# Logic, Mathematics, and General Agency

Johan van Benthem

## 1 Mathematics and common sense: two competing paradigms?

If logic is the general study of a priori valid reasoning, then where is the paradigmatic area where we see this reasoning in its full glory? To some, this is clearly mathematics, where precision is relentless, and strings of inferences are taken to impressive lengths. But on another view, the highest form of reasoning is displayed in the ordinary world of common sense – say, when engaging in conversation about something that matters, where pure information is deeply intertwined with evaluation and goals, and where, crucially, we are surrounded by further agents like us that we must interact with. On the first view, to simplify things a bit, logic is about mathematical proof and related processes like computation, making mathematical logic and foundations of mathematics the heart of the field. Agency is not even needed, and no human aspects are modeled. On the second view (frankly speaking: my own), logic is about interactive agency and all that entails, making philosophical logic and much more equally central to the discipline. The purpose of this brief note is to bring the two perspectives together – though admittedly, only in a light and preliminary manner.

But before I do, let me make sure that I am not setting up the wrong debate. First, from the viewpoint of agency, there is no competition. Mathematics is an important special form of human cognitive behaviour – and the fact that it has developed historically out of our daily social planning abilities does not detract from its power and importance. Any general logic of agency must come to terms with our mathematical activities. Moreover, one can even grant that agenda contraction and restriction to a subdomain can be a winning move in terms of scientific progress: the more specialized concerns of the foundations of mathematics have had immense benefits for logic in general.

Also, a distinction needs to be kept in mind here. It might well be that mathematical logic should still be the hallmark of logic at a meta-level, in

P.-E. Bour, M. Rebuschi & L. Rollet (eds.), *Construction, 277–296.*

terms of the *methods* and standards that it provides for system building.[1] But that does not imply, at the object level of reasoning practices, that the mathematical activity itself should be the paradigmatic area of study for logicians. So much more is worthy of our admiration!

But even with these reasonable distinctions, the contrast may just be overstated. Looking more closely at what professional mathematicians actually do, we see about every feature of general agency: they have knowledge, but also expert beliefs, their research is guided by values they put on results and excitement about new questions, and despite occasional fads of social ineptitude, they manage to interact very successfully. In this lively setting, classical logic has made some extreme abstractions. A 'theory' is a set of formulas in some formal language, a 'proof' is a string of formulas satisfying simple combinatorial criteria. No agents enter the story: only the products of their activities matter. Perhaps surprisingly, these abstractions have been successful. When all traces of human activity are stripped off, we find fundamental insights like Gödel's Theorems, or other major results that have set logic on its modern course.

But now let's move beyond this austere format. Even Euclid's *Elements*, a key document in the history of foundational research, has many further lively aspects that seem crucial to mathematics as understood and practised. There is an active role for definitions, proofs come in a task-oriented format, theorems often come hand in hand with algorithmic constructions, there are tantalizing glimpses of dual methods of 'analysis' and 'synthesis', and so on. Despite centuries of increased formalization, the reality of mathematics today is still of this richer sort. Say, an area like 'Arithmetic' is much more than a formal system of Peano axioms plus first-order inference rules: it is also an agenda of questions, a set of methods, skills, and also, styles of interacting with other mathematicians. These richer intellectual features of mathematical activity have been noted by many authors, from Lakatos in the delightful *Proofs and Refutations* [19] to Brouwer's view of the creative mathematical activity, or in terms of more-agent interaction, Lorenzen's dialogue systems underpinning logical laws.[2]

Indeed, perhaps paradoxically, interactive human activity is a major motivation for the process of formalization itself. Formalizing scientific reasoning and raising precision are all about providing more precise *intersubjective* styles of communication.

In this brief note, I put together current logics of agency with mathematical activities, and discuss what issues arise. I have no deep results to offer, and indeed, I mainly find challenges to my own dynamic logics, rather than

[1] I also see drawbacks to its 'systems' methodology, but will not raise them here.

[2] [24] even claims an actual dialogical origin for Euclid's format and terminology.

sweeping insights into mathematics. But this is just an opening round, and I make some broader suggestions at the end.

## 2 Dynamic logics of agency

Before making concrete comparisons, here is a very brief tour of some recent dynamic logics of agency, as these are much less-known than standard logical frameworks.

**Rational agents** Let us first look at what a logic of full-blooded agency involves.[3] Rational agents are endowed with a number of powers and can perform many cognitive tasks. I think of them as a next stage after Turing machines, that were simple robot-like agents for basic computational tasks. Here are some core features of agency that have turned out amenable to logical investigation. First of all, agents exercise *informational powers*, through external acts of observation, or internal acts of inference, introspection, or memory retrieval. In doing so, they change their knowledge, but also other attitudes that guide behaviour, such as their beliefs. But this information gathering is not a blind process: it has a *direction*, given by an agenda of current 'issues', and the agenda items are steered by agents' questions, and other acts. In a stronger sense, these directions are tied up with genuine goals, having to do with agents' preferences and *evaluation* of situations, another crucial aspect of rational agency. Without the latter, there is just logical 'kinematics', but no deeper explanatory 'dynamics' of behaviour. And finally, human agency is crucially *interactive*, largely taking place in social settings. As in physics, where many-body interaction is the key, strategic many-mind interactions drive logical behaviour, including conversation, argumentation, or more general games. Thus, we get a picture of individual agents endowed with a set of core capacities, involved in dynamic transitions of various kinds from one state to another, and in the process, creating long-term practices over time, with larger groups of participants.

**Logical dynamics of information** While all this might read like an empirical account of human behaviour, the point is that this picture of agency also admits of normative logical study. In particular, information flow though observation or communication of facts is the area of modern *dynamic-epistemic logics*, where successive events of public or private observation or communication change a current epistemic information state, represented by some standard epistemic model $\mathcal{M}$ for one or more agents.

[3] This section is an executive summary of [7], a book that sets out the program of Logical Dynamics and the technical results cited in what follows in great detail.

Here is a standard example of this methodology, concerning an act

$!\varphi$ of public observation, or public announcement,

that the proposition $\varphi$ is currently true. The resulting *update* trims the current epistemic model $\mathcal{M}$ to the model $\mathcal{M}|_\varphi$ retaining just the worlds that satisfy $\varphi$. This shrinking of one's current epistemic range by events of 'hard information' makes information flow through reduction of uncertainty. Characteristic logical laws of such updates are recursion equations telling us what agents $i$ know after some informational event has taken place, in terms of a standard epistemic modality $K_i\psi$ ('agent $i$ knows, or is informed that $\psi$'). Letting a further dynamic action modality $[!\varphi]$ refer to the new model $\mathcal{M}|_\varphi$ arising here, the following key equivalence then holds:

$$[!\varphi]K_i\psi \leftrightarrow (\varphi \rightarrow K_i(\varphi \rightarrow [!\varphi]\psi))$$

Thus, events of hard information change agents' current knowledge, and therefore also, in the process, epistemic statements may change their truth-values. The recursion principle stated just now then reduces knowledge after the event to conditional knowledge before, while taking proper care of these possible truth-value changes.

We will not go into details of these systems, which can also handle more sophisticated events with private information. Our point here is that these phenomena admit of logical study in terms of mathematical systems obeying the usual criteria of the discipline.[4] And therefore, bringing logics of common sense agency to bear on mathematical practice is not a matter of informal talk, but an appeal to concrete logical systems.

**Logical dynamics of belief** Similar logical principles govern further informational acts, and other attitudes that agents can have, such as their beliefs. Consider doxastic-epistemic 'plausibility models' where epistemic equivalence classes are now ordered by a relation of relative plausibility. In such models, an agent believes that $\alpha$ if $\alpha$ is true in the most plausible epistemically accessible worlds. The more general notion needed in such a doxastic setting is that of *conditional belief* $B^\psi\alpha$, which says that the formula $\alpha$ is true in all most plausible epistemically accessible worlds that satisfy $\psi$.

In this setting, beliefs can change in at least two ways. First, they can change under the above events $!\varphi$ of hard information, validating the following recursion equation:

$$[!\varphi]B^\psi\alpha \leftrightarrow (\varphi \rightarrow B^{\varphi\wedge[!\varphi]\psi}[!\varphi]\alpha)$$

[4] We will mostly drop agent subscripts $i$ in the rest of this paper, for greater readability.

But the setting is richer now, and there are also events $\Uparrow\varphi$ of *soft information*, that do not eliminate worlds, but merely change the plausibility order, making (former) $\varphi$-worlds more plausible than the $\neg\varphi$-ones. The following recursion equation then states how an agent's conditional beliefs $B^{\psi}\alpha$ change systematically:

$$[\Uparrow\varphi]B^{\psi}\alpha \leftrightarrow$$
$$(\Diamond(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{\varphi\wedge[\Uparrow\varphi]\psi}[\Uparrow\varphi]\alpha) \vee (\neg\Diamond(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{[\Uparrow\varphi]\psi}[\Uparrow\varphi]\alpha)$$ [5]

We state this rather technical axiom here, not for further use in what follows, but to stress an earlier point. Logic of agency is not primarily about mathematical activity, but its methods are mathematical. Indeed, dynamic logics of agency often merge ideas from 'philosophical', 'mathematical', and even 'computational' logic, making all these labels somewhat obsolete as separate subdisciplines. All are parts of the same story.

**A general dynamic turn** For the purpose of this paper, it is enough to see the general Dynamic Turn in logic at work here. In every province of agency, we look for the crucial events that drive it, and then model these explicitly in the logic, including the recursion laws that specify how agents' attitudes change under these triggers. By now, dynamic logics have been written for many other items in the above picture of agency, including events of *preference change* (such as commands by an authority) that affect our evaluation of worlds, or events of *issue management* (such as questions changing the current agenda of issues), or indeed inference itself (see below).

**Interaction, games and groups** Finally, the preceding laws merely describe single steps of information flow or attitude change of single agents. In general dynamic logics, these are just building blocks for two further levels. One is *longer-term temporal patterns* of behaviour, with moves made in response to others, as in argumentation or *games* in general. Significantly, games are a powerful paradigm in logic,[6] and the above dynamic logics analyze their fine-structure. The other aggregation level is that of *larger groups of agents* engaging in shared activity, such as coalitions in games, or communities of speakers and hearers in communication. Epistemic logic has long studied common knowledge and other crucial informational notions concerning groups, and the dynamic perspective adds issues like the formation of group knowledge and group belief through communication and interaction. While there is some awareness of the role of process structure

[5] In this axiom, $\Diamond$ is the existential modality associated with the earlier epistemic knowledge operator $K$.

[6] [5] is a history of Brouwer's ideas on foundations of mathematics up to modern game semantics of computation and linear logic. [4] is an extensive survey of 'logic games', and conversely, of 'game logics' applying logic to general games.

in general logic,[7] groups have been less of an explicit theme. But clearly, much of human reasoning is a social group activity, in the form of argumentation,[8] and much of science is even a group process par excellence: 'organized rationality'.

**From common sense to science** Now, what does general agency have to do with something as pure as mathematics? I myself think: a lot, because I fail to see any strict boundary between science and common sense.[9] Science is just one striking form of cognitive behaviour of our human species, using general cognitive skills honed first in biological survival, and then refining them for specific purposes. Indeed, the picture of agency that I have sketched in all its aspects, with systematic information gathering, goal management, evaluation, and temporal social structures, seems also a realistic description of science as a rational activity. Despite some undeniable differences in emphasis, theme and structure of the community,[10] no principled border-line seems to separate general intelligent action from mathematical proof or other scientific activities.

But if this is so, can the dynamic logics discussed here throw new light on scientific activity, the same way that the traditional more austere logical analysis has done? And can benefits flow the other way? What can a logic of general rational agency learn from a study of mathematical activity? The sections that follow discuss a few encounters.

## 3 Dynamics of inference, and mathematical proof

**Information and knowledge** Empirical science involves two main information sources in general cognition: experimental observation – or if you wish: questions to Nature – entangled with deductive, and perhaps also other styles of inference.[11] This interplay has long been emphasized by logicians like Hintikka (cf. [17]). But the case of mathematics is special, since there is no empirical observation – if one disregards some recent com-

[7] [6] is a study of this theme in the modal semantics of intuitionistic logic.

[8] One can even defend the view that single-agent reasoning is a mere limit case of the multi-agent scenario, with different voices in my head stating relevant assertions and objections. Manuel Rebuschi reminds me of Plato's *Sophist* here: "*Stranger*: Well, then, thought and speech are the same; only the former, which is a silent inner conversation of the soul with itself, has been given the special name of thought. Is not that true?" (Soph. 263e).

[9] This is not the place for a detailed assessment, but all the usual criteria for making a sharp distinction in an influential book like [20] seem a matter of degree to me.

[10] The latter may be less complex in terms of dimensions involved (compare proving a theorem to writing a successful application letter), but it will be much more focused and probing combinatorially.

[11] [8] brings dynamic-epistemic views to the philosophy of science. [2] have a concrete new dynamic logic analysis of quantum mechanics.

putational experimental developments whose status is not yet clear. What happens to our dynamic logics in this setting? Can there still be knowledge update, belief revision, goals, and even broader concerns?

**Inference and mathematical knowledge** A first striking feature of mathematics at once poses a challenge to the dynamic logics given here, connecting up with a major issue in epistemology. There does not seem to be any broadly accepted model for mathematical knowledge that would allow us to state simple truths like "I don't know if Goldbach's Conjecture is true". At least, epistemic logic has no obvious format for this purpose, as the requisite semantic variety cannot arise in its models. For, mathematical statements are either true in all worlds, or false in all of them. And in line with this lack of a paradigm, there is no accepted model of mathematical acts of knowledge dynamics.[12]

**Inference and fine-grained information** One factor is that the *syntactic information* provided by inference is not at all the same as the *semantic information* derived from observation or related acts. [10] show how this problem occurs much more broadly in logic, starting from the tension between the usual notion of validity: 'valid conclusions add no semantic information to premises', and the feeling that, on the other hand, valid inference is undeniably useful in 'unpacking information'. While there is no consensus on how to best draw the distinction, most approaches to inferential information make an appeal to syntax, one way or another.[13]

The perspective that we will use here stays with the earlier dynamic-epistemic logics. We now assume that epistemic worlds $w$ come with sets of syntactic formulas $E_w$ *explicitly entertained* at them by the agent.[14] These formulas can be true or false, in line with the fact that, in inference, the manipulated formulas need not be true. This extension suggests enriching the usual epistemic language with a new operator

> $E\varphi$, the formula $\varphi$ is in the current entertainment set.

Now we can define new epistemic attitudes that go beyond the implicit knowledge $K\varphi$ of epistemic logic. One such new notion is

> *explicit knowledge* $EX\varphi$, defined as a conjunction $K\varphi \wedge EK\varphi$: the agent knows that $\varphi$ implicitly, and is aware of this knowledge.

[12] [6] shows how the usual semantics of *intuitionistic logic* may be (re-)interpreted as an *implicit* account of various kinds of informational action in mathematics.

[13] While the syntax level has had a bad press in philosophical analysis as being overly detailed, it is of course the crucial medium for subtleties of formulation and procedure.

[14] True entertained formulas model the *explicit access* an agent has to the current world.

For this 'introspective strengthening' of semantic knowledge, cf. [14], [11]. Again, this is not a technical paper with details, but complete proof systems for such extended logics are easy to find. I will assume in the rest of this discussion that agents have implicit knowledge of what they entertain: i.e., the entertained formulas are the same in all worlds that an agent finds epistemically indistinguishable. 'Implicit introspection' $E\varphi \to KE\varphi$ seems quite reasonable to me.

**Syntax dynamics** But the job of analyzing inference in our present style is not yet done. The dynamic-epistemic logics of the preceding section enumerate validities that hold about informational acts, but they do not address the *dynamics of inference itself.* To make the latter explicit, we need more fine-grained events, beyond the earlier ones that changed domains of worlds or plausibility relations over these. In particular, we now need syntactic update of entertainment sets, and a typical example will be an act

$+\varphi$, adding formula $\varphi$ to all current sets $E_w$
('awareness raising').

In general, such an awareness raising act will not occur randomly. It might be induced an act of inference: say, drawing a conclusion $\varphi$ from premises that we already knew. Or it could be licensed by an act of introspection, or by memory search. And once we have such model-changing actions available, we can write a dynamic logic describing their effects on agents' attitudes: semantic implicit knowledge, syntactic entertainment, and mixed notions defined from these such as the above explicit knowledge.

**A complete dynamic logic of semantic and syntactic information** More precisely, the earlier logical methodology still applies in this extended setting. Say, a typical recursion axiom for an act $+\varphi$ would now be the following equivalence:

$$[+\varphi]E\psi \leftrightarrow E\psi \vee \psi = \varphi$$ [15]

Here are three other recursion laws, with implicit knowledge and semantic update. Two of them say that syntactic and semantic update work only within their own domain:

$$[+\varphi]K\psi \leftrightarrow K[+\varphi]\psi$$
$$[!\varphi]K\psi \leftrightarrow (\varphi \to K(\varphi \to [!\varphi]\psi))$$
$$[!\varphi]E\psi \leftrightarrow (\varphi \to [!\varphi]E\psi)$$

While these laws are extremely simple, they can analyze somewhat interesting notions. For instance, one law of the system is this:

[15] The second disjunct involves some abuse of notation, being a syntactic identity.

$$\varphi \to [+\varphi]\varphi$$

that is, entertainment acts have no side effects on the truth of the formula involved.[16] Using this first observation, here is a second basic validity:

$$K\varphi \to [+K\varphi]EX\varphi$$

This says that an 'entertainment act' for an implicit knowledge statement turns the latter into explicit knowledge. Here is a formal derivation:

| | |
|---|---|
| $K\varphi \to [+K\varphi]K\varphi$ | (by the preceding observation) |
| $[+K\varphi]EK\varphi$ | (by one of our dynamic axioms) |
| $K\varphi \to [+K\varphi](K\varphi \wedge EK\varphi)$ | (using propositional logic) |

But the calculus can also analyze more standard issues concerning inference:

*Example* The missing action in logical closure.

Consider the vexed problem of logical omniscience. The following closure principle holds in our logic for implicit semantic knowledge, as it quite properly should:

$$(K\varphi \wedge K(\varphi \to \psi)) \to K\psi$$

But what does not hold, and should not hold is

$$(EX\varphi \wedge EX(\varphi \to \psi)) \to EX\psi$$

Entertaining the premises of Modus Ponens does not imply entertaining its conclusion (yet). But in our dynamic perspective, neither of these assertions touches the crux of the matter. That is rather that agents can come to explicit knowledge if they are willing to make an effort. Thus, the above implication contains a '*gap for an action*' [...]:

$$(EX\varphi \wedge EX(\varphi \to \psi)) \to [...]EX\psi$$

And indeed, our logic proves the following:

$$(EX\varphi \wedge EX(\varphi \to \psi)) \to [+K\psi]EX\psi$$

*Example* From proof to refutation.

But the usual discussions of omniscience are biased, since they only emphasize one role of deductive inference: taking us from known truths to known truths. That is not even its only function inside mathematics. How do we analyze a refutation, going from known falsehoods to new ones? Simply in terms of dynamic validities like this:

[16] The reason is this: the formula $\varphi$ has no sub-formulas long enough to be affected by $E\varphi$'s becoming true.

$$(EX(\varphi \to \psi) \wedge EX\neg\psi) \to [+K\neg\varphi]EX\neg\varphi$$

Thus, our dynamic logic can a least get some basic features right for acts of inference and the two sorts of logical information, in one simple framework.[17]

## 4 From general agency to mathematics, and back

How does the dynamic-epistemic logic of inference apply to mathematical reasoning, that takes place in the setting of mathematical theories?

**Inference and syntax dynamics** Semantically, one can represent the mathematical theory one is working with as the set of its models. Then, to deal with inferential information, these models have to be multiplied, as each now comes with its syntactic 'access set' of currently considered facts.[18] There are even two versions for this:

Some mathematical theories describe one standard model, say, the natural numbers $\mathbb{N}$. This is like the 'actual world' in epistemic logic, and the agent already knows $\mathbb{N}$ implicitly, while the real task is to enrich its syntactic description. Thus, the universe would consist of pairs $(\mathbb{N}, X)$, with $X$ a set of arithmetical formulas. There would be no equivalent of public announcement then, since we will never rule out the topic $\mathbb{N}$ from consideration. One might think that representing $\mathbb{N}$ is redundant then, but we may keep it around when comparing different mathematical theories of this kind.

But when the mathematical theory consists of, say, the axioms for groups, we want many different models $(\mathcal{M}, X)$ for different groups $\mathcal{M}$ – and adopting new mathematical axioms (say, specializing to commutative groups) would now be like the earlier public announcements of new semantic facts.[19] Either way, the earlier framework applies.

As in the above dynamic logic, knowledge growth by deduction can be represented as extension of the current access set through acts of entertainment, without change in the class of models. But mathematics may have other awareness-raising dynamic events as well, beyond inference: our

[17] I am not entirely happy with the current proposal, since explicit knowledge may be more than implicit knowledge plus 'thinking about it'. I could be thinking about my implicit knowledge of your salary, without being aware that I in fact know it. Epistemic 'awareness that' might be a stronger notion, sui generis, and then the dynamic logic needs to be extended accordingly. This raises some unsolved problems with awareness-raising acts for complex epistemic statements. But it may be the better account of what happens when we consciously *draw* a conclusion.

[18] We can think of some of these as explicitly known, while others are just open problems currently under investigation, as in the earlier-mentioned syntactic notions of agenda and issue management.

[19] Manuel Rebuschi has pointed at the intermediate case of mathematicians reasoning about some 'generic model' that looks singular, but stands for a whole family of structures.

dynamic logic is neutral on this. Candidates for such events are acts of geometrical intuition making us aware of some truth already implicit in our semantic information. Still, all this stays close to reformulation – and at present, I can only offer the above system as a perhaps illuminating way of recasting things.[20] I have no applications yet, and the reason may be the poverty of our model so far, ignoring the rich structure of dynamic acts that create and modify mathematical proofs.

**Higher-order knowledge** More subtle features of our dynamic-epistemic logics have to do with what agents know 'socially', not about facts, but about each others' knowledge and ignorance. Observational update becomes exciting precisely because truthfully stating that something is true may have dynamic effects on epistemic statements, changing the original situation. A famous scenario of this kind are true self-refuting

Moore-style sentences $\neg Kp \wedge p$ ("you do not know it, but $p$")

that become false upon being stated. This shows the subtleties of complex epistemic assertions, and a theory of short- and long-term update behaviour has taken off here. In particular, the non-monotonicity of ignorance statements drives crucial information flow in communication. But in mathematics, facts about epistemic states of the reasoner (either her knowledge or ignorance) are not part of the mathematical theory itself.[21]

Still, I am not completely satisfied with this negative assessment either. First, we all agree that truly *competent* scientists are those who also know what their community does not know. And also, mathematics seems the subject par excellence where meta-statements about provability, unprovability, and consistency can be coded back into plain mathematics, and hence be part of arithmetic, set theory, or other theories with enough coding power. But this raises large issues of operator treatments of knowledge versus predicate-based ones (cf. [13]) that would lead me too far afield here.

Let me now turn the other way, and ask what new things a closer study of mathematical reasoning has to offer to the dynamic logic of general agency as sketched in the above. I will only mention two themes here that make the point:

**Dynamics of proof** The High Mass of dynamic analysis is finding the natural repertoire of acts or events that change the relevant information states. But what are the natural dynamic steps in deductive inference?

[20] But cf. [25] for a more detailed epistemic-dynamic logic of inference.

[21] Admittedly, intuitionistic logic gives an epistemic flavour to logical constants, and hence also to mathematical statements containing these. But intuitionism is only about 'monotonic' established knowledge, and not about 'non-monotonic' ignorance statements.

There are acts of 'drawing a conclusion', and maybe we have thrown some light on these. But there are also 'making an assumption', 'refuting a claim', and others. In fact, deduction is so interesting precisely because it is a rich cognitive practice full of rather subtle actions. And these actions also come with a rich repertoire of epistemic attitudes. Deduction is not just about knowledge or belief, but about entertaining hypotheses, and further ways of having propositions in mind that never made it into standard philosophical logic. Finding dynamic logics for this rich cognitive practice is a challenge, and it would require a fresh look at Proof Theory. But it has not yet been done in the logics of agency that I have presented.

There is even one more good reason for doing so:

**Proofs and skills** Analyzing inference steps still does not come to grips with the crucial role of proof in mathematics. Proofs generate evidence, but they do much more than that, being also generic methods that can be reapplied in other settings than those where they were first constructed. Going beyond mathematics, much learning is about general cognitive *skills*, but dynamic-epistemic logic has not had anything substantial to say about this 'know-how' versus 'know-that'.[22] What seems missing in our earlier picture of rational agency is an account of *methods*, in inference, but also for computation, and other tasks. Some of this is happening in dynamic logics of games [5] that also contain explicit strategies or plans of action. But we have no good account of the dynamics of creating and modifying plans. What we need is an integration of proof theory and dynamic logic – but this is an open problem, also in other settings.[23]

**Conclusion** We have seen how logics of agency and mathematical proof can be put side by side, but in doing so, we mainly discovered new open problems for investigation.

## 5 Further dynamic patterns in mathematics: from proof to belief revision

Information, knowledge and proof are just one aspect of mathematical activity. We can use our more general picture of agency to unveil more of its interesting features.

**Belief revision** In daily life, knowledge is usually too hard a currency. Most of what we say and do is driven by *beliefs*. And this is not a concession to stupidity and ignorance, since we have sophisticated ways of revising beliefs when they go bad. True rationality shows in adversity. Indeed, be-

[22] See [15] for a pioneering discussion of this important point in epistemic logic.

[23] Cf. [1] on proof theory vs. model theory in studying processes, and the related distinction between 'logic about process' and 'logic as process'.

liefs are much too important to leave to the psychologists, or the popular press. True, but does pure mathematics involve beliefs in any but an autobiographical way? Can we find a foothold for belief revision theory concerning mathematical theories? I think we can, because belief-contravening surprises and the resulting theory revisions are an essential part of science, too. The quality and power of science shows precisely in the way it learns from mistakes, and corrects itself, and revision mechanisms are therefore essential. I think this is true even for mathematics: incorrect proofs get re-analyzed, problematic theories get changed, and these processes and dynamic practices seem as important in understanding the stability of the discipline as any Hall of Fame of established theorems.

Against this background, here is a modest goal: can we extend our dynamic logics for belief revision to deal with mathematical reasoning? We briefly discuss one way:

**A first attempt** A first semantics for belief change might work as before in the epistemic case. We enrich worlds with sets of syntactic formulas, and distinguish implicit beliefs $B\varphi$ from, say, explicit beliefs $B\varphi \wedge EB\varphi$, where the latter grow through acts of inference or other awareness-raising events. This is feasible, with logics as before, now merging the earlier dynamic doxastic systems with awareness structure.

But this is still not fine-grained enough. We would be analyzing beliefs about the world, based on incoming information, and the extent to which we have made these beliefs explicit to ourselves. Here crucially, our account of belief change presupposed genuine variation in the underlying sets of worlds, ordered by relative plausibility, and belief change was about changes in that world ordering. But this picture fails for mathematics when we focus on one particular model. In particular, our earlier weak introspection condition that sets of entertained formulas be the same between epistemic alternatives, implies that, when we know a single target structure already (say, again the natural numbers $\mathbb{N}$), there can be no further variation in associated sets of formulas. But this is wrong. We want to be able to say that we believe that Goldbach's Conjecture is true, as a statement about one single world $\mathbb{N}$. How can we achieve that?

**Plausibility syntax models** We just sketch a format, without complete definitions. We now allow any set of formulas attached to our worlds. The sets of formulas $X$ in these pairs can be seen as the formulas 'considered true' at $(w, X)$. In full generality, there need not be any systematic coherence constraints on these sets, but we can think of the whole as a possibly *nonstandard valuation* sending all formulas in $X$ to 1, and those in its complement to 0. Viewed in the latter style, this format has the same generality as *impossible worlds* in paraconsistent logics.[24] Of course, a model need not contain all pairs $(w, X)$, and in excluding some, it may already encode constraints on what agents know. If all sets in the family of available pairs contain some formula $\varphi$, then we may consider $\varphi$ as already explicitly known about the underlying structure.

Next, we postulate *plausibility relations directly between pairs* $(w, X)$, $(v, Y)$, not reduced to any relations on separate components.[25] Agents' beliefs are then expressed by formulas that are present in the sets of what we consider *the most plausible pairs* $(w, X)$. An immediate question is if there is a mathematical basis for such plausibility relations. We do not have a concrete proposal, but will mention one option below.

Again, this setting invites a look from different directions. We start with a theme in mathematics viewed from our logics of agency. A central issue with knowledge update was finding the right dynamic acts. So, what natural acts create or modify belief? In mathematics, inference comes to mind. For a start, acts of hard information in our new setting are *deductive inferences* $\mathbf{P} \Rightarrow C$, placing the conclusion in all awareness sets. Knowledge grew when the premises already occurred in all sets present in the model. But for the purpose of generalization, we reformulate the mechanism slightly:

**Classical inference as hard information** With worlds viewed as possible non-standard valuations, think of an inference rule as a *constraint* between truth-values for the premises and the conclusion. Adopting an inference rule $\mathbf{P} \Rightarrow C$ is then the *hard public announcement* that only valuations remain where truth of the premises implies truth of the conclusion. If the access set $X$ contains all formulas from $\mathbf{P}$, it should also contain $C$. Thus, we now have introduced a relation between truth values for premises and conclusions as explicitly represented in our worlds.[26,27] On this basis, we can now move on:

**Default inference as soft information** For belief change, one interesting act is still inference, but this time not deduction, but non-monotonic *default inference*. Intuitively, a default inference does not say that the conclusion $C$ must always hold, but that, given the premises $\mathbf{P}$, drawing the inference *makes it more plausible* that $C$ holds. Thus, a default inference is like the earlier upgrade act $\Uparrow\varphi$: no worlds are eliminated, but *the plausibility ordering changes* in favour of $\varphi$. Likewise, as a first stab,

[24] The set of all formulas is a possible $X$, modeling inconsistency of our theory.
[25] This would also be an interesting generalization to explore in the epistemic case.
[26] We omit more detailed comparisons with our earlier formulation of inferential update.
[27] We forego some complications with modeling *refutational uses* of inference acts, where we may have to work with formulas that are 'accepted', 'refuted', or 'neither'.

among pairs $(w, X)$ with $\mathcal{M}, w \vDash K\&\mathbf{P}$, and $\mathbf{P} \subseteq X$, a default inference $\mathbf{P} \Rightarrow C$ makes pairs with $C \in X$ more plausible than pairs where $C$ is absent from $X$.[28]

This is only one of several formulations that come to mind. But that is fine, since the precise ways in which this can be done will show the same variety as in belief revision policies: it would depend on how much force we assign to the particular default rule.

Default inferences are important in common sense reasoning, and in science (witness the non-monotonic nature of the usual accounts of confirmation or explanation), Do they also make sense in a mathematical setting? [12] argue that classical dialectics assumed that statements become more plausible, even in a deductive setting, when they have survived a new round of attempted refutation.

We will not develop all these ideas in any further technical detail, but hope they are suggestive. See [26] for a further development, including the soft informational role of default inference and its effect on beliefs.

**From mathematics to agency: revolutionary revisions** Conversely, the concrete domain of mathematical reasoning again offers new ideas for a logic of general belief-revising agents, for instance, by paying attention to the procedural details of how they do so. As I said before, there is much fine-structure to scientific reasoning that we tend to neglect in logic or epistemology. Think of acts of *suspending belief* in hypothetical reasoning, or to a researcher's attitude of *being in two minds* when simultaneously exploring proofs and counter-examples for an assertion.[29] But here I conclude with another challenge, the striking phenomenon of *inconsistency* in mathematical theories:

Suppose that deduction has found a contradiction in our current theory, a trigger for belief revision if ever there was one. In terms of the earlier model, we would now only have worlds $(w, X)$ left, where $X$ contains some formula $\varphi$ and its negation.[30] This challenges our dynamic approach to belief change so far. Clearly, the contradiction can no longer be modeled by a mere plausibility reshuffling of worlds. The model itself becomes a point of contention. We now have to revise the conceptual framework it was based on, throwing away axioms, or even changing the whole language.

To me, this calls for a *revolutionary belief revision* in a Kuhnian sense, as opposed to normal science-type belief revisions that can be dealt with by plausibility changes of given models in the above dynamic logics. But

[28] Technically, this is a special case of the earlier $\Uparrow\varphi$, only in a definable subdomain.
[29] For an implementation of this dual method, cf. the well-known semantic tableaux.
[30] We could then add all formulas, but this would be just uninformative 'rubbing in'.

I admit that these are just names, not solutions – and this remains to be incorporated in our logics of agency.

The aspect of language change in all this is quite faithful to mathematics:

**Language and conceptual dynamics** A crucial aspect of mathematical practice is the creation of new notions, hand in hand with proof. It has often been observed that, despite an emphasis on valid consequence as the measure of all things, the reality of modern logic has long put *definability* and meaning on an equal footing with deduction.[31] For instance, returning to the foundational question of consistency, many of our best correction moves involve changing the language, or a whole conceptual framework. We often resolve contradictions in discourse by sharpening meanings, and contradictions in scientific theories are often resolved by new distinctions (cf. [27]).

As we have admitted, no dynamic logic so far sheds any deep light on this phenomenon – though the systems that we gave are certainly compatible with language change.[32]

## 6 Further aspects of agency in mathematics

I am almost done, but will just list some further topics that would merit systematic comparison. Science creates more complex *theories* than the (presumably) simple common sense knowledge that guides our daily lives. Even so, structured notions of theory structure that have been proposed for agency would bear comparison with those in mathematics. Likewise, we emphasized the importance of *questions* to rational agents, to give direction to what they are doing. But in mathematics, too, actions are not blind uses of available inference steps. Proof search has a purpose, and it proceeds on the basis of beliefs and experience.[33] Mathematical research comes with both local and global agendas of issues to be resolved, and we should understand the dynamics of that, too.[34] Next, we have seen that one only gets to an explanatory dynamics of human behaviour by considering the crucial phenomenon of *evaluation* that guides our choices and actions. At some level, this is also crucial to mathematics. While people are fond of saying that mathematical truth is objective, and achievement an 'absolute' feature, the reality is that 'importance' drives mathematical progress and esteem, just as much as in other areas of intellectual activity. Papers

[31] The third major theme since the 1930s is surely the theory of computation, of which our logical dynamics of agency is a successor – in a suitably modern sense of computing.

[32] There are a few attempts at incorporating language change in logic: cf. [21].

[33] Similar points have been made recently by Jaakko Hintikka on the importance of 'strategic aspects' in reasoning.

[34] Cf. the Stanford course of George Smith on 17th century physics, [22].

in mathematical journals get rejected for incorrectness, but much more often, for lack of importance. Careers are made in terms of importance of contributions, as judged by the community. Thus, there is a dynamics of preference and taste underlying the field. A final aspect of agency that needs to be mentioned is its *interactive social* character. While science is often associated with individual insight into the truth, separate from the usual social graces, the reality is the opposite. Science is one of the most evident and successful forms of social organization that humanity has developed. And mathematics is no exception. Theories are community constructs, and the certainty of mathematics has much to do, not with the brain power of individuals, but the ever-turning grind-stone of many minds absorbing and using new propositions. Indeed, [23] has proposed that formalization is the ultimate form of 'democratization' of science, serving the primary purpose of communication and reproduction of thoughts in other minds.

## 7 Conclusion

Logics of general agency meet mathematics at two levels. First, dynamic calculi strive for the same technical standards as their 'static' predecessors, and thus mathematics is essential to their design and study. The less obvious encounter arises when we view the mathematical activity itself through the lense of dynamic logics. I have suggested that new features become visible then that are worth contemplating. In doing so, it soon became clear that this is not simply reforming mathematical logic with agent logics. In terms of immediate benefits, agent logics rather seem to learn from mathematical practice, since it offers such a rich and well-defined set of cognitive skills. More concretely, it suggests a procedural fine-structure underneath existing logics of agency.

Still, I would hope also for a beneficial converse aspect, changing the unthinking identifications of logical analysis of mathematics with foundational research and formal systems. Using logic to get closer to practice would have great benefits, if only to make mathematicians feel that logic actually talks about their discipline *at all*, instead of some self-created world of formal systems (cf. the criticisms in [18]). More concretely, I have suggested that a logic of belief revision and theory correction can contribute to old foundational questions, by giving a better account of the *dynamic stability* of mathematics. Like in general agency, theories that stand refuted are replaced by more sophisticated ones – and it is in that much richer rational process that the safety and stability of science resides. By contrast, Hilbert's Program of proving consistency both asks too much and does too little, as it does not analyze the former phenomena.

The dynamic agenda extension also shifts traditional battle lines. In the philosophy of science of the 1960s, people felt one had to choose between neo-positivist logical analyses of reasoning and theory formation versus Kuhn's historical and sociological accounts of normal scientific activity and occasional framework-changing 'revolution'. Faced with that dilemma, many chose against logic. But revolutions are not necessarily irrational phenomena: they clearly involve belief revision, language change, and agenda change. But these are all crucial features of rational agency, and there is no reason at all why they could not be incorporated into a modern logical view of science.

Finally, expanding the logical agenda also has a social benefit. On a narrow conception of logic, the purest form of rationality is mathematical proof, and everything else is either a watered down approximation of that ideal, or just an instance of 'irrationality'. But that is dangerous, since it surrenders to irrationality most of the world of ordinary human behaviour, while rationality gets just a tiny rarified corner.[35] By contrast, I am an optimist, pleading for the opposite cutting of the cake, seeing that there is an enormous amount of rationality to our ordinary lives – while mathematics shows what we can achieve when we harness some of that general intelligence to one fixed purpose.

**Acknowledgment** I thank Manuel Rebuschi for several useful comments. In addition, Franck Lihoreau suggested many features of actual reasoning that challenge my simple dynamic model of awareness raising steps. They all point to the need for a richer inferential structure, including evidence, and perhaps a more structured background as in argumentation theory. I find this persuasive, but it seems a subject for another paper – one that I would love to write.

**Dedication** Together with Gerhard Heinzmann, I have engaged in pleasant and useful enterprises. In particular, with our friend and colleague Henk Visser, we once edited a book called *The Age of Alternative Logics* [9], documenting a lively Nancy conference organized by Gerhard on how the current plethora of alternative logics might come to influence the philosophy of mathematics. This paper could have fit there, but I do not see things so much in terms of 'alternatives'. Logical dynamics is not alternative medicine: it rather proposes an agenda extension of logic, while sticking to classical standards. I believe that Gerhard's work represents similar views. And also, while I am all for a dynamic turn in logic, some things in life had better remain static: let our friendship persist!

[35] I also take this to be a central point in [16], leading to a theory of dialogue postulates for genuine conversation and 'communicative competence'.

# BIBLIOGRAPHY

[1] S. Abramsky & J. van Benthem, in progress, 'Logic as Game versus Logic of Games: Categorial versus Modal Approaches', working paper, ILLC Amsterdam & Computing Lab, Oxford University.

[2] A. Baltag & S. Smets, 2008, 'A Dynamic-Logical Perspective on Quantum Behavior', In L. Horsten & I. Douven, eds., Special Issue on Applied Logic in the Methodology of Science, *Studia Logica* 89, 185-209.

[3] J. van Benthem, 2001, *Logic in Games*, Lecture Notes, ILLC Amsterdam. Revised version to appear in *Texts in Logic and Games*, Springer.

[4] J. van Benthem, 2007, 'Logic Games, From Tools to Models of Interaction', in A. Gupta, R. Parikh & J. van Benthem, eds., *Logic at the Crossroads*, Allied Publishers, Mumbai, 283 – 317.

[5] J. van Benthem, 2008, 'Een Postzegel Vol Logica', *De Gids*, Amsterdam, March 2008, 191–205. Also available as 'A Stamp Full of Logic', to appear in V. Hendricks, ed., *Yearbook of Philosophical Logic*, Automated Press, Copenhagen.

[6] J. van Benthem, 2009, 'The Information in Intuitionistic Logic', *Synthese* 167:2, 251–270.

[7] J. van Benthem, 2010, *Logical Dynamics of Information and Interaction*, Cambridge University Press, Cambridge.

[8] J. van Benthem, to appear, 'The Logic of Empirical Theories Revisited', *Synthese*.

[9] J. van Benthem, G. Heinzmann, M. Rebuschi & H. Visser, eds., 2006, *The Age of Alternative Logics*, Springer, Dordrecht.

[10] J. van Benthem & M. Martinez, 2008, 'The Stories of Logic and Information', in P. Adriaans & J. van Benthem, eds., *Handbook of the Philosophy of Information*, Elsevier Science Publishers, Amsterdam, 217–280.

[11] J. van Benthem & F. Velazquez-Quesada, 2009, 'Inference, Promotion, and the Dynamics of Awareness', to appear in *Knowledge, Rationality & Action*.

[12] B. Castelnérac & M. Marion, 2009, 'Arguing for Inconsistency: Dialectical Games in the Academy', Philosophical Institute, University of Montréal.

[13] P. Égré, 2004, *Attitudes Propositionnelles et Paradoxes Épistémiques*, Ph.D. dissertation, Université Paris 1 & IHPST.

[14] R. Fagin & J. Halpern, 1987, 'Belief, Awareness, and Limited Reasoning', *Artificial Intelligence* 34:1, 39–76.

[15] P. Gochet, 2006, 'La Formalisation du Savoir-Faire', Lecture at Pierre Duhem Colloquium IPHRST Paris, Philosophical Institute, Université de Liege.

[16] J. Habermas, 1971, 'Vorbereitende Bemerkungen zu einer Theorie der Kommunikativen Kompetenz', in J. Habermas & N. Luhmann, eds., *Theorie der Gesellschaft oder Sozialtechnologie*, Suhrkamp, Frankfurt, 101–141.

[17] J. Hintikka, I. Halonen & A. Mutanen, 2002, 'Interrogative Logic as a General Theory of Reasoning', in D. Gabbay, R. Johnson, H. Ohlbach & J. Woods, eds., *Handbook of the Logic of Argument and Inference*, Elsevier, Amsterdam.

[18] Ph. Davis & R. Hersh, 1980, *The Mathematical Experience*, Birkhäuser, Basel.

[19] I. Lakatos, 1976, *Proofs and Refutations*, Cambridge University Press, Cambridge.

[20] E. Nagel, 1961, *The Structure of Science*, Harcourt, Brace & World, New York.

[21] R. Parikh, 2009, 'Beth Definability, Interpolation and Language Splitting', CUNY New York. To appear in J. van Benthem, Th. Kuipers & H. Visser, eds., Proceedings Beth Centenary Symposium 2008, *Synthese*.

[22] G. Smith, 2009, Seminar on 17th Century Astronomy, Program in History and Philosophy of Science, Stanford University.

[23] F. Staal, 2007, 'The Generosity of Formal Languages', in *Proceedings of the Second Workshop on Asian Contributions to the Formation of Modern Science*, Amsterdam, May 18-20, 2006. *Journal of Indian Philosophy* 35/5–6: 405–626.

[24] A. Szabó, 1969, *Anfänge der Griechischen Mathematik*. R. Oldenbourg, München & Akadémiai Kiadó, Budapest. 1978 English edition: *The Beginnings of Greek Mathematics*, Reidel, Dordrecht.
[25] F. Velazquez-Quesada, 2009, 'Inference and Update', *Synthese (Knowledge, Rationality and Action)* 169: 2, 283–300.
[26] F. Velaquez-Quesada, 2010, *Small Steps in the Dynamics of Information*, Dissertation, ILLC, University of Amsterdam.
[27] O. Weinberger, 1965, *Der Relativisierungsgrundsatz und der Reduktionsgrundsatz – zwei Prinzipien des dialektischen Denkens*, Nakladatelství Ceskoslovenské akademie Ved, Prague.

Johan van Benthem
Amsterdam & Stanford
Johan.vanBenthem@uva.nl