existing beliefs. In particular, the idea of language change goes beyond standard theories of belief revision. It seems closer to the more dramatic shifts in conceptual frameworks of scientific theories under pressure from recalcitrant facts, as described by Kuhn. Rott 2001B has a discussion of belief revision in the presence of language change.

From secure foundations to intelligent repair Belief revision is one more current arena where philosophy meets with logic, language, and computation. Add modern theories of learning to the picture, and we have one of the liveliest interfaces today. In the light of all this, the original emphasis in the foundations of mathematics may have been misguided. Frege wanted absolute security, otherwise "mathematics will come down like a house of cards". Gödel's results show that such security cannot be had, except in tiny corners. But the key phenomenon in understanding rational behaviour does not seem to lie here anyway. Nobody believes that the discovery of a contradiction in a mathematical theory is the end of the world. In fact, every time this has happened (and it does happen), a misunderstanding was exposed, a better and richer theory was found, and a deeper understanding resulted. What seems crucial about us is not the use of infallible methods, but reasoning with whatever means we have, plus an amazing facility for belief revision, i.e., *coping with problems* as they arise. It is this dynamic feature of human rationality which logicians and philosophers should try to understand better. Logic is not some vaccination campaign eradicating all diseases once and for all. It is rather *the immune system of the mind*!

7 Dynamic logic, communication, action, and games

A logical consequence is often described as involving a conclusion which 'adds no new information to the premises'. Indeed, logic is deeply tied up with extracting and modifying information, as well as knowledge and other epistemic attitudes arising in these informational processes. Defining a precise notion of 'information' serving all logical purposes is no trivial task (cf. Adriaans & van Benthem, eds., 2005, *Handbook of the Philosophy of Information*). There are several different strands in the literature (cf. Carnap 1947, Barwise & Seligman 1995), whose connections are not always obvious (van Benthem 2005B). The story-line in this section follows one of these, viz. information as a range of options for the actual world that are still compatible with what agents know. This was the original idea in Carnap's *Meaning and Necessity*, and it became central later on in the classical work on epistemic logic (Hintikka 1962). Epistemic logic is an interesting place to start a story line, since many people consider it a typical case of a dead-end as an interface between logic and philosophy. This is as far from the truth as it gets, and we give a short summary of van Benthem 2005A.

Statics: epistemic logic and information states Epistemic logic started as an account of the philosophical notion of 'knowledge', for which we have famous accounts by Plato or Descartes. The second half of the 20th century added a host of interesting alternatives, including contributions by Gettier 1963, Hintikka 1962, Dretske 1981, and Nozick 2001. Hintikka's main idea was this. An agent *knows* those propositions which are true in all situations compatible with what she knows about the actual world:

$$\boldsymbol{M}, s \models K_i\phi \quad \text{iff} \quad \text{for all } t \sim_i s: \boldsymbol{M}, t \models \phi$$

New information will tend to shrink this range, perhaps until the actual world is all that is left open to our scrutiny. There are similar accounts of *belief*, and other epistemic attitudes. The power of this formalism shows in simple scenarios. Consider the simplest case of genuine information flow between agents:

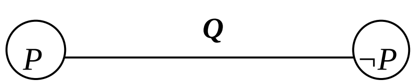
Q asks a question *"P?"*, and ***A*** gives a true answer "Yes".

If this is all honest Gricean communication, then the question itself indicates that the ***Q*** does not know the answer (base-level factual knowledge), but also, that she thinks ***A*** might know (knowledge about other people's information). The answer conveys that that Answerer knows that *P*, and its public announcement in the group *{**Q**, **A**}* makes sure that ***Q*** now also knows that *P*, that both agents know this about each other, and so on. In terms of Lewis 1969, following the episode, the agents have achieved *common knowledge* of *P*. Epistemic logic as it stands can describe all relevant epistemic properties at the three stages of this process: before, during, and after. And of course, it can do a lot more, witness the further literature on this subject (Hendricks 2005).

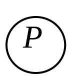
Now, in this area, one of the most spectacular 'crossings' took place. First, in the mid-1970s, epistemic logic emerged independently in game theory (Aumann 1976, Osborne & Rubinstein 1994), as economists developed epistemic scenarios for justifying the usual Nash equilibria of strategies. The idea was to 'lock' players 'into' equilibria in terms of mutual expectations (via knowledge or belief) concerning each other's rationality. Also, by the 1980s, epistemic logic crossed over to computer science (Fagin, Halpern, Moses & Vardi 1995). Here it was the evident power of human-style models for understanding protocols in distributed systems that started a vast literature at the interface of computer science, philosophy, and game theory. It is about agents who (want to) know things, and communicate accordingly. The TARK conferences (*Theoretical Aspects of Rationality and Knowledge*) have brought together these three communities since 1983, and the new interdisciplinary journal *Knowledge, Rationality, and Action* shows the contact is still very much alive.

In this section, we just sketch one particular strand in this story, tying it up with the logical dynamics that emerged in Section 6. Knowledge of a proposition comes about by certain *actions*, be it an inference, asking or answering a question, or performing an observation of some event in general. As before, the structure of the knowledge and the nature of those actions are intertwined. Thus, to really get at what happened in our question-answer scenario, we must 'dynamify' standard epistemic logic!

Updating information Speech act theory in the philosophy of language tells us that linguistic communication triggers systematic changes in information states of language users. For instance, the above question-answer episode might start from the following smple epistemic model, where the line between the two alternative worlds indicates my uncertainty:



Since you have no uncertainty lines in either world, you know whether *P* throughout, and I actually know that you know, as it is true in both of my alternatives. Now, your answer to my question *updates* this model, ruling out the ¬*P*-world, to obtain:



Here both you and I know that *P*, and we know this about each other, and so on to further mutual stages: *P* has become common knowledge between the two of us.

Notice what happened here: epistemic logic now has to deal with dynamic actions, while there is also an irreducible multi-agent 'social' character to the setting.

Dynamic-epistemic logic Both social and private epistemic actions can be described as first-class citizens in dynamified epistemic logics that take a leaf from computer science. Our epistemic actions that eliminate worlds are really model-changers, and the this process involves constant changes in truth value for relevant assertions. Similar phenomena occur with computer programs, recipes, or any sort of structured action. Such phenomena have been studied in *dynamic logics*, whose languages involve mixed assertions with action expressions and propositions. In particular, a dynamic modality $[a]\phi$ says that, after every successful execution of action or program *a* from the current state of the world, or our device, ϕ is the case. In this style, we would get

$[A!]\phi$ after public announcement of *A*, formula ϕ holds.

This mixed proposition/action language can express epistemic effects such as

$[A!]K_jB$ after public announcement of *A*, agent *j* knows that *B*

$[A!]C_GA$ likewise, *A* has become common knowledge in the group of agents *G*.

There are complete and decidable dynamic-epistemic logics of this kind, allowing us to reason systematically about the epistemic effects of assertions. Here is an example of a valid 'update law' – which is as central to dynamic-epistemic logic as an introspection principle would be to static epistemic logic:

$$[A!]\, K_i\, \phi \quad \leftrightarrow \quad (A \rightarrow K_i\, [A!]\, \phi) \qquad \#$$

Note the intellectual history concentrated in formulas like this. Speech acts come from the philosophy of language and linguistics, epistemic logic from epistemology, and their combination in a dynamic logic comes from computer science. This harmonious encounter is a perfect illustration of the story lines of this paper, and the axiomatic calculus suggests new contacts between these fields. E.g., philosophers of action will note that the axiom # expresses an interchange between knowledge that we have beforehand about an action and knowledge that we have after the action. This is a non-trivial feature, which a game-theorist would call *Perfect Recall*: public announcements are epistemically transparent. By contrast, e.g., an action like drinking would not satisfy this axiom. I know that drinking will make me boring – but after drinking, alas, I do not know that I am boring. Thus, the dynamic turn in epistemic logic greatly enriches the agenda of interesting issues for philosophers, allowing them to move beyond worries about Omniscience, Introspection, and the other usual suspects.

More complex communication over time Dynamic-epistemic logics do not just deal with single assertions. They can also handle complex conversational instructions involving typical program structures, such as "First say this, then that", "Say 'Thanks' if you are grateful, and 'No thanks' otherwise", "Keep saying nice things until the dean grants your promotion", or "Answer all together". Finally, they can deal with much more complex, and realistic, communicative actions which involve hiding of information, or partial observation. This includes whispering in lecture rooms, looking at your hand in a card game, or many further informational moves in games that people play. Some basic sources are Baltag, Moss & Solecki 1998, van Ditmarsch, van der Hoek & Kooi 2005, van Benthem 2002B, 2005E. In particular, dynamic-epistemic logic can deal with misleading actions (hiding, lying, cheating) as well as truthful ones – leading to a classification of epistemic action as to expressive power and thresholds of complexity. Also, the framework links up with *belief revision*, *learning*, *probabilistic update*, as well as infinite processes and evolution over time.

Philosophy again Update logic may seem new-fangled, but it connects up with well-established issues in philosophy. Van Benthem 2004 discusses the age-old Moore-sentences of the possibly true but pragmatically infelicitous form "*p*, but I do not know it". These turns up in a surprising new setting in epistemic dynamics. It seems evident at first that public announcements of true propositions. *A* lead to common knowledge of *A*. Indeed, this seems to be the purpose of public announcements, as described by speech act theorists. But now consider an assertion like

"you don't know if *p*, but *p* is true".

This can be perfectly true, and yet its announcement makes *p* common knowledge, thereby invalidating the first conjunct. Thus, this proposition is 'self-refuting'. This failure may seem like a trick, but announcements of uncertainty leading eventually to knowledge occur quite often in famous puzzles, such as The Wise Men, or the Muddy Children. True but self-refuting assertions are also at the heart of current discussions of the Fitch Paradox (Brogaard & Salerno 2002) threatening the Verificationist Thesis that "everything which is true can be known".

More generally, dynamic-epistemic logics shift the traditional boundary between semantics and pragmatics in interesting ways. This demarcation issue is on the agenda in both linguistics and philosophy of language these days. Another interface issue concerns the topic of theory change in Section 7. The internal information flow of a scientific community can be analyzed now, not just the theories produced by that community. What are the systematic relationships between the short-term multi-agent dynamics of scientific debate and the long-term dynamics of theory change?

Games In its current manifestations, logical dynamics is making one more move, toward *game theory*. Both knowledge and communication really make sense only against the background of goal-directed interactive social processes between rational agents. Van Benthem 1999-present, 2005D provide an extensive panorama of how games interface with logic. This includes 'logic games' for argumentation (Lorenz & Lorenzen 1978) and semantic evaluation (Hintikka 1973), showing how these core tasks of logic may be cast as interactive processes between different agents. But the interface also involves modern 'game logics' that have been used to throw light on the rationality of players in general games (Aumann 1976, Stalnaker 1999 and earlier publications, Pauly 2001, de Bruin 2004). In these game logics, many of our earlier systems come together: dynamic logic, epistemic logic, but also logics of counterfactuals, and belief revision. Games started in philosophy with Wittgenstein's language games, and they were advocated vigorously by Hintikka through his 'game-theoretic semantics' and '*IF*-logic' (Hintikka & Sandu 1997). But it seems we have only seen the beginning of their true impact in a broader sense (cf. Hendricks 2005).

Thus again, we see how the current philosophical agenda could be enriched by fully absorbing the fundamental nature of multi-agent information and interaction.

9 Further themes

The themes in this paper are just a selection from a much longer list, including several survey papers written by this author in other volumes. Other story lines linking logic, philosophy and computer science for which the material lies at hand include the following. The references given are just pointers – not a complete portal to the fields:

(a) *structure of scientific theories* (cf. the list of 'Semantic Parallels'
in van Benthem 1989, which is still only partially explored),
(b) *knowledge and belief* (van Benthem 2005A),
(c) *information and computation* (van Benthem 2005C),
(d) *probability*,
(e) *temporal logic* (van Benthem 1995),
(f) *deontic logic, norms and obligations* (van Hees & Pauly 2004,
as well as the Proceedings of the *DEON* conferences, operating
since the late 1980s at the interface of computer science and logic)
(g) *philosophy of action* (Belnap et al. 2001),
(h) *infinite processes and dynamical systems* (Skyrms 2004),

(i) *optimality and evolutionary games* (van Rooy 2004).

Thus, the number of strands connecting logic and philosophy is large, and the fabric of the interface is hard to tear accordingly.

10 Conclusion: logic and philosophy once more

The upshot of this paper is simply this. The history of interfaces between logic and philosophy is rich and varied, especially when we describe it through themes, rather than formal languages or systems. Our story lines are just sketchy first illustrations, but they demonstrate the broad spirit. Before we can have any considered opinion of the current state of affairs between logic and philosophy, we need such a rich view of their past and present interactions. The resulting picture may not be quite like received opinions as to what are the most crucial topics – but it seems definitely much richer than what would be found in any standard text book. It may even help get rid of some internal research programs in 'philosophical logic' that seem to have run their course.

As for more practical conclusions, past meetings between logic and philosophy have been successful, and there seems every reason for pursuing these contacts. Moreover, these contacts are often enhanced through a third party, viz. computer science, which acts as a laboratory for continuing philosophy 'by other means'. Interdisciplinary conferences like TIME, CONTEXT, or TARK lead the way in this menage à trois.

But perhaps most of all, we have written on purpose in a non-conventional style. Much intellectual history seems to make what happened Platonic, heroic, inevitable, and thus: hard to identify with. But our founding fathers were just ordinary people like us, with all sorts of options open to them, including chance encounters, and outside events – and mathematical logic or philosophical logic, though shining with the halo of time, are still open to significant changes, both in terms of interpretation and in terms of real change. Just read Stelzner 1996 on how the optical firm of Carl-Zeiss Jena sponsored Frege's tenure case (with the physicist Ernst Abbe as a guardian angel), and your view of the birth of modern logic, the genesis of *Begriffsschrift*, and its famous metaphor of Microscope versus Eye, will never be the same again!

Acknowledgment I would like to thank Hannes Leitgeb, Gabriel Sandu, and Frank Veltman for their useful and congenial comments.

11 References

P. Adriaans & J. van Benthem, eds., 2005, *Handbook of the Philosophy of Information,* Elsevier, Amsterdam, to appear.

C. Alchourron, P. Gärdenfors & D. Makinson, 1985, 'On the Logic of Theory Change: Partial Meet Contraction and Revision Functions', *Journal of Symbolic Logic* 50, 510–530.

A. Aliseda-Llera, 1997, *Seeking Explanations: Abduction in Logic, Philosophy, and Artificial Intelligence*, Dissertation, Department of Philosophy, Stanford University & ILLC, University of Amsterdam.

R. Aumann, 1976, 'Agreeing to Disagree', *Annals of Statistics* 4, 1236–1239.

A. d'Avila Garcez, K. Broda & D. Gabbay, 2002, *Neural-Symbolic Learning Systems: Foundations and Applications,* Springer, London.

A. Baltag, L. Moss & S. Solecki, 1998, 'The Logic of Public Announcements, Common Knowledge and Private Suspicions', *Proceedings TARK 1998,* 43–56, Morgan Kaufmann Publishers, Los Altos.

E. Barth, 1974, *The Logic of the Articles in Traditional Philosophy*, Reidel, Dordrecht.

J. Barwise & J. Perry, 1983, *Situations and Attitudes*, The MIT Press, Cambridge (Mass.).

N. Belnap, M. Perloff & M. Xu, 2001, *Facing the Future,* Oxford University Press, Oxford.

J. van Benthem, 1981, 'Historische Vergissingen? Kanttekeningen bij de Fregeaanse Revolutie' (*Historical Errors? Marginal comments on the Fregean Revolution in Logic*), *Kennis en Methode* 5:2, 94-116.

J. van Benthem, 1982, 'The Logical Study of Science', *Synthese* 51, 431-472.

J. van Benthem, 1985, 'The Variety of Consequence, According to Bolzano', *Studia Logica* 44:4, 389-403.

J. van Benthem, 1986, *Essays in Logical Semantics*, Reidel, Dordecht.

J. van Benthem, 1987, 'Meaning: Interpretation and Inference', *Synthese* 73, 451-470.

J. van Benthem, 1989, 'Semantic Parallels in Natural Language and Computation', in H-D Ebbinghaus et al., eds., *Logic Colloquium. Granada 1987,* North-Holland, Amsterdam, 331-375.

J. van Benthem, 1991, *Language in Action*, North-Holland, Amsterdam. Reprinted 1995, The MIT Press, Cambridge (Mass.).

J. van Benthem, 1995, 'Temporal Logic', in D. Gabbay, C. Hoggar & J. Robinson, eds., *Handbook of Logic in Artificial Intelligence and Logic Programming*, Volume 4, Oxford University Press, 241-350.

J. van Benthem, 1996A, *Exploring Logical Dynamics*, CSLI Publications, Stanford.

J. van Benthem, 1996B, Inference, Methodology and Semantics, in P. Bystrov & V. Sadofsky, eds., *Philosophical Logic and Logical Philosophy, Essays in Honour of Vladimir Smirnov,* Kluwer Academic Publishers, Dordrecht, 63-82.

J. van Benthem, 1999A, 'Modal Foundations for Predicate Logic', in E. Orlowska, ed., *Logic at Work*, To the Memory of Elena Rasiowa, Physica Verlag, Heidelberg, 39–54.

J. van Benthem, 1999B, 'Wider Still and Wider: resetting the bounds of logic', in A. Varzi, ed., *The European Review of Philosophy*, CSLI Publications, Stanford, 21–44.

J. van Benthem, 1999–present, *Logic in Games*, Electronic lecture notes, ILLC University of Amsterdam & Department of Philosophy, Stanford University. http://staff.science.uva.nl/~johan/Teaching

J. van Benthem, 2002A, 'Mathematical Logic and Natural Language', in B. Löwe, W. Malzkorn, & T. Rasch, eds., *Foundations of the Formal Sciences II: Applications of Mathematical logic in Philosophy and Linguistics*, Kluwer, Dordrecht, 25–38.

J. van Benthem, 2002B, 'One is a Lonely Number', ILLC Report 2002-07, University of Amsterdam. To appear in P. Koepke & W. Pohlers, eds., *Colloquium Logicum,* ASL Publications, New Providence.

J. van Benthem, 2003A, 'Fifty Years: Changes and Constants in Logic', in V. Hendricks & J. Malinowski, eds., *Trends in Logic, 50 Years of Studia Logica*, Kluwer, Dordrecht, 35–56.

J. van Benthem, 2003B, 'Is there still Logic in Bolzano's Key?', in E. Morscher, ed., *Bernard Bolzanos Leistungen in Logik, Mathematik und Physik. Beiträge zur Bolzano-Forschung,* Bd.16, Academia Verlag, Sankt Augustin 2003, pp.11-34.

J. van Benthem, 2003C, 'Logic and the Dynamics of Information', in L. Floridi, ed., *Minds and Machines* 13:4, 503–519.

J. van Benthem, 2003D, 'Logic in Philosophy', electronic course notes, http://staff.science.uva.nl/~johan/Teaching/ , Stanford University.

J. van Benthem, 2003 E, 'Structural Properties of Dynamic Reasoning'. in J. Peregrin, ed., *Meaning: the Dynamic Turn*, Elsevier, Amsterdam, 2003, 15-31.

J. van Benthem, 2003F, 'The Categorial Fine-Structure of Natural Language', Tech Report PP-2003-20, ILLC Amsterdam. In C. Casadio, P.J. Scott, R.A.G. Seely, eds., *Language and Grammar*: *Studies in Mathematical Linguistics and Natural Language*, CSLI Publications Stanford (2005), 3–29.

J. van Benthem, 2004, 'What One May Come to Know', *Analysis* 64 (282), 95–105.

J. van Benthem, 2005A, Epistemic Logic and Epistemology. The State of their Affairs', to appear in V. Hendricks, ed., special issue of *Philosophical Studies* on Bridges between Formal and Mainstream Epistemology.

J. van Benthem, 2005B, 'Information as Correlation and Information as Range', Working Paper, ILLC, University of Amsterdam. To appear in L. Moss, ed., *Memorial Volume for Jon Barwise*.

J. van Benthem, 2005C, 'Information Is What Information Does', draft editorial, *Handbook of Philosophy of Information*, ILLC University of Amsterdam.

J. van Benthem, 2005D, 'Open Problems in Logic and Games', Working paper, ILLC University of Amsterdam, website http://www.illc.nl/~glc To appear in H. Barringer et al., eds., *Festschrift for Dov Gabbay 60*.

J. van Benthem, 2005E, 'Open Problems in Logical Dynamics', ILLC Preprint:, Institute for Logic, Language and Information, University of Amsterdam, DARE electronic archive 148382. To appear in D. Gabbay, S. Goncharov & M. Zakharyashev, eds., *Mathematical Problems from Applied Logic*, Russian Academy of Sciences, Novosibirsk & Springer, New York.

.J. van Benthem & A. ter Meulen, eds., 1997, *Handbook of Logic and Language*, Elsevier, Amsterdam.

E. Beth, 1949, 'Towards an Up-to-Date Philosophy of the Natural Sciences', *Methodos1*, 178–185.

B. Blanshard, 1964, *Reason and Analysis*, Open Court Publishers, La Salle (Ill.).

B. Bolzano, 1837, *Wissenschaftslehre*, Buchhandlung Seidel, Sulzbach. Translated as *Theory of Science* by R. George, University of California Press, Berkeley & Los Angeles, 1972.

B. Brogaard & J. Salerno, 2002, 'Fitch's Paradox of Knowability', Stanford Electronic Encyclopedia of Philosophy, http://plato.stanford.edu/entries/fitch-paradox/.

B. de Bruin, 2004, *Explaining Games*, Dissertation 2004-03, ILLC Amsterdam.

R. Carnap, 1947, *Meaning and Necessity*, University of Chicago Press, Chicago.

CONTEXT, 2005, 5th International and Interdisciplinary Conference on Modeling and Using Context, July 5-8, Paris, France, www.context-05.org.

D. Davidson, 1967, 'Truth and Meaning', *Synthese* 17, 304-323.

D. Davidson & G. Harman, eds., 1972, *Semantics of Natural Language*, Reidel, Dordrecht.

H. van Ditmarsch, W. van der Hoek & B. Kooi, 2005, *Dynamic Epistemic Logic*, Kluwer Academic Publishers, Dordrecht, to appear.

K. Dosen & P. Schröder-Heister, eds., 1994, *Substructural Logics*, Clarendon Press, Oxford.

F. Dretske, 1981, *Knowledge and the Flow of Information*, Chicago University Press, Chicago .

M. Dummett, 1973, *Frege, the Philosophy of Language*, Duckworth, London.

C. Dutilh-Novaes, 2003, 'Medieval Obligationes as Logical Games of Consistency', Philosophical Institute, University of Leiden.

G. Englebretsen, 1981, *Three Logicians*, van Gorcum, Assen.

J Etchemendy, 1990, *The Concept of Logical Consequence*, Harvard University Press, Cambridge (Mass.).

R. Fagin, J. Halpern, Y. Moses & M. Vardi, 1995, *Reasoning about Knowledge*, The MIT Press, Cambridge (Mass.).

H. Field, 1972, 'Tarski's Theory of Truth', *Journal of Philosophical Logic* 69, 347-375.

P. Flach & A. Kakas, eds., 2000, *Abduction and Induction. Essays on their Relation and Integration*, Kluwer Academic Publishers, Dordrecht.

L. Floridi, ed., 2004, *The Blackwell Guide to the Philosophy of Computing and Information*, Blackwell, Oxford – New York.

D. Gabbay & F. Guenthner, eds., 1983–1988, *Handbook of Philosophical Logic*, Reidel, Dordrecht. Second revised and enlarged edition starting from 1996, Kluwer Academic Publishers, Dordrecht.

GAMUT, 1991, *Logic, Language and Meaning*, Chicago University Press.

P. Gärdenfors, 1987, *Knowledge in Flux. Modeling the Dynamics of Epistemic States*, The MIT Press, Cambridge (Mass.).

P. Gärdenfors & H. Rott, 1995, 'Belief Revision', in *Handbook of Logic in Artificial Intelligence and Logic Programming*, Volume 4, ed. by D. M. Gabbay, C. J. Hogger, and J. A. Robinson, Oxford University Press, Oxford.

P. Geach, 1972, *Logic Matters*, Blackwell. Oxford.

J. Gerbrandy, 1999, *Bisimulations on Planet Kripke*, dissertation DS-1999-01, Institute for Logic, Language and Computation, University of Amsterdam.

E. Gettier, 1963, 'Is Justified True Belief Knowledge?', *Analysis* 23, 121–123.

B. Geurts & F. van der Slik, 2004, 'Monotonicity and Processing Load', *Journal of Semantics* 22, 97-117.

A. Gillies, 2004, 'New Foundations for Epistemic Change', *Synthese* 138, 1–48.

H. Grice, 1975, 'Logic and Conversation', in P. Cole & J. Morgan, eds., *Syntax and Semantics III*, Academic Press, New York, 41–58.

J. Groenendijk & M. Stokhof, 1991, 'Dynamic Predicate Logic', *Linguistics & Philosophy* 14, 39-100. Also in *The Philosopher's Annual*, 1992.

A. Grove, 'Two Modelings for Theory Change', *Journal of Philosophical Logic*, 157-170, 1988.

M. van Hees & M. Pauly, 2004, 'Logical Constraints on Aggregating Judgments', Department of Philosophy, University of Groningen & INRIA Toulouse. To appear in the *Journal of Philosophical Logic*.

P. Hempel, 1965, *Aspects of Explanation and Other Essays in the Philosophy of Science*, The Free Press, New York.

V. Hendricks, 2002, 'Active Agents', PHILOG Newsletter, Roskilde. In J. van Benthem & R. van Rooy, eds., special issue on Information Theories, *Journal of Logic, Language and Information* 12:4, 469–495.

V. Hendricks, 2005, *Mainstream and Formal Epistemology*, Cambridge University Press, New York.

J. Hintikka, 1962, *Knowledge and Belief*, Cornell University Press, Ithaca. Reprint 2005, V. Hendricks & J. Symons, eds., King's College Publications, London.

J. Hintikka, 1973, *Logic, Language Games and Information*, Oxford University Press, Oxford.

J. Hintikka & G. Sandu, 1997, 'Game-Theoretical Semantics', in J. van Benthem & A. ter Meulen, eds., *Handbook of Logic and Language*, Elsevier, Amsterdam, 361–410.

J. Hobbs, M. Stickel, D. Appelt & P. Martin, 1990, *Interpretation as Abduction*, SRI Technical Note 499, SRI International, Menlo Park, California.

W. Hodges, 2001, 'Formal Features of Compositionality', *Journal of Logic, Language and Information* 10, 7 – 28.

W. Hodges, 2005, 'Opening Two Windows: Arabic Logic and Cognitive Science', Department of Mathematics, Queen Mary and William's College, London. To appear in D. Gabbay, S. Gontcharov & M. Zakharyashev, eds., *Logics for the XXIst Century: Mathematical Problems From Applied Logic*, Russian Academy of Sciences, Novosibirsk & Kluwer/Plenum Publishers, New York.

Th. Janssen, 1981, *Foundations and Applications of Montague Grammar*, Dissertation, Mathematical Institute, University of Amsterdam.

D. Jacquette, ed., 2002, *A Companion to Philosophical Logic*, Blackwell, Oxford.

H. Kamp & U. Reyle, 1993, *From Discourse to Logic*, Kluwer, Dordrecht.

H. Katsuno & A. Mendelzon, 1991, ''Propositional Knowledge Base Revision and Minimal Change', *Artificial Intelligence* 52, 263–294.

E. Keenan & D. Westerståhl, 1997, 'Quantifiers', in J. van Benthem & A. ter Meulen, eds., *Handbook of Logic and Language*, Elsevier, Amsterdam, 837– 893.

K. Kelly, 1996, *The Logic of Reliable Enquiry*, Oxford University Press, Oxford.

K. Kelly, 2002, 'Knowledge as Reliably Inferred Stable True Belief', Department of Philosophy, Carnegie-Mellon University, Pittsburgh.

S. Kripke, 1975, 'Outline of a Theory of Truth', *Journal of Philosophy* 72, 690 – 716.

Th. Kuipers, 2000, *From Instrumentalism to Constructive Realism*, Kluwer Academic Publishers, Dordrecht.

Th. Kuhn, 1962, *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago.

H. Leitgeb, 2004A, 'Carnap's Logischer Aufbau Revisited', Department of Philosophy, University of Salzburg.

H. Leitgeb, 2004B, *Inference on the Low Level. An Investigation into Deduction, Nonmonotonic Reasoning, and the Philosophy of Cognition*, Springer, Berlin.

H. Leitgeb, 2005, 'Gärdenfors' Impossibility Argument Defused', working paper, Department of Philosophy, Stanford University & Fachbereich Philosophie, Universität Salzburg.

D. Lewis, 1969, *Convention*, Harvard University Press, Cambridge (Mass.).

D. Lewis, 1973, *Counterfactuals*, Blackwell, Oxford.

S. Lindström & W. Rabinowicz, 1992, 'Belief Revision, Epistemic Conditionals and the Ramsey Test', *Synthese* 91,195-237.

K. Lorenz & P. Lorenzen, 1978, *Dialogische Logik*, Wissenschaftliche Buchgesellschaft, Darmstadt.

B. Mates, 1953, *Stoic Logic*, University of California Publications, Berkeley & Los Angeles.

J. McCarthy, 1980, 'Circumscription – A Form of Nonmonotonic Reasoning', *Artificial Intelligence* 13, 27–39.

J. McCarthy, 2005, 'The Philosophy of AI and the AI of Philosophy', Department of Computer Science, Stanford University. To appear in P. Adriaans & J. van Benthem, eds., *Handbook of the Philosophy of Information*, Elsevier, Amsterdam.

R. Montague, 1974, *Formal Philosophy*, *Selected Papers of Richard Montague*, R. Thomason, ed., Yale University Press, New Haven.

R. Nozick, 1981, *Philosophical Explanations*, Harvard University Press, Cambridge (Mass.).

D. Osherson, M. Stob & S. Weinstein, 1988, *Systems that Learn*, The MIT Press, Cambridge (Mass.).

P. Parikh, 2001, *The Use of Language*, CSLI Publications, Stanford.

M. Pauly, 2001, *Logic for Social Software*, dissertation DS-2001-10, Institute for Logic, Language and Computation, University of Amsterdam.

D. Pearce & V. Rantala, 1984, 'A Logical Study of the Correspondence Relation', *Journal of Philosophical Logic* 13, 47-84.

F. Pereira & D. Warren, 1983, 'Parsing as Deduction', *Proceedings of the 21st Meeting of the Association for Computational Linguistics*, 137--144.

S. Peters & D. Westerståhl, to appear, *Quantifiers in Language and Logic*, Philosophical Institute, University of Gothenburg & Department of Linguistics, Stanford University. To appear with Oxford University Press.

M. Przelecki, 1969, *The Logic of Empirical Theories*, Routledge and Kegan Paul, London.

H. Putnam, 1975, *Mind, Language, and Reality,* Cambridge University Press, Cambridge.

F. Ramsey, 1931, *The Foundations of Mathematics and Other Essays,* Kegan Paul, London.

R. Reiter, 2001, *Knowledge in Action. Logical Foundations for Specifying and Implementing Dynamical Systems*, The MIT Press, Cambridge (Mass.).

W-P de Roever, F. de Boer, U. Hannemann, J. Hooman, Y. Lakhnech, M. Poel & J. Zwiers, 2001, *Concurrency Verification: Introduction to Compositional and Noncompositional Methods*, Cambridge University Press, Cambridge.

R. van Rooy, 2004, 'Signalling Games Select Horn Strategies', *Linguistics and Philosophy* 27, 493-527.

R. van Rooy & K. Schultz, 2004, 'Pragmatic Meaning and Non-Monotonic Reasoning: The Case of Exhaustive Interpretation', Working Paper, ILLC University of Amsterdam. To appear in *Linguistics and Philosophy*.

H. Rott, 2001A, *Change, Choice, and Inference*, Clarendon Press, Oxford.

H. Rott, 2001B, 'Theoretical Concepts in Flux: Conceptual Knowledge and Theory Change', in R. Eckardt, K. von Heusinger & Ch. Schwarze, eds., *Words in Time – Diachronic Semantics from Different Points of View*, FB Sprachwissenschaft der Universität Konstanz, 279-301.

M. Ryan, 1992, *Ordered Presentations of Theories: Default Reasoning and Belief Revision*, Ph.D. thesis, Department of Computing, Imperial College.

V. Sanchez Valencia, 1991, *Studies in Categorial Grammar and Natural Logic*, Dissertation, ILLC, University of Amsterdam. To appear with MIT Press.

M. Shanahan, 1997, *Solving the Frame Problem*, The MIT Press, Cambridge (Mass.).

Y. Shoham, 1988, *Reasoning about Change. Time and Causation from the Standpoint of AI*, The MIT Press, Cambridge (Mass.).

B. Skyrms, 2004, *The Stag Hunt and the Evolution of Social Structure*, Cambridge University Press, Cambridge.

P. Smolensky & G. Legendre, 2005, *The Harmonic Mind. From Neural Computation to Optimality-Theoretic Grammar*, The MIT Press, Cambridge (Mass.).

F. Sommers, 1982, *The Logic of Natural Language*, Clarendon, Oxford.

R. Stalnaker, 1999, ‘Extensive and Strategic Form: Games and Models for Games’, *Research in Economics* 53:2, 93-291.

W. Stelzner, 1996, *Gottlob Frege. Jena und die Geburt der Modernen Logik*, Verein zur Regionalforderung von Forschung, Innovation und Technologie, Jena.

F. Suppe, ed., 1977, *The Structure of Scientific Theories*, University of Illinois Press, Urbana.

A. Tarski, 1956, *Logic, Semantics, Metamathematics*, Oxford University Press, Oxford.

P. van Ulsen, 2000, *Evert Willem Beth als Logicus*, Dissertation, Institute for Logic, Language and Computation, University of Amsterdam.

F. Veltman, 1996, 'Defaults in Update Semantics', *Journal of Philosophical Logic* 25, 221–261. Also appeared in *The Philosopher's Annual*, 1997.

P. Wason & P. Johnson-Laird, 1972, *Psychology of Reasoning. Structure and Content*, Batsford, London.

O. Weinberger, 1965, 'Der Relativisierungsgrundsatz und der Reduktionsgrundsatz – Zwei Prinzipien des Dialektischen Denkens', Czecho – Slovakian Academy of Sciences, Prague.

W. Wundt, 1880/3, *Logik. Eine Untersuchung der Prinzipien der Erkenntniss und der Methoden Wissenschaftlicher Forschung*, Ferdinand Enke, Stuttgart.

E. Zalta, 1993, 'A Philosophical Conception of Propositional Modal Logic', *Philosophical Topics* 21:2, 263 – 281.