Logic and philosophy, a sea of stories

Johan van Benthem, http://staff.fnwi.uva.nl/j.vanbenthem

Abstract

The interface of logic and philosophy is diverse, and has always been so. Understanding the interplay raises issues on which some philosophers have strong a priori views. My own preference as a working logician is to proceed by actual history of ideas, something still largely to be written for modern logic. I will present a few lines connecting logic and philosophy around the theme of *implication* and *consequence*, and show how rich the agenda is as it keeps evolving with contributions from many disciplines. This is a light survey paper, but I supply references for readers who want to see more of the nitty-gritty details.

Logic and philosophy

No one knows exactly how logic arose as a scientific discipline around 500 BC, and various hypotheses have been put forward. These hypotheses include a history of gradual reflection on success factors in much older reasoning traditions such as mathematics or the law (dating back to the 3d millennium BC), the tradition of live public intellectual debate in various cultures, or the advent of new recording technologies making it possible to look at pieces of reasoning as codified objects. But what seems indubitable is the long-standing connection of logic and philosophy, prominent with Plato and Aristotle in Greece, and in other forms, also to be found in the Indian and Chinese cultures. Since that age, the relation between logic and philosophy has undergone many changes, where logic has sometimes been seen as a favorite tool of philosophers, but sometimes also, as an intellectual yoke to be thrown off by truly creative minds. However this may be, over the centuries, logic has formed fruitful partnerships with many other disciplines besides philosophy, such as mathematics, computer science, or linguistics – and the list of partners is still growing.

Reflecting this diverse history, there are different views on what logic 'really is', especially among philosophers. Personally, I have little sympathy with such essentialist positions toward anything, let alone my own field, as these seem a remnant of an Aristotelean stance that has been superseded by modern science (Beth 1959). Even though I have worked technically in exploring the issue what makes a notion 'logical' in terms of invariance (van Benthem 2002), I prefer to not 'philosophize' the interface of logic and philosophy. Instead, I would rather advocate a much more down-to-earth approach, namely, charting the history of ideas concerning major notions. What one will find then is the following. Reality is much richer and less predictable than what one might think, there is a large web of stories concerning major notions, and when one travels along these, surprising twists and turns happen on the way, right across different academic disciplines.

I will demonstrate the story-telling approach advocated in this article for one major theme connecting logic and philosophy, namely, views of the notions of *implication* and *consequence* over the last century. Traveling its roads without preconceptions will reveal exciting vistas and connections, often quite surprising, showing how logic connects effortlessly across the academic spectrum – and at the end, we will sit back and draw some conclusions.

There is a lot of material concerning logic and philosophy that would support many further such travels, witness the *Handbook of Philosophical Logic* (Gabbay & Guenthner, eds., 1983–1988), the *Handbook of the Philosophy of Logic* (Jacquette 2006), the *Handbook of the History of Logic* (Gabbay & Woods, eds., 2004–present), the interdisciplinary anthologies van Benthem & Gupta, eds., 2011, van Benthem & Liu, eds., 2013, as well as the many logic-related entries in the *Stanford Encyclopedia of Philosophy* (http://plato.stanford.edu/). Together, these sources extensively document such fields as modal logic, epistemic and doxastic logic, temporal logic, conditional logic, deontic logic, logics of questions, logics of action, and so on. However, the specific theme-oriented, rather than field-oriented approach taken in the present article comes from van Benthem 1989, 2006, where the interface of logic and philosophy is discussed, not in terms of standard areas, but of major *themes* and their natural development over time.

The surplus of implication

Anyone who take a first course in modern logic is exposed to the truth-table account of material implication. On this account, a proposition $\varphi \to \psi$ ("if φ, then ψ") is true if and only if either φ is false or ψ is true. This is only a bleak approximation to implication as one would ordinarily use it in our common natural language, but the truth table view does have a pedigree going back to the Stoics in Greek Antiquity, when it was said that "even the sparrows on the rooftops in Athens are discussing the meaning of implication" (Kneale & Kneale 1961, Bochenski 1961). I do not know what the Chinese birds on the Tsinghua campus are discussing these days, but if they were short of topics, the analysis of implication is still a live subject well-worth their attention.

At the other extreme of the truth table account of implication lies the classical logical notion of *valid consequence*, where we say that ψ follows logically from φ (in notation, $\varphi \models \psi$) if, in every situation where φ is true, ψ is also true. It is standard logical methodology to separate these two extremes severely, and we all grade our students on this, failing them if they confuse the two. An implicational proposition lives in the 'object language', a statement of valid consequence lies in the 'meta-language', as an assertion *about* object-language propositions. But as with many methodological distinctions, things turn out more blurred, and richer, in practice. Much of this article is about views of implication that intuitively lie in between, where $\varphi \to \psi$ has something to do with what happens in some relevant *range of cases*. This theme will return at length below.

Just as a warm-up, and also to shake up readers used to thinking of logic as purely normative, consider how human subjects actually test implications (Wason & Johnson-Laird 1972). There is always a surrounding universe of cases. Suppose we give a set of cards with a letter on one side and a number on the other, and let φ stand for "one side of the card has a vowel", and ψ for "one side of the card has an even number". We show people four cards: *A K 4 7.* Which ones should be turned to check whether the implication holds? Most subjects (including some professional logicians) name only the *A*-card, many fewer give the correct answer that one should also turn the 7-card. We can see this as an issue of what are the relevant cases. Seeing one face of a card is a test where it lies in the spectrum of 4 possibilities {vowel-even, vowel-odd, consonant-even, consonant-odd}. The only counter-example to the proposed rule is 'vowel-odd'. If we see the letter *A*, we learn that we are in the set {vowel-even, vowel-odd} and a further test is needed to rule out the counter-example. If we see the letter *K*, we learn that we are in the set {consonant-even, consonant-odd} which does not contain the counter-example, and no further test is needed. If we see the number 7, then we are in the set {vowel-odd, consonant-odd} which does contain the counter-example, and again we need one more test of turning a card to decide whether the implication holds. [1]

Now let me embark on the story of this paper. I called material implication a bleak – though it is of course at the same time a wonderfully abstract and technically highly useful – approximation of a real implication "if φ, then ψ". Now, what is the surplus of the latter notion, what does it really mean? There is actually a wide variety of notions of conditional assertion in reasoning, even in our common natural language, with a wide family of related terms such as "implies", "entails", or "supports". Therefore, a unique answer need not be expected.

Let me start with a family member from the philosopher's jargon. We could say that an implication holds if the proposition φ is a *sufficient condition* for ψ. This is richer than material truth-table implication, and it can be cashed out in several ways. [2] What extra is involved when we say that one proposition is a sufficient condition for another? Intuitions that have been advanced run in several directions. Authors such as Barwise & Perry 1983 or Barwise & Seligman 1995 have stressed objective informational dependencies in nature that underpin, or are expressed explicitly by true conditionals. One can even go further than this and think of the antecedent `causing' the consequent, though natural dependencies might also just be correlations (van Benthem & Martinez 2008). But the surplus in one proposition entailing another may also be more subjective, referring to the activities of cognitive agents that supply reasons or provide explanations, and in doing so, often engage in argumentation with other agents, whose reasons need not be the same.

These two views of reason-based implication are not in conflict (van Benthem 2015): objective implications are there for subjective human agents to find and wield, while the most effective reasons to use in the long run are those that are based on objective realities. Indeed, a study of the two aspects: objective informational dependency in the world, and humans reasoning, convincing, refuting, or doing all the things intelligent agents do, occur intertwined throughout the history of logic. In line with this observation, many of the accounts of implication that we will discuss below have this double flavor, and can be looked at from both sides.

Even with this broad canvas, however, we are not going to discuss every existing view of what makes an informational dependency an implication. Other famous intuitions include relevance, or degrees of strength as measured by probability. We will encounter yet others as we proceed, but no matter how many views, every journey is selection.

---

[1] Why do human reasoners often forget the latter case? I will not give an explanation, except to say that it leads us straight to well-known basic issues in philosophy, such as the advocacy of refutation rather than proof as the driving force of science in Popper 2002. Perhaps what we see in the Wason task is that learning from errors does not come as naturally to us as philosophers of science may have thought. However do look at the delicate discussion of actual reasoning with implications in van Lambalgen & Stenning 2008 to better understand the subtlety of the cognitive mechanisms at work here.

[2] The term 'sufficient condition' seems to favor φ's point of view as a benefactor supporting others. But we could just as well say that ψ is a *necessary condition* for φ, meaning that ψ sticks to φ like glue. In case the consequence ψ is bad in some way, that would lead us to pay more attention to the refutation cases in the above card example. I will not pursue the second, sometimes more refutation-oriented, perspective systematically in this paper, though many colleagues think it should be there always in logical systems, as a sort of counterpart to the positive forward view of implication (cf. Wansing 1993).

We end this section and all later ones with a 'resting point', taking stock of what we have seen.

Where we stand. The intuitive notion of implication is not quite the implication in truth tables, despite the many technical uses of the latter. However, a mismatch of formalism and intuition need not be a reason for rejection of logical theory, but it can rather be a pushing force for conceptual innovation. In particular, we are led to think about the surplus needed for real implication. And ideas about that come even in two varieties, forming two naturally entangled dual perspectives through all of logic: the study of consequence as an objective link between situations or phenomena in the world, and the study of reasoning by subjective agents.

Extending the range of cases: modal entailment

A major stream in thinking about implication was started in Lewis 1918 who reintroduced the notion of intensionality into logic after its removal by Frege and Russell in their famous systems of mathematical logic. We can think of a classical truth-table implication $\varphi \to \psi$ as expressing just the statement that $\neg(\varphi \wedge \neg\psi)$: that is, the antecedent is not true while at the same time, the consequent is false. To get closer to the intuitive meaning of implication, this statement can be strengthened by adding a modal notion that comes from philosophy: the antecedent *cannot* be true while the consequent is false: $\neg\Diamond(\varphi \wedge \neg\psi)$ – with $\Diamond$ for the modal expression "it is possible that". Equivalently, this is a strengthened modal assertion that the truth-table implication is *necessarily* true: i.e., $\Box(\varphi \to \psi)$, now with $\Box$ for the modal expression "it is necessarily the case that".

This brings us to the area of modal logic, whose most common and simple semantics, going back to Leibniz, reads a necessity modality as a universal assertion over some relevant range of options – as viewed from the current situation (Blackburn, van Benthem & Wolter, eds., 2006, van Benthem 2010). Thus, modal entailment is a parametrized implication, where the additional strength comes from the fact that we can use the regularity in a lot of cases. What these options are can depend entirely on the setting: ontological possible worlds, epistemic alternatives, or perhaps just all cases that form the range of applicability of our current theory.

Thus, the same modal laws represent patterns of reasoning that we expect to recur across very different areas of philosophy, whether endorsed or objected to, and in fact, they do. Two well-known modal patterns are the following two: Weakening $\Box\varphi \to \Box(\varphi \vee \psi)$ and Conjunction $(\Box\varphi \wedge \Box\psi) \to \Box(\varphi \wedge \psi)$. Ontologically, these say that necessary truths are closed under consequence and conjunction, while epistemically, they make the claim that these same properties hold for agents' information.

Modal logic then studies the general laws of reasoning with this strengthened notion of implication, and what its having different interpretations means is the usual virtue of abstraction: the same logical forms will occur across metaphysics, epistemology, or whatever field the model is appropriate for. Whether one considers them valid or not in these areas can be a matter of debate, but then, focusing debates, rather than settling them forever, is in fact an important role of logic.

Where we stand. To deal with genuine implication, we need at least a notion of range of relevant cases beyond the actual world that is intermediate between truth in a single case and validity. This perspective fits well with the standard semantics of modal logic, and hence many theories of implication have a modal flavor. Moreover, the modal stance supplies logical forms, that is, possible patterns of reasoning with implication that works across, and sometimes reveals analogies between what are usually considered quite different areas of philosophy.

Logical properties of implication

But we have already seen in our story that logic does not come with one single methodology toward its objects of study. In the preceding section we have proceeded semantically, reading an implication across a range of objects in some semantic model, representing some relevant part of reality in the reasoning practice under study. In this manner, the philosophically relevant question to ask is whether the sort of models one proposes faithfully represents some philosophical intuition, or more down to earth, some conceptual view. But one can also proceed differently, and thinks in terms of axioms or *postulates*: global properties, or rules as such, that any notion of implication, no matter how defined semantically, should satisfy.

Here are a few properties that many people have considered typical for implication. The first is that any proposition implies itself ('reflexivity'): $\varphi \to \varphi$. The second is an automatic chaining of consequences ('transitivity'): $\varphi \to \psi$ and $\psi \to \chi$ imply that $\varphi \to \chi$. The third property says that adding premises cannot cancel earlier conclusions ('monotonicity'): $\varphi \to \psi$ implies $(\varphi \wedge \chi) \to \psi$. Once we have these general properties about what one might call the 'transmission process' of valid deduction, we can derive further ones purely combinatorially, supplementing informational inspection of semantic models with purely formal proof steps.

These properties look innocent, but they are strong. It can be proved (Scott 1971) that any abstract relation between propositions satisfying reflexivity, transitivity, and monotonicity can be represented as implication in the earlier semantic sense, where $\varphi \to \psi$ says that all φ–situations are ψ–situations. Van Benthem 1996A has many such 'representation theorems' for different sets of constraints on implication relations. But for us here, technical details do not matter, and the main points to note are the following three general observations.

The first point is that we can develop logical intuitions both in a semantic and in a postulational manner, and indeed, both methods are found at the interface of logic and philosophy. The second point is that these two approaches, semantic and syntactic, can be in harmony, witness the result just noted – which is really just a very special case of what a logician would call a 'completeness theorem'. Finally, as a smaller point, the above tells us that any notion of implication that truly wants to generalize from the semantic picture given so far will have to drop at least one of the three stated intuitions. Soon we shall see how this is precisely what has happened in the history of the field.

We conclude by observing that, in cognitive practice, both perspectives play together. When solving a problem, we usually both form a mental picture of the situation, closer to a model for the particular situation at issue, but also appeal to general rules, at least, when we know them about the general domain of cases that our problem belongs to. This interplay seems to have some clear cognitive value qua efficiency, though the balance is ill-understood.

Where we stand. Implication can be studied from two stances that occur throughout logic: providing semantic models for what implications mean, or listing broad postulates on how implications function in proof. These are not conflicting methodologies, since the postulates may constrain the choice of a semantics that validates them, and more generally, the two stances provide two footholds for developing philosophical intuitions about many subjects.

Extended alternatives: counterfactuals and scientific explanation

A next crucial step in modern theories of implication was taken not inside logic itself, but in the philosophy of science. When thinking about what makes a statement in a scientific theory 'law-like' rather than a mere generalization from observed cases, philosophers such as Goodman 1947 noted that there seems to be a further modal force to a genuine law, that goes beyond what is to what could be or what might have been. If I explain the falling to the ground of the cup that I dropped by invoking the law of gravity, I am not merely saying that all objects in this world happen to fall to the earth. I seem to be saying that, even if I had dropped something else like my saucer (which I did not), it, too, would have fallen to the ground. The law even works in a sort of neighborhood of the actual world that did not come about.

This observation led to the study of so-called 'counterfactual conditionals' of the syntactic form "if φ had been the case, then ψ would have been the case", where we know the antecedent to be false, and are yet we are saying something obviously significant. It is strong conditionals like this that drive scientific explanations, but just as well, our common sense justifications in daily life of what we choose to do ("otherwise, this or that bad consequence would have occurred"). [3] Note how this moves the range of relevant cases for an implication from hypothetical, but usually consistent ones to cases that can even turn out demonstrably false. [4]

More generally, there is a connection here with the influential 'deductive-nomological account' of *scientific explanation* proposed by Hempel & Oppenheim in the 1950s (cf. Hempel 1965). We have a scientific theory *T* and an observed fact *E*. We now formulate a hypothesis *H* and show that, given *T, H* implies *E* 'under normal circumstances'. This may be viewed as a generalized implication with a finer description of all aspects that are in play: antecedent facts, but also general regularities. The resulting schema of reasoning is reminiscent of Peirce's notion of abduction, but also of an 'argument schema' in the argumentation theory of Toulmin 1958. [5] There is a consequent *E*, an antecedent *H*, and a background theory *T*, and the force of the implication from *H* to *E* is only across a range of 'normal cases', determining the force of the implication. Thus, it was philosophers of science who opened up a much richer view of what implication and consequence involve. For a study of explanation schemas from the philosophy of science in terms of logic, see van Benthem 1996B, Aliseda-Llera 2006.

Where we stand. Major new ideas in logic can also be triggered by developments in other fields. The philosophy of science is a case in point for the study of conditionals: it introduced seminal new notions such as counterfactuals, and the importance of a restriction to 'normal cases', as well as a sensitivity to the range of reasoning activities, with explanation as a paradigmatic example. More generally, major intellectual figures in the 19th and 20th century still cannot be classified very precisely as being either a logician or a philosopher of science, witness important authors like Bolzano, Mill, Peirce, or Carnap, many of whose works are an inextricable mixture of both. [6]

---

[3] The study of 'dispositional predicates' in Carnap 1967 ("this glass is brittle", meaning it would break if hit, even if it never does get hit in the course of history) points in the same counterfactual direction.

[4] This move does not take us outside of the realm of logical reasoning. The same occurs, of course, when giving a piece of hypothetical reasoning that establishes some conclusion by a reductio ad absurdum.

[5] More can be said about the Toulmin connection: it also concerned the dynamics to be discussed later, in Toulmin's case, the role of 'formalities' in proper juridical reasoning procedure.

[6] The fields of logic and philosophy of science have been closer and then again further apart over the

Conditional logic: from 'all' to only 'the best'

In the late 1960s, counterfactuals inspired a new treatment of implication in philosophical logic, with a crucial new feature beyond a mere range of relevant possibilities. Lewis 1973 and Stalnaker 1968 introduced a comparative *ordering* of worlds (often thought of as 'relative similarity' between worlds) so that it makes sense to talk of the closest worlds to a given world *s*: those leaving as many things as possible normal from the standpoint of *s*. A counterfactual implication $\varphi \rightarrow \psi$ is then true in world *s* if, in all worlds *closest* or 'most similar' to *s* where φ holds, ψ holds as well. [7]

The logic of counterfactuals deviates considerably from the earlier simple classical entailment approach. While reflexivity remains valid, both transitivity and monotonicity fail. For instance, closest worlds among the worlds where a conjunction $\varphi \wedge \psi$ holds need not all be closest worlds where φ holds. However, this does not mean that no valid reasoning principles hold for conditionals generalized in this way. In fact, there are complete axiom systems for conditional logics with varying assumptions on the comparative order (see Lewis 1973 and its large follow-up literature), and even, as in modal logic, the basic conditional logic is decidable.

One interesting fork of the road should be noted here. The subsequent study of counterfactuals has largely moved away from the philosophy of science, and ideas from the above similarity semantics became very influential in the study of conditional expressions in natural language, of which there are quite a few (from 'counterfactual' to 'indicative'). Many innovative ideas in later years came from papers in the area of linguistic semantics, such as the analysis of the meaning of the verbs 'want' and 'desire', important drivers of our actions, by Stalnaker (see Heim 1992 for an influential version) that mixes agents' preferences and beliefs in a conditional-like manner. The same mixture of influences from both philosophy and linguistics was characteristic for many areas of philosophical logic in the 1970s, and it has persisted until the present day. [8]

In particular, the idea that we are not looking at *all* a priori options in the current range, but only at *the best* options as defined by some ordering (closest, or otherwise most relevant or desirable worlds) has emerged independently in many fields of philosophical logic. A particular instance where 'best' really means best is *deontic logic*, the formal study of reasoning about obligations, permissions, commitments and norms. Hansson 1969 is an early treatment of deontic logic based on relative betterness order of worlds, where we must try to realize those things φ that are true in the normatively best worlds (an obligation $O\varphi$). Here, as in conditional logic, the total deontic betterness order also allows for 'conditional obligations' $O^{\psi}\varphi$ with an antecedent ψ. For instance, given circumstances that in themselves might not be morally optimal but just happen to hold, even murderers still have moral obligations, Chisholm 1963. Deontic logic is also closely related to *preference logic* (Hanson 2001, Liu 2011), and indeed, concentrating on the best options, either in an absolute or a conditional sense, is standard in fields such as decision theory.

---

course of the 20th century, but they recently joined forces again in `mathematical philosophy'.

[7] There are some complications with lifting this truth condition to work well on infinite models, but these can be solved in various ways, and hence the issue of infinite models will be ignored in what follows.

[8] For a broad survey of non-monotonic reasoning in the field of linguistics, see Thomason 1997.

This idea is not peculiar to logic of course: 'optimization' as maximization along some ordering lies at the heart of mathematical fields such as decision theory for rational agents, but even at that of natural sciences (witness shortest-path explanations of physical phenomena).

Where we stand. Formal systems of conditional logic can be designed for philosophical intuitions about maximal or closest worlds to be considered when evaluating an implication. Complete logics can be found for such notions just as for classical mathematical logics, and hence the standard mathematical methodology of logic has much further reach than may have been thought. What we see at the same time is that deviating from the standard laws of classical logic does not mean 'no logic', but 'more logic'. This fits with the richness of conditional expressions in natural language, which have continued to inspire work in the philosophy of language until today.

We will pursue one example of these ordering and maximality ideas inside philosophy, where an interesting parallel development took place in the 1970s and 1980s whose similarities with the above may not have been highlighted enough until recently. After that, we will move to concurrent developments in computer science, the other main track where the above ideas have taken hold.

Knowledge as robustness and tracking

As we noted, one of the striking phenomena in the intellectual history we are tracing here is the emergence of similar ideas in different fields of philosophy. Moving from considering all alternatives to special relevant sets of these has also been a conspicuous feature of models for knowledge in epistemology. In the original modal approach to knowledge of Hintikka 1962, knowing that φ meant that φ is true in all worlds that belong to the epistemic range of the agent. But this has several undesirable consequences, such as the automatic closure of knowledge under deduction: $K\varphi$ and $K(\varphi \rightarrow \psi)$ imply $K\psi$, a principle that has been questioned by philosophers from many points of view. One may see this as an unwarranted assumption of 'logical omniscience' giving agents the fruits of deduction for free. But also, even for 'ideally astute agents' that are perfect reasoners, given that skeptics may come up with new scenarios all the time, ranges of epistemic alternatives may be open-ended and impossible to manage. How do I rule out that I am a brain in a vat, or that everything that happens to me is a theatre show arranged by some malevolent demon?

Dretske 1970 proposed an alternative account where knowledge of φ means that we have ruled out all *relevant alternatives* to φ. What these relevant alternatives are may depend on context, on the agents doing the knowing, or on the structure of the known proposition φ. Later versions of this theory have involved comparative order again: relevance means being maximal in some ordering of worlds or possibilities according to their 'relative importance'. Moreover, Nozick 1981 added the important further idea of *tracking* where knowing that φ means that φ is true, we believe that φ, but we are also well-attuned in the sense that if φ had been false we would not have believed that φ (or even believed that not-φ), plus some similar assumptions of 'robustness' across situations. By now, there is a whole spectrum of such epistemological views of knowledge.

Relevant alternatives theories or tracking theories of knowledge (and its companion notion of belief) have been studied extensively in Holliday 2012, which brings to light similarities with conditional logics, as well as striking formal analogies between relevant-alternatives and tracking accounts. The upshot of this body of work for our current story is this. Again we see how ideas from the study of implication: a range of options, and maximizing along some ordering, pervade some of the key notions of one more area of philosophy, help formulate them in new ways, and as a result, help make different epistemological positions comparable, and some-times even inter-translatable.

The above is largely a semantic account of knowledge. However, there is also a more proof-theoretic account in terms of intuitions that philosophers have about inference with knowledge. This approach via accepting or rejecting inferential properties called 'epistemic closure properties', such as $K\varphi \rightarrow K(\varphi \vee \psi)$, or $(K\varphi \wedge K\psi) \rightarrow K(\varphi \wedge \psi)$, is a proof-theoretic alternative to the semantic interface with logic. [9] [10]

Where we stand. We have seen one more, perhaps surprising, parallel development toward ordered-range semantics in another area of philosophy during the 1970s, this time, in relevant-alternatives theories in epistemology. Logic may have served here as a messenger of similar ideas elsewhere, facilitating the flow of ideas from one field to another. At the same time, the continuing formal analogies across subfields like philosophy of science, philosophy of language, deontic logic, and epistemology show logic in another useful role: as a force toward the methodological unity of philosophy.

Artificial intelligence and non-monotonic reasoning

Still roughly in the same time period, around 1980, philosophy and linguistics were joined by another interface for theorizing about reasoning, namely, computer science. In particular, by the end of the 1970s, researchers in Artificial Intelligence started studying styles of reasoning that would fit the design of programs or machines that can solve significant puzzles, as a pilot for complex decision or planning problems. As people realized, making such machines perfect mathematical reasoners with classical logic would be infeasible for reasons of high computational complexity, and also, it would also ignore a major available source of information, namely, the subtle and often quite effective reasoning strategies that humans have when dealing with solving problems and engaging in planning in practice.

Reflecting on these practices, McCarthy 1980 proposed 'circumscription', a form of reasoning where agents only look at smallest or otherwise most relevant models for a set of premises defining the task at hand. They can then draw conclusions that are true in all such models, saving on a costly survey of vast spaces of options. Stated in different terms, by looking at such small models, circumscriptive agents capitalize on their current ignorance of exceptions: what they do not know does not exist.

Circumscription is a new form of inference that does not obey all classical structural rules. In particular, it is *non-monotonic*: learning new premises may change the set of minimal models. But this is exactly what we also saw with conditional logic in philosophy – and this analogy

---

[9] There are interesting similarities between the discussion of closure principles in modern epistemology and those of the earlier structural rules in conditional logic, that remain to be explored in full.

[10] Monotonicity may fail for knowledge when we weaken a current knowledge claim with new (and irrelevant) propositional material: this is reminiscent of Bolzano's insistence, to be mentioned later, on relevance as a key characteristic of philosophical reasoning.

should not come as a surprise. As has been said on the basis of many examples (cf. van Benthem 1990), 'Artificial Intelligence is a continuation of philosophy by other means'. The technical connection between circumscription and other AI techniques in reasoning for planning and problem solving was clarified in Shoham 1988, which showed how the underlying pattern for circumscriptive inference is maximization along an ordering of abstract preference between models. This of course is exactly what we have seen in other settings above.

Since the 1980s, these ideas have coalesced into the area of *non-monotonic logic* which encompasses a vast array of reasoning styles, from problem solving to the practical default inferences we all draw when interpreting a situation, or planning the stream of actions that constitute our working day. Of course, many more ideas have gone into the modeling of default reasoning than we can survey here: two recommended sources are Veltman 1996 and Horty 2012.

What Artificial Intelligence, and in recent years also the study of 'agents' in ordinary computer science, added in this mix was not so much the practical benefit of building reasoning expert systems (though such practical benefits are not to be disdained), but rather a sense of the importance of how we represent the information at hand, the computational complexity of the ensuing reasoning tasks, and philosophically perhaps more important, a systematic awareness of the *nature of the agents*, the epistemic subjects if you wish, that have to do the reasoning.

Where we stand. Artificial Intelligence and computer science added new impetus to the study of consequence and reasoning around 1980, and injected new ideas into their study: such as task-dependence, computational process structure, complexity, and in the end, agency.

Varieties of reasoning: Bolzano, Peirce, and logical pluralism

While our story of implication has moved steadily from the 1950s to the 1980s, it is time for a historical flashback to at least a century earlier. The work from the eighties that we discussed introduced the idea that there are different styles of reasoning, each with their own logical laws, appropriate to some specific area of application. And we do not even have to think of computers only when making that observation. Other well-established reasoning styles with their own peculiarities arise in professional practices such as the law (Prakken 1997). Yet further instances of task-dependence can be found in the empirical psychology of reasoning (Wason & Johnson-Laird 1972). [11]

But this variety of reasoning styles is not just a modern interest. It goes right back to the work of some great innovating logicians in the 19th century whose influence is still felt today. Bolzano 1837, the original source of the modern semantic notion of validity, saw the task of logic as charting the natural reasoning styles that humans have available. For Bolzano, these styles included general purpose inference, but also mathematical and probabilistic reasoning, while he accorded a special place to philosophical reasoning, that he thought needed to satisfy the strictest standards of all in its insistence on 'relevant' trains of thought. [12] But beyond this pluralistic view in itself, Bolzano also noted that different styles of reasoning may satisfy

---

[11] The issue of what actually guides chains of inferences has come up recently in cognitive science, and one answer is that acts of inference are always triggered by decision problems: see Icard 2014 who connects these issues with the foundations of probability, scientific methodology, and decision theory.

different formal properties, in modern terms he was concerned with structural rules (and their entanglement with the language of forms that inferences are couched in) – something with which he was almost two centuries ahead of his time.

Another major source for logic viewed as a science of natural reasoning styles is C.S. Peirce, a contemporary of Frege, co-discoverer of predicate logic, co-founder of algebraic logic, but also a great philosopher with a broad range of interests. [13] Especially influential, right until today, has been Peirce's division of reasoning into three main categories: 'deductive', 'inductive', and 'abductive'. In particular, abduction as inference to the best explanation is still a live subject in argumentation theory, logic, artificial intelligence, and even linguistics and cognitive science.

It has taken a while before variety of reasoning styles and their different formal properties became an accepted topic for philosophers to ponder, but this emancipation has taken place in the program of 'logical pluralism'. [14] While some versions of logical pluralism only considers classical logic, intuitionistic logic, and relevant logic as serious contenders, i.e., mainly a few styles of reasoning known to mainstream philosophers, it is easy to see to place this in a more general setting going back to earlier parts of this paper. As Beale & Restal 2006 emphasize, varieties of consequence are engendered by the set of 'cases' one takes the consequence to range over. Now, if we also analyze and then modulate what these cases depend on, we get much more variety. If the relevant cases are given absolutely, we stay close to standard conditional logics. But if the set of options can depend on the content of the antecedent proposition, or even the consequent proposition, then many more options arise, and a broad spectrum of reasoning styles emerges.

Where we stand. Variety of reasoning styles was a major theme in 19th century logic. It disappeared for a while in the mathematical phase of Frege and his successors, where classical implication was the norm, with a little ecological niche reserved for intuitionistic logic (Troelstra & van Dalen 1988). By now, this broader pluralism toward logical reasoning styles seems a well-accepted position, as logicians broaden their territory to also study other professional practices such as the law, and actual cognitive reasoning behavior. While this may mean that some of the laws of classical implication are abandoned, what is not given up here is the mathematical methodology of modern logic, which is probably much more characteristic for the field than any one specific 'logical law'.

Belief revision theories

We have suggested that logic can study various ways in which human agents can reason, against the background of the objective logical consequences that hold in the world. Philosophy and its subdisciplines themselves provided various instances of this. But this perspective does not stay here: and it raises some fundamental new perspectives on implication. A fundamental innovation was made, again, in the 1980s. [15] If we view the absence of classical structural rules merely as danger signs, our focus is on what can go wrong. In particular, non-monotonicity means that we may have jumped to conclusions that are no

---

[12] The idea that philosophy has its own characteristic repertoire of formal reasoning patterns had also been formulated by some modern philosophers: see the various chapters of Passmore 1961.

[13] Some essential writings from his immense oeuvre can be found in the anthology Peirce 1992.

[14] This program goes back to Restal 2000, where structural rules for logical systems are a crucial engine powering deduction that can be modulated for various reasons, having to do with inferential sources.

[15] Precursors include Weinberger 1965 on logical patterns in reasoning with contradictions.

longer supported by new evidence that we received. But if we just view a reasoning practice in terms of its admissible or inadmissible inferences (admittedly, the way many philosophers think), we ignore a crucial feature. Dangerous forms of inference can be highly successful because there is another human capacity that goes hand in hand with inference, namely, a talent for *correction*, which consists in revising our former conclusions or beliefs when they are contradicted by new incoming information.

While learning by trial and error had already been a major ingredient in some areas of the philosophy of science (Popper 2002), informational actions as such only became a major explicit topic in philosophical work in Alchourron, Gaerdenfors & Makinson (cf. Gaerdenfors 1987), inspired by epistemology and philosophy of science, and partly also by the study of juridical reasoning. In this so-called 'belief revision theory', acts of update, retraction and revision are axiomatized in postulational style, while matching semantic models picturing belief changes were proposed in Grove 1988. Now, we see how reasoning is viewed as a process of forming beliefs, where beliefs can be shown wrong by new incoming information, and have to be retracted.

Not surprisingly, these developments turned out related again to models for conditional logic and nonmonotonic logic, but now used in more radical manners, as a record of what agents currently believe, a record that can change as it gets updated by new information and matching acts of self-correction. By now there is a convergent field of nonmonotonic logic and belief revision theory, where contributions come from philosophy, computer science, but also from quite different areas such as economics and cognitive science. This paradigm has absorbed insights from many disciplines, and it has become a sort of standard for logical approaches to practical reasoning. For a state of the art, see Gärdenfors & Rott 1995, Rott 2001. The field is still expanding: e.g., Kelly 2014 gives a new bridge to longer-term processes of inquiry in learning theory in epistemology.

Where we stand. Reasoning styles are performed by agents, and these agents have a much broader repertoire of epistemic acts of inference, update, correction, and learning. This repertoire lends itself to logical and philosophical study – and thus, the eventual true story of implication may well be one that should include both inference and learning from errors.

Agent dynamics of information and belief

The preceding can be taken one step further, which will also be the last step in the story of this paper. If we see the study of implication as lying in between objective truths and dependencies in nature and reasoning styles of subjective agents, then our conception of logic may well change. What we want to understand now is the total structure of information-processing agents, and what are relevant acts and events here needs to be approached with an open mind (Barwise & Perry 1983, van Benthem 2011). So, what are the basic epistemic actions, and how do agents stay attuned to objective information in the world? Van Benthem 2011 cites an old Chinese example from Moist logic (Liu, Seligman & Zhai, eds., forthcoming), where information comes from three sources, that can be viewed as communication, experience, and demonstration. But many others can be found if one really starts looking at cues from cognitive practice. One basic point to note here is that this richer repertoire of action still satisfies logical laws, and hence we can use the tools of logic to also bring out the abstract principles governing observation, questions, acts of revision, and many others, leading to a general theory of 'information dynamics.

We will not survey this movement in this paper, except for noting a few things. First, the richer the repertoire of informational events and actions, the richer the repertoire of attitudes that agents can have: belief and knowledge are well-known examples, but many other cognitively relevant attitudes have been emerging in recent work in formal epistemology and philosophy of language, including various forms of discourse commitment, preference, and intention. In particular, in this setting, inference is just one talent that human agents have, attuned to just one sort of information in a broad spectrum of varieties of information that are relevant to logic (van Benthem & Martinez 2008). Indeed, explicit inference seems related to raising awareness or attention to particular facts that agents already knew implicitly (van Benthem & Velazquez-Quesada 2011).

Viewed in this way, it may look as if our emphasis on implication and consequence has been just a small corner of the total world of logic and philosophy. But viewed in another way, we are still at the heart of things. Even other information-producing acts such as observation or communication, obey laws, and stating these laws and probing what follows from them is again an exercise of our logical reasoning abilities. Nevertheless, the introduction of a wider repertoire of epistemic acts also may have consequences for how we study implication and consequence. For instance, one now gets options in studying the non-standard nonmonotonic reasoning styles introduced earlier. One might also view a nonmonotonic inference as an act of belief formation, and then decompose what used to be nonmonotonic inference with a non-classical logic into a theory of acts of belief revision on top of a standard classical logic (van Benthem 2011). And this is just one instance of modern perspectives. Current views of inference also bring in the role of *decisions* and *planning* (van Lambalgen & Stenning 2008, Icard 2014), and from there, it just a small step to *game-theoretic* views of reasoning where strategic multi-agent interaction is essential (van Benthem 2014). All these topics go far beyond the compass of this paper, but they do illustrate where the study of implication, consequence and reasoning stands conceptually today.

Where we stand. Reasoning is part of a much wider repertoire of informational actions for human agents, whose process structure can be modeled and whose laws can be studied by logical techniques. The very status of implication, consequence, and inference, and how to best conceptualize them then becomes a matter of logical debate: but going beyond technicalities, also of philosophical debate. Part of that debate may also be the issue of where to locate the border line between the study of implication and consequence and that of agency in general. One wants to let a broader world enter the house of logic, one does not want to claim that logic is the study of everything.

Small conclusions: implication

In this paper, we have traced a small history of implication, whose general line was a move from considering ‘all cases’ where the premises hold to only ‘the best’ cases as measured by some ordering over models or worlds. We have seen that this idea arose in many areas independently, something which we would see as a sign of stability and content, somewhat along the lines of parallel inventions in the mathematical theory of proof or computation. On the way, we noted many other features with broader methodological impact, such as the way in which semantics and proof analysis offer equally valid dual perspectives on how to study an phenomenon from a logical point of view. Next, we saw a further trend toward making the agents explicit behind implication and consequence: their reasoning styles, their further abilities (e.g. for correction), in short: the nature of the reasoning subjects entered the study of reasoning. Finally, we saw how all this plays against two perennially entangled aspects in logical history: what agents do, and how nature is.

Our story is by no means complete. Further perspectives on implication that we have ignored, but that would fit in is the modern conglomerate of proof theory, construction of objects, algorithms, and computation. Likewise, there is the world of probabilistic reasoning, that connects to information theory, to the numerical qualitative boundary, and to strength of inferences. And finally, we have ignored how reasoning is not just something that takes place in a fixed given language, but rather something that plays a role in processes of language creation. Logic is not just the spinning out of what follows from given conceptual frameworks, it is also an actor in creating these.

Big conclusions: logic and philosophy

Personally, I see philosophy and logic as separate fields: logic is much broader than philosophy, and philosophy is much broader than logic. But the two have been lifelong friends and allies. In this contact, logic can help to clarify philosophical views and make them precise, but at the same time, in doing so, it also enhances creativity, as abstraction is always a source of new possibilities. Logic also helps observe analogies across different fields of philosophy: as it cuts across pairs such as epistemic-ontological, facts-norms, information-action. In doing so, it offers new options for developing philosophical frameworks and positions. Finally, the congeniality also shows in that both fields often seem to be undergoing the same outside influences at the same time. Right now, for instance, there is pressure of taking the empirical facts of human cognition seriously in logic, instead of hiding timidly behind the usual normative-descriptive barrier, and likewise, there is a constant backflow of ideas into pure logic coming from applications. But the very same is true for many areas of philosophy, especially, the ones that we have mentioned in this paper.

Can a purely formal discipline like logic really deliver all of the listed benefits? As the reader may have sensed in reading this paper, there is no easy separation of descriptive content and formal tools in logic. One can influence the other – and even highly formal concerns can become very practical realities, as we saw in our discussion of Artificial Intelligence. [16] All this dynamic is as we also see it in the historical entanglement of technology and art: reality inspires theory, but theory also creates reality. Logic will not do all the work for philosophers, but it is a potent stimulant.

Finally, I have not made any attempt at proposing profound insights or deep mathematical theories in this paper. I mainly tried to sensitize the reader to how much is going on at the interface of logic and philosophy. This paper already brings together an impressive amount of material and references just on the theme of implication and consequence, and not even all of that theme was pursued at length here. I submit that it is time to do more of this tracking of broad ideas, try to see the total landscape that has been developing over the last decades, and only then try to draw general conclusions and make general predictions. [17]

---

[16] Compare also the development of the modern computer out of mechanizing ideal logical reasoning.

[17] I refer once again to van Benthem 2006 for many further examples. Incidentally, this thematic approach would also work for interfaces between logic and other disciplines, such as mathematics, linguistics, or computer science, which have sometimes been `ideologized' too much as well.

References

P. Adriaans & J. van Benthem, eds., 2008, *Handbook of the Philosophy of Information*, Elsevier Science, Amsterdam.

A. Aliseda-Llera, 2006, *Logical Investigations Into Discovery and Explanation*, Springer Science, Dordrecht.

J. Barwise & J. Perry, 1983, *Situations and Attitudes*, The MIT Press, Cambridge MA.

J. Barwise & J. Seligman, 1995, *Information Flow, Logic and Distributed Systems*, Cambridge University Press, Cambridge UK.

J.C. Beal & G. Restal, 2006, *Logical Pluralism*, Oxford University Press, Oxford.

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

J. van Benthem, 1990, Kunstmatige Intelligentie: Een Voortzetting van de Filosofie met Andere Middelen, *Algemeen Nederlands Tijdschrift voor Wijsbegeerte* 82, 83-100.

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 Honor of Vladimir Smirnov,* Kluwer Academic Publishers, Dordrecht, 63-82.

J. van Benthem, 2002, 'Invariance and Definability: Two faces of logical constants', in W. Sieg, R. Sommer & C. Talcott, eds., *Reflections on the Foundations of Mathematics. Essays in Honor of Sol Feferman*, ASL Lecture Notes in Logic 15, 426–446.

J. van Benthem, 2006, Logic in Philosophy, in D. Jacquette, ed., 65–99.

J. van Benthem, 2010, *Modal Logic for Open Minds*, CSLI Publications, Stanford University.

J. van Benthem, 2011, *Logical Dynamics of Information and Interaction*, Cambridge University Press, Cambridge UK.

J. van Benthem, 2014, *Logic in Games*, The MIT Press, Cambridge MA.

J. van Benthem, 2015, 'Fanning the Flames of Reason', *Nieuw Archief voor Wiskunde*, Leiden, and *Studies in Logic*, ILC Guangzhou.

J. van Benthem & A .Gupta, eds., 2011, *Logic and Philosophy Today*, College Publications, London.

J. van Benthem & F. Liu, eds., 2013, *Logic Across the University: Foundations and Applications*, College Publications, London.

J. van Benthem & M. Martinez, 2008, 'The Stories of Logic and Information', in P. Adriaans & J. van Benthem, eds., 217–280.

J. van Benthem & F. Velazquez Quesada, 2010, 'The Dynamics of Awareness', *Synthese* 177:1, 5–27.

E. W. Beth, 1959, *The Foundations of Mathematics*, North-Holland, Amsterdam.

P. Blackburn, J. van Benthem & F. Wolter, eds., 2006, Handbook of Modal Logic, Elsevier Science, Amsterdam.

M. Bochenski, 1961, *A History of Formal Logic*, Notre Dame University Press, Notre Dame, Indiana.

B. Bolzano, 1837, *Wissenschaftslehre*, Seidelsche Buchhandlung, Sulzbach.

R. Carnap, 1967, The Logical Structure of the World, The University of California Press, Berkeley CA.

R. Chisholm, 1963, 'Contrary-to-Duty Imperatives and Deontic Logic', *Analysis*, 24: 33–36.

F. Dretske, 1970, 'Epistemic Operators', *The Journal of Philosophy* 67:24, 1007–1023.
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.
D. Gabbay & J. Woods, eds., 2004–present, *Handbook of the History of Logic*, Elsevier Science, Amsterdam.
P. Gärdenfors, 1988, *Knowledge in Flux*, The MIT Press, Cambridge MA.
P. Gärdenfors & H. Rott, 1995, 'Belief Revision', in D. M. Gabbay, C. J. Hogger & J. A. Robinson, eds., *Handbook of Logic in Artificial Intelligence and Logic Programming* 4, Oxford University Press, Oxford.
N. Goodman, 1947, 'The Problem of Counterfactual Conditionals', *Journal of Philosophy* 44:5, 113–128.
A. Grove, 1988, 'Two Modelings for Theory Change', *Journal of Philosophical Logic*, 157–170.
S. O. Hanson, 2001, 'Preference Logic', in D, Gabbay & F. Guenthner, eds., *Handbook of Philosophical Logic IV*, second edition, Kluwer, Dordrecht, 319–393.
B. Hansson, 1969, 'An Analysis of Some Deontic Logics', Nous, 3, 373–398.
I. Heim, 1992, 'Presupposition Projection and the Semantics of Attitude Verbs', *Journal of Semantics* 9:3, 183–221.
P. Hempel, 1965, *Aspects of Explanation and Other Essays in the Philosophy of Science*, The Free Press, New York.
J. Hintikka, 1962, *Knowledge and Belief*, Cornell University Press, Ithaca NY.
W. Holliday, 2012, *Knowing What Follows. Epistemic Closure and Epistemic Logic,* Dissertation, Department of Philosophy, Stanford University & Dissertation DS-2012-09, ILLC, University of Amsterdam.
J. Horty, 2012, *Reasons as Defaults*, Oxford University Press, Oxford.
Th. Icard, 2013, *The Algorithmic Mind, A Study of Inference in Action*, Ph.D. Thesis, Department of Philosophy, Stanford University, Dissertation 2014-02, ILLC, University of Amsterdam.
D. Jacquette, ed., 2006, *Handbook of the Philosophy of Logic*, Elsevier Science, Amsterdam.
K. Kelly, 2014, 'A Computational Learning Semantics for Inductive Empirical Knowledge', in A. Baltag & S. Smets, eds., *Johan van Benthem on Logic and Information Dynamics*, Springer Science Publishers, Dordrecht, 289–337.
W. & M. Kneale, 1961, *The Development of Logic*, Oxford University Press, Oxford.
M. van Lambalgen & K. Stenning, 2008, *Human Reasoning and Cognitive Science*, The MIT Press, Cambridge MA.
C. I. Lewis, 1918, *A Survey of Symbolic Logic*, The University of California Press, Berkeley CA.
D. Lewis, 1973, *Counterfactuals*, Blackwell, Oxford.
F. Liu, 2011, *Reasoning About Preference Dynamics*, Springer Science Publishers, Dordrecht.
F. Liu, J. Seligman & J. Zhai, eds., forthcoming *Handbook of the History of Logical Thought in China*, to appear with Springer Asia Publishers, Beijing.
J. McCarthy, 1980, 'Circumscription – A Form of Nonmonotonic Reasoning', *Artificial Intelligence* 13, 27–39.
R. Nozick, 1981, *Philosophical Explanations*, Harvard University Press, Cambridge MA.
J. Passmore, 1961, *Philosophical Reasoning*, Duckworth, London.
C.S. Peirce, 1992, *The Essential Peirce*, 2 vols. Edited by Nathan Houser, Christian Kloesel, and the Peirce Edition Project (Indiana University Press, Bloomington, Indiana).
K. Popper, 2002, *The Logic of Scientific Discovery*, Routledge, London.

G. Restal, 2000, *An Introduction to Substructural Logics*, Routledge, London.

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

H. Prakken, 1997, *Logical Tools for Modeling Legal Argument. A Study of Defeasible Reasoning in Law*, Kluwer, Dordrecht.

D. Scott, 1971, 'On Engendering an Illusion of Understanding', *Journal of Philosophy* 68, 787–807.

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

R. Stalnaker, 1968, 'A Theory of Conditionals', *Studies in Logical Theory, American Philosophical Quarterly* (Monograph Series, 2), Blackwell, Oxford, 98–112.

R. Thomason, 1997, 'Nonmonotonicity in Linguistics', in J. van Benthem & A. ter Meulen, eds., *Handbook of Logic and Language*, Elsevier Science, Amsterdam, 777–831.

S. Toulmin, 1958, *The Uses of Argument*, Cambridge University Press, Cambridge UK.

A. Troelstra & D. van Dalen, 1988, *Constructivism in Mathematics*, North-Holland, Amsterdam.

F. Veltman, 1996, 'Defaults in Update Semantics', *Journal of Philosophical Logic* 25:3, 221–261.

H. Wansing, 1993, *The Logic of Information Structures*, Springer Science Publishers, Berlin.

P. Wason & Ph. Johnson-Laird, 1972, *Psychology of Reasoning: Structure and Content*, Harvard University Press, Cambridge MA.

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