Thank you for agreeing to participate in *Formal Philosophy: Interviews*. Please answer in whatever detail you deem relevant the 5 questions below bearing in mind that the interview should be self-containing. With approximately 20 interviews we estimate that the complete manuscript comprises between 150-250 pages. Once completed please send the questionnaire either electronically or by ordinary mail (fax) to either one of the editors. Contact information is to be found in the footer.
Yours truly,
Vincent F. Hendricks & John Symons

Φ

Name (to appear in publication): Haim Gaifman

Position (to appear in publication): Professor of Philosophy

Affiliation (to appear in publication): Columbia University

The 5 Questions

1. Why were you initially drawn to formal methods?

As an undergraduate I specialized in mathematics, philosophy and physics. My first result (the equivalence of context–free and categorial grammars) was a mathematical answer to a linguistically motivated question. After graduation I planned to write a thesis in the foundations of probability, sponsored by Abraham. Robinson, while being Carnap's research assistant—a situation that reflected well my dual interests. My eventual dissertation under Tarski was however on different, purely mathematical subjects (not even, properly speaking, in mathematical logic), and in most of my academic career I was based in the mathematics department of the Hebrew University. I researched at the same time, and wrote occasionally on philosophical subjects, holding various visiting positions in philosophy and computer science. Given this spectrum of interests, the use of formal methods in philosophy seemed a natural choice. It was not ideologically motivated; when I was a student Spinoza was my favorite philosopher and Nietzsche was a looming figure. The choice was dictated by the necessity to work on subjects in different fields that formed somehow a continuous group. There are so many things that one can do.

2. What example(s) from your work illustrates the role formal methods can play in philosophy?

Strictly speaking, the use of formal methods means that some formal (or mathematical, or semi–formal) setup is offered in order to model, or analyze a given subject. More broadly, it can mean the inclusion of a sufficient amount of formal or mathematical items, which cannot be avoided if the subject requires it. Thus, a philosophical analysis of Gödel's result (exemplified in my "What Gödel's Incompleteness Result Shows and Does Not Show", Journal of Philosophy 2000) must address some technical aspects, but, does not amount to "use of formal methods". My paper on Dummett ("Is the "Bottom-Up" Approach from the Theory of Meaning to Metaphysics Possible?" *The Journal of Philosophy*, 1996) certainly does not, nor do my recent works in the philosophy of mathematics (also on my website). Among more recent works, the role of formal methods in the strict sense, is exemplified in:

"Pointers to truth" (1992 Journal of Philosophy)

"Pointers to propositions" 2000 in *Circularity*, *Definition and Truth* (ed. Chapuis and Gupta)

"Vagueness, Tolerance and Contextual Logic" (Since 2002 on my website http://www.columbia.edu/~hg17, I still have to get down to publish it).

"Reasoning with Bounded Resources and Assigning Probabilities to Arithmetical Statements" (2004, Synthese).

I should clarify that, in general, formal modeling are not offered as realistic pictures of actual reasoning. It is not claimed, in my probabilistic modeling (in earlier works), that human beings actually compute Bayesian probabilities. Neither is it claimed that the pointer–evaluation algorithms is how in fact we judge truth and falsity in complicated situations. The point of the model is to uncover certain basic mechanisms that, in principle, are operative in our thinking. This applies across the board, the prime examples being the modeling of logical reasoning in Boole and Frege. Of course, the model can evolve a life of its own: in practical applications (e.g., the use of Boolean algebras in computer science), or in the service of a metaphysical program (e.g., *Principia Mathematica*).

3. What is the proper role of philosophy in relation to other disciplines?

I take it that we should view this from the point of view of analytic philosophy. But since the phrased question is about philosophy *tout court* let me first take the philosophical license of considering the question itself. 'Philosophy' is now applied to a congeries of writings and discussions, for which the question of role cannot be answered except in vague generalities. What ball park are we in? this is up for grabs; we have "philosophy of friendship", "philosophy of sport" and what not. Some is done by clever writers, a sort of intelligent journalism, which can make interesting reading on a plane. Adhering to a more traditional framework, the spectrum is still extremely broad, from Frege to Heidegger (not to mention trendy concoctions by Lacan and Zizek). The beginning of this story, in Greece, may therefore be used to give us orientation. Originally, philosophy was a total enterprise, a "theory of everything". (It was also, for some groups, a practical way of life—an aspect I shall, for obvious reasons, ignore here.) It aimed, naively and profoundly, at a picture of the world. When Hamlet says "there are more things in heaven and earth, Horatio, then are dreamt of in your philosophy" he means by 'philosophy' a general frame of knowledge; 'natural philosophy' is another surviving testimony to the original philosophical ambition. What distinguished it from verbal art was the use of the thinking faculty as the major tool in constructing the picture. This overall conception involved, as it must have, systematic use of metaphor (analogical thinking). Nietzsche claimed that that type of knowledge is based on worn out metaphors and is therefore doomed to error. Large scale metaphor is indeed essential to science, which constructs the unfamiliar from familiar materials. But in science, systematic repetitive use, the very thing that kills metaphors in art, gives them life. Their survival, moreover, depends on their passing the repeated severe test of success; metaphors are refined and modified in a continuous feedback; as new forms become familiar to the professional, they serve as a basis for newer more abstract ones. (Mathematics requires a different analysis, the fundamental role there is played not by metaphors, but by structures, or basic modes of organization; I cannot however enter into it here.) It has been often observed that scientific disciplines, at least some of them, have their roots in philosophy. One can say that philosophy, properly speaking, is the enterprise that, avoiding the sacrifice involved in narrow specification and the methodology of experimental success, remains faithful to the goal of "true picture", or "basic account". This still leaves open the kind of building blocks and the kind of tools employed in giving the account. One major divide is between those that place human experience and interests at the center, most notably phenomenological philosophy (exemplified by Husserl and Heidegger) and those who make place for brute scientific facts, as an independent ingredient in the picture. To put it bluntly and somewhat naively, whatever the insights in Heidegger's account, or whatever it truths, it is highly probable that 10 million years ago there was no *Dasein* , i.e., no basic structures of the existential *human*, and no Being of entities (unless the dinosaurs, or some extraterrestrials, were the source of some such structures); and it is highly probable that *Dasein* and Being will not subsist, two hundred million years from now (unless, again, humans migrate to other planets, or there are extraterrestrials, etc.). But there were and will be stars and galaxies, and the truth of '$2^{25964951}-1$ is a prime number' will not be affected (the truth of certain English statements does not require that English, or any language, exist at the time in question). Heidegger may regard this naïve claim as fundamentally misguided, since the entities of physical science are derivative: the outcome of the "objectifying" processes in the *Dasein*; or he might regard it as philosophically uninteresting. But I cannot help being dogmatically impressed by such objective facts and by science's claims (at least some of them) to reveal non–trivial truths . This does not mean that I endorse "scientism"—a reduction of "everything" to science. *Meaning is* basic; but whatever the picture, it will have to accommodate some scientific truths as a self– standing elements, not merely as human constructs.

The science–friendly (and mathematics–friendly) attitude inclines one naturally to emphasize clarity and precision in the choice of tools. The prime example is Frege, whose analysis of logical structure, in thought and language, was achieved by setting up, in his words: "a formula language, modeled upon that of arithmetic, for pure thought". This does not obviate the use of metaphors, of which Frege availed himself freely, e.g., the distinction between *saturated* and *unsaturated* entities, which marks the essential difference between objects and concepts (or, more generally, functions). His proposed analogy between *moon*, *moon's projection in the telescope*, *moon's projection on one's retina*, on one hand, and *reference, sense, subjective associated idea*, on the other, is a rather crude didactic aid, which is dispensable. Not so the subtle metaphor that underlies his characterization of the sense of an expression as "the mode of presentation", or "the way the reference is given to us". (This metaphor, I think, is often misunderstood.)

To come back to the question, the proper role of philosophy derives from what remains of its initial ambitions, given the accumulated intellectual history of the human race, including of course science. It is *au fond* a way of knowing, of apprehending of "getting it". What can this way do for us now? Wittgenstein, in the *Tractatus*, boils it down to "elucidations", which "…make clear and delimit sharply the thoughts which otherwise are, as it were, opaque and blurred." In his later period he seems, in some pronouncements, to restrict it further to a therapeutic activity that cures philosophers by dissolving their confused questions. I think that "elucidations" is nearer to the mark, provided that we interpret the terms broadly, unencumbered by the Tractarian framework. Elucidation can be a great creative project, like the elucidation of logical categories and logical structure, by Boole, Frege, Peirce and others; or the elucidation of what *algorithm* means, proposed by Church and Turing. Naturally, I give examples that are nearer to my professional interests, but the principle is wide. Conceptual clarification, an analysis of plausible approaches and how they are related, of what is implied by this or that view, can be invaluable in ethics as it is in probability or foundational physics. Philosophy has moreover a similar task with regards to its own history; it rediscovers, reflects on, and critically reconstructs its past. It goes without saying that nothing is implied here concerning formal tools, whose justification depends on how and where they are used.

The answer, given from the perspective of analytic philosophy, makes no claim of exclusivity. In a more general perspective, any "systematic" thinking that shows us something significant, worth getting, through the use of intellectual metaphors can qualify as fulfilling the task. It may use a specially designed, dense vocabulary (the reader can judge at the end whether his or her effort was worth it). And it can be "systematic" in being non–systematic, like Wittgenstein's *Philosophical Investigations*, which tries to deliver the picture by an assortment of observations, thought experiments and little fables.

4. What do you consider the most neglected topics and/or contributions in late 20th century philosophy?

I must qualify my answer in two respects: "Most neglected… in late 20th century philosophy" is too dramatic and sweeping for my present mood. Naturally, my views relate to philosophy that treats of logic, language, mathematics, epistemology and metaphysics; that is: not ethics (except that I take understanding and aiming at truth as basic values), not political philosophy, and not aesthetics. In areas related to my own activities, the abundance of published (and e–published) material is such that one always risks missing relevant works. The following is a list of important topics that deserve more philosophical effort than they got.

Waismann's work on open texture merits a follow up and a more systematic analysis that relates it to the analysis of vagueness. The borderline between the two is itself vague, yet paradigmatic examples clarify the distinction. Whether the same sort of fine–grained analysis (not to speak of formal modeling), which has been applied to vagueness, is feasible here remains to be seen. Open texture can make for a better understanding of the analytic/synthetic distinction than Quine's form of behaviorism. In fact, Waismann anticipated Putnam's observations about the splitting of concepts under the impact of new discoveries.

Related to this is the wide subject of analogical thinking, in empirical science and mathematics. In empirical science (perhaps also in certain aspects of mathematical thinking) we have also metaphors. It is to be sure is a difficult subject, but at least we have a lot of examples to go on; even a preliminary sorting of basic parameters should be of great value. Works on mathematical heuristics have been written by mathematicians, but the philosophical task remains. I should note that analogies, which are often marked by "similarly", "by the same token", "in the same way" and their like, can be quite precise; yet, their recognition often amounts to having major insights. Formal logic results, in fact, from recognizing and simulating patterns of reasoning; the back–and–forth moving between different levels of language, "from within" and "from without", is a common technique in mathematical logic. The proofs of Gödel's incompleteness results, as well as the results concerning V=L, are based on these insights. The difficulty of simulating some of these proofs in a powerful automated theorem prover such as *Isabelle* indicates how deep these human insights are.

In mathematics, the subject connects naturally with that of mathematical intuition, including in particular, geometric intuition; it also includes the use of paradigmatic examples. Here is a simple illustration. An interval (by which I mean a *closed* interval) consists of two points and all the points between them (in the strict sense of 'between'). If between every two points there is a point, then the union of two disjoint intervals is not an interval. These concepts are easily accessible, and the claim is immediately recognized as true by anyone (say, a seven–grader) who considers the drawing: A——B - - - - C——D; using another drawing, he or she will also see that the union of two intersecting intervals *is* an interval. We can logically derive the claim from the axioms of linear order, $<$, and the definition: X is between A and B iff $A < X < B$

or B < X < A. The proof—and I am not speaking of a formalized proof—is easy, but requires going through various cases and using repeatedly the properties of <. I noted that, without sufficient mathematical training, highly intelligent college students find the construction of the proof quite difficult. Now, in mathematical practice, some "seeing" of this kind is applied to highly abstract structures; proofs are then written as a check against error and for purposes of communication.

This brings me to the more general topic of *understanding*. Great philosophical efforts were centered around belief and knowledge, not so around understanding. More recently, there have been attempts, mostly in the AI community, to treat is as competence to perform various derivations, pertaining to syntax and semantics, which can be simulated in certain computational networks. But what I mean here is a subject that tries to address questions such as: What does it mean to understand a mathematical proof? (It certainly does not mean that one has gone over a formalized version and checked every step). The point is to investigate understanding without being bogged down by questions about consciousness. I do not know how much a philosophical analysis can accomplish, but it is worth trying.

The general notion of *proof* is another highly deserving topic. Roughly, and in all generality, a proof is an object whose presence in a given context serves as justifiable grounds for belief in a certain statement (where "justifiable" may include probabilistic estimates). Developments in theoretical computer science provide new interesting angles on this. One is the concept of zero–knowledge proofs. Here, the setup consists of two parties, the prover and the verifier; the prover aims to convince the verifier of the existence of a mathematical object satisfying certain constraints (e.g., a proof of a given theorem, or a solution of some equation), without revealing anything else about the object (this can be given precise probabilistic meaning). Both parties follow a certain protocol of exchanging messages, in which the verifier can use random numbers. If the prover has performs successfully to the end, the verifier concludes that the object exists with very high probability, which can be as near to 1 as wished by increasing the number of steps in the exchange. Another, more recent development gives rise to PCP (probabilistic checkable proofs); it establishes the possibility of encoding candidates for proofs in such a way that their correctness can be verified in very short time by methods of statistical sampling (with probabilities as near 1 as wished).

There is place for philosophical work on the concept of *algorithm*. Some work has been done on the Church–Turing thesis, but I think that more is required, including, in particular, evaluations of Gurevich's *Abstract State Machines* and of Moschovakis' *recursors*.

In the foundations of probability, the question: what is a good Bayesian prior? deserves more philosophical attention than given to it. An argument first made by Putnam and further developed in my work with Snir ("Probabilities over rich languages, testing and randomness" 1982, JSL), suggest that priors that enable us to learn more from experience involve an inevitable price in complexity. This shows that there must be a limit to the use of Bayesian methods. At some junctures we must jump to conclusions not via conditionalization. There are theorems to be found, which establish more down–to–earth estimates of the complexity cost for being non–dogmatic. Philosophically, the significance of the limits, in principle, on deductive capacity and the value of *ignoring* information, have been largely underappreciated. Within the framework of probability theory there have been proposals by Hacking, Garber, and lately by myself (in my paper in *Synthese*) for incorporating bounds on resources in the general probabilistic picture. Much more needs to be done, both technically and non–technically.

We still need more comprehensive accounts of the relations between various conceptions of probabilities (subjective, objective, dispositions, frequentists, von Misses's collectives), an account that will also do justice to the fact that in practice we do not have infinite sequences and we decide via methodologies of significance levels and confidence intervals. In my paper "Towards a unified concept of probability" (Proceedings of the 1983 Congress for Logic Methodology and Philosophy of Science), I have sketched a way of "deriving" objective probabilities from subjective ones. Whether this is accepted or not, more needs to be done.

In the philosophy of mathematics we need accounts that reflect more the actual ways in which mathematics has been practiced. The general items mentioned above have direct bearing on this, but let me specify more. The point is not to give an entertaining account of mathematics as a social cultural phenomenon, or of "the mathematical experience", laced with interesting stories, but to do real philosophical work. The fact that there were many mathematical errors and that mathematical history contains significant shifts should be addressed, but at greater depth than the shallow conclusion that mathematical beliefs, like all other human beliefs, are "eternally corrigible". And I do not think that mere history can lead to insights, unless it is guided by a philosophy based on an internal grasp of the subject. Here are some other directions that are important in my view.

It appears that the classical debates between the various brands of constructivism and realism (or Platonism) have come to an impasse. By far and large, mathematics, contrary to the initial expectations of intuitionists and constructivists, uses classical logic and has no qualms about infinities. The reason is that adopting a realistic stance within a classical framework, and accepting various infinities as is needed, is so much more convenient, hence more efficient. The foundational questions remain however. Many philosophers shy away "on principle" from a critical evaluation of mathematical practice, contending themselves with giving some plausible philosophical picture of whatever is accepted by the majority of mathematicians. The worth of this picture, which depends on the presupposed philosophical vocabulary, varies. But in the long run , this tendency is likely to produce a self–contained industry, which settles its internal debates by superficial reliance on what mathematicians do. A philosopher can however, without settling foundational questions, map what he or she considers feasible positions, and what they imply; and here one *does* adopt a critical stance based on a good grasp of the subject. In this direction I would like to mention two projects.

In the foundations of set theory we are in dire need of a good account that explains recent developments and their philosophical significance. Addressed to the non–expert who knows some set theory, this, by itself, would amount to a major philosophical enterprise. In the last forty years set theory has grown in technical sophistication to such an extent that mathematical logicians, who are not among the small number engaged in cutting edge research, find it extremely difficult. At the same time the subject is of high philosophical interest, and there are clear philosophical motivation driving some major technical works. A good in depth explanation is needed also because some published philosophical papers—which try to use recent set theory in support of philosophical claims—are of embarrassing quality, displaying technical errors, as well as a miscomprehension of the role of quantification over classes in set theory. (There is also a wide spread philosophical misconception of second order quantification in general, but this is a somewhat different matter.)

The second topic is strict finitism, the view that certain numeric terms, e.g., $2 \uparrow 5$ (which is 2 with four 2's stacked on top, it comes to 2^{65536}; the definition is: $2\uparrow 1 = 2$, $2\uparrow (n+1) = 2^{2\uparrow n}$), fail to denote existing numbers. Is such a position feasible, and if so what does it imply? In his impressive *Predicative Arithmetic*, Nelson has proposed a framework based on this view, motivated by nominalism: the scarcity of physical objects that can underpin extremely large segments of natural numbers (the predicativity constraint is related, but does not, in itself, imply strict finitism; it only implies inability to prove, on the basis of addition and multiplication, closure under exponentiation). The work motivated related technical research by logicians and computer scientists. Nelson suggested a modified Hilbert program, whose goal is a demonstration of the system's consistency, which does not transgress its basic principles. In an *Erkenntnis* paper from 2000, Iwan marshals previous results in mathematical logic to argue convincingly that *that* goal is unachievable. Questions concerning the feasibility of strict finitism and its philosophical implications remain however. I think that strict finitism should make its claim explicit by incorporating in the system statements of the form "t exists" and their negations, where t is a numerical term (this calls for some version of free logic). It should moreover include closure principles of form: "for all n, $f(n)$ exists" (which means: for all existing n, $f(n)$ exists), where $f(n)$ is a sufficiently slow–growing function; e.g., for all n, $n+1$ exists, or, for all n, n^2 exists. For an appropriate term t, this can be compatible with "t does not exist", since the derivation of t's existence from the closure principles is so long so that it is non–existent. This line of thought can be combined with Yessenin Volpin's suggestion of treating the natural numbers as a bounded vague totality; it is obviously related to the Sorites. The idea was considered by Dummett in his paper "Wang's Paradox"; instead of 'existing number' he used the predicate 'small number'. He argued that this still leads to paradox; but in fact it does not (his argument rests on an interesting oversight). While $2\uparrow5$ is "too small" for a non–existence claim (since it can be reached in $2\uparrow 4$ steps of repeated–squaring), arguments given by Nelson suggest that '$2\uparrow(2\uparrow 5)$ might be a good candidate (perhaps we will have to go higher). I believe that even a strict finitist is committed to some form of realism, that is, to evidence–transcendent truth. The subject merits both technical and philosophical effort.

I believe that questions in the foundations of mathematics are metaphysical; they touch on ultimate aspects of reality. I hesitate to use 'metaphysical' since nowadays it may serve as a rubber stamp, intended to confer gravity on any kind of enquiry. The positivists tried to rule out metaphysical questions as confusions due to misuse of language. They were wrong, but their demise opened the floodgates, and a lot of what passes under 'metaphysics' is the result of pushing forms and metaphors from everyday language, beyond their domain of significance. Historically such theorizing can be very fruitful, e.g., ancient atomism. But we are not in ancient Greece and elementary particles are essentially different entities, though many philosophers do not appreciate this. An attempt to systematize everyday conceptions can be interesting and rewarding. But reading metaphysical debates in mereology whether there is a possible world made of gunk (atomless stuff), I know that something wrong has happened to 'metaphysics'.

5. What are the most important open problems in philosophy and what are the prospects for progress?

I shall point out one: the mind–body problem. I do not think we will ever give it a satisfactory answer. And I shall make one observation about reductionism.

When we reduce physical heat to molecular velocities, we are working within the same framework, and this enables us to derive physical heat from velocity–distribution. We can predict from the distribution what the heat will be, even in situations not encountered before. In other examples, the derivations are not practical, due to complexity. The output (or "move") of a complicated computational interactive system is derivable from its basic state at that moment , which boils down to an assignment of binary values to all memory locations. The enormity of the setup blocks any attempt to specify, even approximately, this state. Yet we know that this is all there is to it, states succeeding states according to the inputs and the system's program. We know how it works, because we have built it. When it comes to human beings, we have no such in–principle derivation. I cannot see any way how something like a succession of enormous array s of 0's and 1's, translates into color sensations or pains. Experimental research might help to set up empirical correlations: such and such a sequence of neuron firings is associated with such and such a sensation. These will make up a list of brute facts. There will be no theory of the kind that derives heat from velocities, or the computer's behavior from its state.

I think there is scope for progress in sharpening the concept of supervenience and clarifying more what is implied by it. We might see more clearly how, in certain respects, supervening entities are different from the ones on which they supervene, or from disjunctions of the latter.

I prefer to stop here, rather than extend my portion of answers, which is a too long already.

I wish to express my appreciation of your project and my thanks for the unique opportunity of thinking and expressing my views on such a wide range of subjects.