

Foundations and Extensions of Logic —An Interview with Timothy Williamson

Hu Liu

Xuefeng Wen

Yuan Ren

Timothy Williamson is one of the most influential contemporary analytic philosophers. He serves as the Wykeham Professor of Logic at the University of Oxford and is a fellow of New College, Oxford. His research has deeply shaped several core areas of philosophy, including philosophical logic, epistemology, metaphysics, and the philosophy of language. He is particularly renowned for three major contributions.

Epistemology (“Knowledge First”): In his groundbreaking book *Knowledge and its Limits* (2000), he flipped traditional epistemology. Instead of analyzing “knowledge” as a combination of belief, truth, and justification, Williamson argues that knowledge is a fundamental, unanalyzable mental state. This framework is widely known as the “Knowledge-First” approach.

Philosophical Logic & Vagueness: In *Vagueness* (1994), he defended a highly striking view called epistemicism. He argues that vague predicates (like “tall” or “bald”) actually have sharp, precise boundaries, but it is inherently impossible for us to know exactly where those boundaries lie.

Modal Metaphysics: In *Modal Logic as Metaphysics* (2013), he explores the logic of necessity and possibility, famously defending necessitism, the view that everything that exists necessarily exists, which carries profound implications for quantified modal logic.

Williamson is also highly regarded for his work on philosophical methodology. In *The Philosophy of Philosophy* (2007), he argues that philosophy is an armchair science but one that is subject to rigorous, formal standards of evaluation, much like any other disciplined inquiry.

This interview focuses on logic, covering fundamental issues in the philosophy of logic, the relationship between logic and metaphysics, philosophical problems in specific areas of logic such as predicate logic, modal logic, the logic of conditionals, and epistemic logic, as well as issues concerning the methodology of logic and research practice in logic.

1 On Logical Pluralism

Q1: In your paper “Is logic about validity”, you argue that the essence of logic is to discover universal, structural truths about the world, instead of discovering valid

Received 2025-11-20

Hu Liu

Institute of Logic and Cognition, Sun Yat-sen University
Department of Philosophy, Sun Yat-sen University
liuhu2@mail.sysu.edu.cn

Xuefeng Wen

Institute of Logic and Cognition, Sun Yat-sen University
Department of Philosophy, Sun Yat-sen University
wxfllogic@gmail.com

Yuan Ren

Institute of Logic and Cognition, Sun Yat-sen University
Department of Philosophy, Sun Yat-sen University
hssreny@mail.sysu.edu.cn

inferences. This view raises a profound question about logical diversity: If classical logic (and other systems you endorse) correctly describes the objective structure of the world, how then should we understand the status of the various non-classical logical systems that have emerged historically and continue to proliferate today, such as intuitionistic or many-valued logics? Should we regard the proponents of these alternative theories as colleagues who have simply made a scientific error, analogous to physicists who once adhered to the phlogiston or ether theories? Or does this proliferation of logical systems reveal something deeper — perhaps a more complex relationship between the objective structure of the world and our human attempts to capture it through language and symbolic systems?

Williamson: I think this attempt to discover universal structural truths about the world is the most basic part of logic. But logic has other branches, and metalogic is one of those. In metalogic, we are interested in discovering which inferences are valid and which are invalid. But the kind of thing that I'm talking about is more basic because, even in metalogic, we are still reasoning, but in the metalanguage. That kind of reasoning has to be correct too. And that has to do with logic which we're using rather than simply talking about.

On the status of non-classical logical systems: Of course, you could have systems where we just mean something different by the logical symbols. But I think that is not the case with, for example, intuitionistic or many-valued logic. They really are intended as more correct alternatives to classical logic. You can see that from the motivation of the people who put them forward. The intuitionists, for example, think that the law of excluded middle, as it is proposed in classical logic and mathematics, is incorrect. It's not just that they want to think in a different way, but that classical mathematicians are making a mistake. With many-valued logics, for example, one motivation is vagueness and the Sorites paradoxes. The reason that people are proposing fuzzy logic is because they think that classical logic is wrong for vague language. They are disagreeing at some level with classical logicians. It is in some ways similar to disagreement in physics. Although some philosophers, like Quine, have attempted to interpret these as not real disagreements, but just a case of using words with different meanings or something like that, that doesn't really fit the way that non-classical logicians have attacked classical logic.

I think we're dealing with genuine disagreements. They're disagreements not simply about the technical notion of validity, but disagreements about the world itself. A classical logician will assert the law of excluded middle, and an intuitionist will deny the law of excluded middle. Of course, intuitionists don't deny particular instances of excluded middle, because that would be inconsistent in intuitionistic logic. But they do deny excluded middle as a universal generalisation. They say it is not the case that for all P , either P or not P . They really are denying what classical

logicians assert.

Q2: I have done things in many-valued logic. One thing I learned is that many-valued logic can be encoded by the two-valued logic. I wonder, what is the usefulness of non-classical logics like many-valued logic?

Williamson: Of course, you're right that simply as a formal system, many-valued logics are well-defined from a classical standpoint. If you're a classical logician, you can do the metalogic of a many-valued logic using classical reasoning in the metalanguage. There's no doubt that these are well-defined formal systems. But that isn't the only use to which they're being put. People who propose fuzzy logic as a solution to the Sorites paradox are proposing it as a better way of reasoning. They're not interested in it just as a formal system. They're interested in it as an alternative way of reasoning in a vague language. It's a genuine disagreement.

In the case of many-valued logic, it can make a difference whether you use many-valued logic in the metalanguage or classical logic in the metalanguage. It seems to me that when people use classical logic in the metalanguage to reason about many-valued logics, they're not taking the many-valued logics very seriously. But if you reason about them using many-valued logic in the metalanguage as well as in the object language, which some people do, then you're taking many-valued logic more seriously. It can make a difference to which principles come out as valid. There are principles that you can use as examples in some many-valued logics. You can show in a classical metalanguage that they're valid on the many-valued semantics. But in a many-valued metalogic, you cannot show that they're valid on the many-valued semantics. It really makes a difference what your metalogic is.

Q3: Logic pluralism has become very popular in recent years. How do you view the competitive relationship between intuitionistic logic, paraconsistent logic, quantum logic, and other deviant logics with classical logic? One view holds that they are each correct within different domains, that different domains require different logics, and that logical validity is only validity relative to its domain and specific models. In this way, these logics do not constitute competition. How do you view this claim? You once said that the proliferation of alternative logics, many of which stem from metaphysical considerations (for example, the rejection of the law of excluded middle or the law of non-contradiction), means that logic is not neutral with respect to metaphysics. So, is the difference between classical logic and deviant logics primarily due to differences in metaphysical stances? Do there exist logical truths independent of any metaphysical reasons (such as the law of identity)? You seem to believe that the reason classical logic prevails is based on abductive reasoning, with one major reason being that classical logic has withstood the test of mathematical history. Do you

think differences in metaphysical stances can be resolved through pragmatic reasons? What is the metaphysical basis of classical logic?

Williamson: There are quite a lot of questions there. One point is, I think all logical truths are metaphysically non-neutral. Even the law of identity, the law that everything is identical with itself, is metaphysically controversial. I had a colleague who was in a German university who told me that it was difficult to have a philosophical conversation with his colleagues who were so-called “continental” philosophers, because they assumed that nothing is self-identical. The law of identity was non-neutral metaphysically. I think they were all mistaken, but I think their wrong metaphysical view was that everything is changing, and that means that nothing is self-identical. I think any logical truth is non-independent of some metaphysical issues.

I don’t think that metaphysical statuses or differences can be resolved just through pragmatic reasons. I don’t think that an abductive methodology is pragmatic. It’s something that is completely consistent with some form of scientific realism about what we’re investigating. It’s just a way of trying to decide which theory is the more promising candidate for being objectively true.

Another question is what is the metaphysical basis of classical logic? I think classical logic in a way is itself metaphysics. Because it’s telling us something structural about how the world is. Even when it says that everything is self-identical, or when it says that everything either has a property or does not have that property, by excluded middle, those are themselves metaphysical claims. It’s not that there’s some deeper metaphysical level below classical logic. Classical logic itself is, in a sense, in this very abstract way, metaphysical. It’s telling us how the world is.

I don’t agree with this idea that all these different logics are each correct in their different domains, because in many cases, they’re competing in the same domain. The case of different logics of vagueness is an example of that. They’re all concerned with the domain given by a vague language. The question is, what’s the correct way of reasoning in this vague language that we have, like English or Chinese? I don’t think that you can resolve the main disagreements just by relativising to domains.

I don’t think that all logical disagreements are primarily metaphysical. Another area where you get a lot of disagreement about logic is in solving the semantic paradoxes, like the liar. The motivations there are not usually metaphysical motivations. They’re more like motivations about what is the smoothest, most elegant way of handling semantic paradoxes. Again, that’s not a pure metaphysical dispute, it’s not metaphysically motivated, although which logical laws you end up accepting is a kind of metaphysical difference, like the difference between asserting or denying the law of excluded middle. But the motivation underlying that may be more logical elegance in some of these cases than it is any difference in metaphysical worldview or something like that.

Q3: Are there many differences from technical reasons?

Williamson: People's choice of logic is usually influenced by technical considerations. Sometimes they're combined with metaphysical views, sometimes with epistemological views. There are many different factors. That's to do with what motivates non-classical logics. But once you have a logic that is non-classical, then there's also some disagreement at the level of metaphysics, as in, for example, with the difference between asserting and denying the law of excluded middle.

Q3: The question asks for the metaphysical justification of classical logic. You said, if I understand right, classical logic doesn't need a metaphysical justification. It itself is already metaphysical.

Williamson: Yes.

2 On Anti-Exceptionalism

Q4: The traditional view of logic holds that logic is an absolutely universal and purely formal normative discipline, with logical truths possessing the characteristics of analyticity and necessity, and the justification of logical truths being non-inferential, a priori, and empirically unrevisable. You evidently oppose such a position and hold an anti-exceptionalist view of logic. According to your view, we can compare theories built around different consequence relations in the same way we evaluate scientific theories, and logic should, like other scientific theories, be subject to empirical revision. Does this mean that logic is closer to natural science than to mathematics? If logical theories accept empirical revision, then is it possible for classical logic to be revised and abandoned? The anti-exceptionalist stance in your recent new book *Good as Usual: Anti-Exceptionalist Essays on Values, Norms, and Action* (Oxford 2024) seems to extend anti-exceptionalism in philosophy and logic to the domain of practical philosophy. Overall, is this a project that permeates naturalism throughout all areas of philosophy?

Williamson: The first point is, does this mean that logic is closer to natural science than to mathematics? No, because I would take the same view of mathematics as of logic. Mathematics depends on logic, and the theorems of standard mathematics are proved using classical logic. If we come to reject classical logic, then we may have to reject some theorems of mathematics. It's not a difference between logic and mathematics. It's just that logic, methodologically, is somewhat closer to natural science than people have thought.

Is it possible for classical logic to be revised and abandoned? Yes, it's possible in the sense that people could do that, and people have proposed doing that. They've proposed revisions of classical logic, which have involved abandoning classical logic. It's possible for people to do that. I regard classical logic as correct, so I think that people who do that are making a mistake, but it's still possible to do it, just as it's possible for people to abandon a correct theory in physics.

Another question is about to what extent anti-exceptionalism is the same as naturalism. Usually when people talk about naturalism, by "naturalism" they mean something like an attempt to reduce everything to natural science, to natural sciences such as physics and biology. I am not trying to reduce logic or philosophy to natural science. In practical philosophy, I'm not trying to reduce normative questions to questions of physics or something like that. I'm just saying that normative questions are more similar than people have thought to physical questions. I take a realist attitude to normative questions, they have right and wrong answers, and the process of finding out which answers are right and which are wrong is not completely methodologically unlike what goes on in the sciences. I think of logic in particular, but philosophy more generally, as a science, but not a natural science. By a science, I mean something much more general. A science is some systematic, critical, evidence-based kind of enquiry. In that sense, logic and philosophy, and also mathematics and history, are all sciences which are not natural sciences. Natural sciences have a methodology that involves experimentation and observation, and I don't think that those methods are particularly helpful for philosophy or for logic, just as they're not particularly helpful for mathematics. But all of these subjects can still be scientific in this broader sense of "science".

Q5: Your anti-exceptionalist view of logic holds that theory choice in logic is continuous with theory choice in the sciences, governed by abductive standards and inference to the best explanation. A crucial component of this theory is the evidence against which theories are tested. While evidence in the empirical sciences is largely perceptual, the nature of logical evidence seems more elusive. One might suggest that our intuitions about the validity of arguments, or the need to resolve paradoxes, constitute the "data" for logic. Indeed, Bertrand Russell argued that a logical theory should be tested by its ability to handle puzzles, just as a physical theory is tested by experiment. However, in the case of logic, intuitions can be unreliable and revisionary — as your own work on vagueness, necessity, knowledge, and conditionals has argued—and paradoxes often admit of multiple solutions. Our question is: How can we, from such materials, establish a sufficiently stable and objective evidential basis for logic, one that is analogous to repeatable experiments in physics? Furthermore, if our logical intuitions are themselves already theory-laden—arguably to a much greater extent than observation is in the empirical sciences—how can we avoid

the charge of vicious circularity when we use them to adjudicate between competing logical theories, for instance, when choosing between classical and intuitionistic logic?

Williamson: Just by looking at this room, we can see that some chairs have four legs, but we can also see by looking at that chair that not all chairs have four legs. That's already some evidence that is relevant, for example, to whether we can infer from "some F are G " to "all F are G ", because we have a counterexample to that just by observing things in this room. I think calling those intuitions is misleading because it makes it seem more mysterious than it really is. These are just observational remarks that we can make by using our senses. But the observations themselves use logical constants like "some" and "not" and "all", so they already supply us with some evidence that we can use in testing logic, so that, for example, we're going to reject a logic which enables us to argue from "some F are G " to "all F are G ".

Of course, there's a possibility that some of our judgments might turn out to be wrong. That's because some of our observations might be wrong. Maybe if we look more carefully we'd find that that chair really does have four legs in some strange way. But the fact that we can sometimes be mistaken and classify false propositions as part of our evidence is a problem in other sciences as well as in logic. It's not a special problem for logic. You say paradoxes often admit of multiple solutions. Of course, it's true with most paradoxes that different solutions have been proposed. But I think that's the same as in natural science, there are many phenomena for which different explanations have been offered, and we have to choose between these different explanations or different solutions to puzzles. We choose using this abductive methodology. Really, there's no deep difference between logic and other sciences in this respect.

Q6: Continuing from the previous question: What precise and complex epistemological role does intuition play in your theoretical framework? On the one hand, abductive reasoning seems to require some pre-theoretic intuitive judgments as initial evidence or data to get the evaluative process started. On the other hand, your practice suggests that many common-sense intuitions should ultimately be treated as revisable theoretical outputs, or even as costs to be paid for the sake of overall theoretical superiority. How do you distinguish between these two kinds of intuitions? Is there a principled criterion for determining which intuitions are qualified to serve as reliable inputs for theory choice, and which should be considered reasonably expendable under the pressure of theory construction?

Williamson: First of all, the category of "intuition" plays no role in my theoretical framework. I'm rejecting that term. But still, some of the questions you're asking

here could be rephrased, not using the word “intuition”. Of course, it’s true that every theoretical enquiry starts from some pre-theoretic position. It starts from a situation in which we already have some knowledge of whatever domain we’re investigating. For example, physics starts with our ability to observe phenomena in the natural world, to observe motion for example. Then it builds on that. In some cases, it may have to end up realising that some of the starting point, the pre-theoretical starting point, is wrong. For example, if you take the case of physics, there is what psychologists call folk physics, which is the pre-theoretic instincts about physical matters that human beings have. For example, if you whirl something round and round and round and then let go, people pre-theoretically expect that its motion will continue to be circular, even though we learn from physics and then from observation that that is false.

There’s no special criterion for determining which parts of our pre-theoretical starting point are immune to revision and which are not. None of them are completely immune in the sense that there’s no aspect of our pre-theoretic practice which is completely immune in principle from revision. We just don’t know in advance which bits we may have to revise and which we don’t. We’re not starting from zero. We’re starting from already having some reasonable ways of finding out about the world. We’re trying to improve those and correct them as we go on. We just don’t know in advance exactly where that process is going to lead. Everything is, in principle, revisable. But at the same time, the default is that we just use our normal methods of acquiring knowledge. Then when we find a need to correct them in some way, we correct them. That’s the same in logic as it is in natural science.

Q6: Is it possible to have another concept instead of intuitions as the “right” concept?

Williamson: A more helpful term is just “pre-theoretic”. Of course, there’s some sense in which we have lots of pre-theoretic judgments. They’re pre-theoretic in the sense that they’re not based on any reflective theorising, but they’re just judgments that human beings make, which may even be exercises of some innate capacities we have. Maybe they’re more culturally influenced than that. Science is this long process of starting from pre-theoretic cognition and then using pre-theoretic cognition to develop more systematic, more reliable, more explanatory kinds of theories. As we go along, we sometimes find that some of our pre-theoretic cognitive dispositions are not a hundred percent correct. We learn that we have to revise them. But that process of starting from the pre-theoretical and then improving and improving, that involves some correction as well as extension. That applies to logic just as it applies to physics.

Q7: Throughout your work, you have, on the one hand, forcefully argued against logical exceptionalism, contending that the choice of a logical theory should, like any

scientific theory, follow a general abductive methodology and be revisable in light of evidence. On the other hand, when confronted with challenges such as paradoxes or vagueness, you seem consistently dedicated to defending the core status of classical logic. A prominent example is your epistemicist theory of vagueness, which, in order to uphold the classical principle of bivalence, introduces what are, for many, quite radical assumptions about our epistemic limitations. This approach seems to demonstrate a strong inclination to prioritize the integrity of classical logic, even at the cost of embracing metaphysical or epistemological commitments that are at odds with our ordinary intuitions. Our question is : Is there a potential tension between these two positions? If, in practice, you consistently tend to revise auxiliary hypotheses rather than logic itself to resolve theoretical difficulties, does this imply that, although logic is not exceptional at a methodological level, classical logic still plays a special, perhaps non-revisable, or at least last-to-be-revised role at the practical level? Does this suggest a de facto preservation of a special status for classical logic?

Williamson: If you look at my book on vagueness, you'll see that classical logic is not given any methodological preference as a starting point. The way that book is argued, I have chapters on various non-classical alternatives to classical logic, such as many-valued logic and supervaluationist logic. I explore those theories in their own terms. I don't just say these are wrong because they reject classical logic. I say, let's see how these work as explanations of the phenomena of vagueness. I argue that they work very badly, that they don't end up clarifying the problem. One of the main reasons for that is higher-order vagueness—borderline cases of borderline cases and that kind of thing. I show that these other theories are pretty bad from an abductive standpoint. I argue that you can give a much better explanation of the phenomena of vagueness by thinking about it using classical logic and using independently motivated epistemological principles about margins for error.

It's not that I'm insisting on some assumptions that you have to make about epistemology. I'm considering an alternative theory of vagueness, which is a theory which embodies classical logic and some epistemological principles. I just argue that on simple abductive grounds, it gives a better explanation of what's going on in vague languages than any of the alternatives that I've examined. There's not really any methodological inconsistency. It's not like a transcendental argument for classical logic. It's an immanent argument for classical logic. When you look at these different explanations, classical logic provides the best one.

3 On Logical Constants and the Metaphysics of Logic

Q8: You argue that logical constants are the result of empirical generalizations, that logic is about the world rather than about a metalanguage, and that logic describes the

most general aspects of the world, enabling us to discover highly general structural principles about how things exist. In this sense, regarding the nature of logic, do you support some form of logical realism?

Williamson: The answer is yes. I completely agree with Bertrand Russell that logic is just as much about the real world as zoology, but just concerned with its more abstract and general features. I think that is a form of logical realism. Anyway, that's one short answer.

Q8: So, logic in your mind is something like *logos* in ancient Greece. Am I right?

Williamson: I'm not a scholar of ancient philosophy, but I suspect that that is right, yes. My guess is that this is a broadly Platonist view, and I'm perfectly happy with that.

Q9: How do you view the neo-psychologism advocated by scholars like van Benthem, which considers logic as a dynamic cognitive immune system capable of correcting errors in real time, as well as the claims of some neo-Kantians that logic is the result of constructive operations based on innate cognitive capacities? Do you believe that cognitive and pragmatic factors play some role in logic? Some scholars argue that we should distinguish between logic as a norm of everyday reasoning (*logica utens*) and logic as a formalized symbolic system (*logica docens*). Do you support this distinction? Do you believe that classical logic serves both as the norm of our everyday reasoning and as the only correct logical theory, such that no distinction between the two is necessary?

Williamson: I'm not a psychologist about logic. I do not think that the main subject matter of logic is psychological. I suspect that van Benthem himself doesn't really mean this in any very strong sense. But we can certainly apply logic to considering questions about how we should reason, at least reason using logical constants. That's a legitimate application of logic. That is normative, which is not purely psychological, because psychology is not primarily normative. But I think the core of logic is no more psychological than the core of mathematics is psychological. In mathematics too, we can talk about how we can reason mathematically, but that's not where the heart of mathematics is.

The distinction between norms of everyday reasoning and logic as a formalised symbolic system is a genuine distinction. Our everyday reasoning is not simply standard classical logic, because, for example, our reasoning with conditionals is more complicated than that. Much of our everyday reasoning involves heuristics, which are efficient but not 100% reliable ways of reasoning. That's very different from

what you get when you formalise classical logic. Everyday reasoning is something very different from classical logic, even though it does to some extent include a lot of classical principles. With everyday reasoning, there is no natural distinction between purely logical reasoning and non-logical reasoning. For example, in everyday reasoning, it's fine to go from "this is scarlet" to "this is red", where scarlet is just a more specific shade of red. That's a piece of good reasoning by everyday standards, but it's not considered on most views a purely logically valid argument, because the connection between the words "scarlet" and "red" is not a logical connection. It might be something that we have very good access to, but it's not purely logical simply because "scarlet" and "red" are just independent non-logical terms, colour terms. There is a gap between everyday reasoning and even a classical symbolic system.

4 On First-Order Logic

Q10: In your two seminal works, "Everything" and "A Completeness Theorem for Unrestricted First-Order Languages," you advance a project on two fronts. On one hand, you offer a powerful philosophical defense for the possibility and necessity of absolutely unrestricted quantification. On the other, you provide a sophisticated formal proof of a sound and complete deductive system for such a logic. Our question concerns the relationship between these two fronts. Your completeness proof relies on a very strong metatheoretic assumption, the "global choice principle", which is a substantive mathematical axiom about the structure of the entire universe. This raises what seems to be a profound question: How ought we to understand the dependency between a universalist logic, which claims to be the foundational framework for all theories, and its own metatheoretic justification? If the proof for this logic's most crucial metalogical property, like completeness, must be secured within a powerful mathematical theory (such as second-order ZFC with global choice), does this imply that the universality of logic is in some sense conditional? Or, does it perhaps suggest that the boundary between logic and mathematics, particularly set theory, is far blurrier than traditionally conceived?

Williamson: I have a few points to make about that. One is that this issue doesn't only arise for unrestricted languages. Even if you look at ordinary first-order logic, where the quantifiers are restricted, the completeness proof for first-order logic depends on a choice principle, although not one as strong as global choice. It also involves set theory, just in the way that the completeness is proved. This is really quite a common phenomenon, that when you look at a logic from outside, from the standpoint of a metalanguage, and then you ask questions about the logic, so that you are adopting some metalogical standpoint, then that's a more sophisticated standpoint than the standpoint of the original first-order object language. New theoretical assumptions

and frameworks need to be used in order to carry out that kind of reflection and answer questions like completeness.

I don't think that makes the universality of logic conditional. The quantifiers in the object language can be completely unrestricted, let's say, first-order quantifiers. But we're adopting a different theoretical standpoint when we start asking metalogical questions about the object language, and we need more theoretical apparatus to ask those more sophisticated questions.

I think the boundary between logic and mathematics is not very deep. I would probably agree that it's blurrier than traditionally conceived, and that there is quite a natural transition between reasoning in the original first-order object language and then moving to the standpoint of a metalanguage and asking metalogical questions about the object language. To answer those, or to some extent even to raise those metalogical questions, we need more powerful apparatus like set theory. If we start from logic and then we start reflecting on what we're doing and wondering on its adequacy, that can easily take us to set theory and mathematics.

5 On Conditionals

Q11: Your book *Suppose and Tell* offers a profound and systematic analysis of conditionals. A long-standing and influential view in the field, famously articulated by Allan Gibbard, is that the significant formal overlap between Adams's logic for indicative conditionals and Stalnaker's logic for subjunctive conditionals is merely a technical coincidence. This "separatist" thesis holds that they are fundamentally distinct: The former concerns degrees of epistemic credence, while the latter concerns metaphysical possible worlds. Your own theory, however, particularly with its emphasis on the cognitive procedure of supposing as central to the evaluation of all conditionals, seems to aim for a much more unified account. Our question is: How exactly does your unified theory respond to Gibbard's separatist thesis? Do you argue that the manifest differences between indicatives and subjunctives arise not from two fundamentally different underlying logics, but rather from the different background knowledge, contextual constraints, and heuristics we apply when running the same core cognitive procedure of supposition?

Williamson: On the view that I propose in the book, the difference between indicative and subjunctive conditionals is semantic. I argue that indicative conditionals are simply in effect the material conditional, and subjunctive conditionals are modalized and contextually restricted versions of indicative conditionals. In English, the way you typically form a subjunctive conditional is by using the word "would". I argue that "would" is itself a modal operator. What we have is a material conditional which is then operated on by a modal operator. The modal operator, the word, is a con-

textually restricted strict conditional. It's a universal quantification over a restricted set of possible worlds. There's a real structural difference between them, but it's one which can be explained in terms of compositional semantics by the interaction of the indicative conditional, which is just the material conditional, with this modal element that is added by the word "would". Of course, different languages handle this distinction in different ways. But I think the underlying cognitive apparatus that we're using to make the distinction is the same for everyone, even though it's implemented in different ways in different languages. It's not just some pragmatic distinction. It's a compositional semantic distinction, roughly between the material conditional and a strict conditional.

Q12: Is your defense of the claim that indicative conditionals are equivalent to material conditionals motivated by a commitment to defending classical logic? It is generally thought that the semantics of counterfactual conditionals typically does not rely on explanations in terms of material conditionals. Can you explain how you achieve a unified interpretive framework for both indicative and counterfactual conditionals? You advocate a "heuristic framework" in the book to account for the differences between the semantics and pragmatics of conditionals. Could you elaborate on this approach?

Williamson: I assume that by "counterfactual" conditionals you mean the same as by "subjunctive" conditionals. I agree that these counterfactual conditionals are not material conditionals. On the view in the book, they are strict conditionals. They are the result of applying this contextually restricted necessity operator, which in English is expressed using the term "would", to the material conditional. The counterfactual conditionals in that sense are semantically more complex. It's true that we use similar heuristics for both kinds of conditional, for the counterfactual subjunctive and for the indicative, but there's an extra layer of abstraction from the actual world with the factual conditionals. Because we're not thinking of the antecedent as just embedded in the actual world, but we're allowing that it can represent a different world from the actual world. That has some specific implications for how we reason with counterfactual conditionals in a slightly different way from the way that we reason with indicative conditionals. But there are a lot of structural parallels between the two. Just a comment about what the difference is the word "would" in English is the past tense of the word "will". But this past tense is being used as what linguists call a "fake past". It's not really concerned with a difference in time. It's using, as it were, the past as a metaphor for a modal difference between the actual world and a counterfactual world. The cognitive apparatus that we're using is repurposing our ability to think about past times and using that to think about counterfactual possibilities. In the book, it gets quite complicated. But that picture is what has been worked out in

the book in detail.

Q13: Metaphysical modality, specifically counterfactual modality, characterizes possibilities in counterfactual situations. Counterfactual situations are typically expressed using subjunctive conditionals. Thus, the logic of conditionals is closely related to metaphysical modality. Could you clarify whether your 2020 work on the logic of conditionals builds upon your 2013 book *Modal Logic as Metaphysics*? How are the two works related?

Williamson: They're consistent with each other, but they're not very closely related. There's more of a connection between the 2020 book and my 2007 book *The Philosophy of Philosophy*. Because in *The Philosophy of Philosophy* what I was interested in doing was explaining metaphysical modality in terms of counterfactual conditionals, or at least our ability to think about metaphysical modality in terms of our ability to think with counterfactual conditionals. Because I wanted to show how our ability to think about something as philosophical and abstract as metaphysical modality was based on our quite common sense capacities to think about counterfactual questions, which are mostly not questions about all possible worlds, or anything like that, but much more restricted.

I slightly revised that 2007 view in the 2020 book *Suppose and Tell* because in 2007 I had been assuming something more like the Lewis-Stalnaker semantics for counterfactuals. When I realised that it was actually better to understand counterfactual conditionals as the result of applying a contextually restricted necessity operator to a material conditional, which is not really consistent with the Lewis-Stalnaker approach, I changed my view — but not in a way which makes very much difference, because still these counterfactual conditionals were contextually restricted strict conditionals. But if you remove all the contextual restrictions or you just treat them as the null or the tautologous contextual restriction, then you get back to a metaphysical modality. You get back to one which corresponds to “in all metaphysically possible worlds”, because once you've dropped all the restrictions, the basic idea that metaphysical modality is simply a special case of the kind of modality that we're thinking about all the time in using ordinary language and in thinking about what people are capable of doing. But the 2013 book is not so much connected with those issues about ordinary cognition of modal matters.

Q14: In *Suppose and Tell*, you anchor the semantics of “if” in the material conditional, while classifying the Ramsey Test, which guides our everyday reasoning, as a practical cognitive heuristic. This distinction raises an interesting question about the relationship between semantic knowledge and cognitive practice. Typically, we consider a semantic theory successful if it can explain how competent speakers use

their vocabulary. However, if speakers' most central cognitive procedures, such as the Ramsey Test, systematically lead them to make incorrect judgments about the truth-values of conditionals, then it seems we must accept that our semantic grasp of "if" is divorced from the cognitive methods we use to process it. Could you elaborate on your view of this semantics-cognition gap? And in your opinion, on what grounds does the empirical adequacy of a semantic theory rest, if that theory treats widespread linguistic intuitions as "useful errors"?

Williamson: You're wondering, given that the main heuristic for assessing conditionals is something like the Ramsey test, why don't we have a semantics that is more closely related to the Ramsey test? The main answer to that is, as I argued, when you look at the general heuristic that we're using, you can see what it is, and that it is implicitly inconsistent with itself. The inconsistency can be shown in various different ways. Some of them are related to David Lewis's results about the probability of conditionals versus conditional probability, but some of them don't depend on that result and just have to do with the principles for assessing conditionals in a non-probabilistic way.

Because our central heuristic for conditionals is implicitly inconsistent, no semantics is going to be able to validate it. Any semantics is going to have to say that we're making mistakes some of the time. I argued that even if you tried to use context dependence or something like that to defend the Ramsey test, or the "suppositional heuristic" as I call it, there would still be no way of semantically vindicating our practice one hundred percent. Any theory is going to have to end up being an error theory about some examples. You then have to look for a more sophisticated relation between the semantics and the use than one might have expected. I argued that the material semantics in certain ways fits the Ramsey test better than any other semantics. It's not in all respects, but in key ways it's more faithful, even though it's not completely faithful.

The general answer is just that the fit between the semantic theory and the linguistic data has to be somewhat more sophisticated, but just in a way that we would expect from most sciences; the underlying explanations are not ones which you can just read off the observable phenomena. You need to have a theory that explains the observable phenomena, and its explanations with a bit of luck will turn out to be the correct ones. The explanation doesn't have to be simply some mirroring of the observation, but it can postulate some non-obvious structure behind these. The structure that I'm using to understand the use is this heuristic, and the heuristic does have some relation to the semantics. But the semantics enables one to see how the practice of using conditionals according to this heuristic will pass on information from speaker to speaker. That's why the book is subtitled *Suppose and Tell*. "Suppose" for the heuristic, "tell" for the communication that preserves this information. One does have to be

more sophisticated about what the relation between theory and data is.

Q15: This is a related question to the last one. In your theory of conditionals, the semantic model itself is kept remarkably simple, while a great deal of the complexity is handled by pragmatics or cognitive heuristics. It's a very elegant architecture. We're very interested to know how you conceive of the interactive boundary between the semantics and these cognitive heuristics. In your view, is a simple semantic core, supplemented by a complex external system of explanation, a more promising theoretical architecture than a semantic model that is itself inherently more complex? Approaches like dynamic semantics attempt to semanticize more pragmatic information. What fundamental difficulties do you believe such attempts ultimately face, which in turn would highlight the value of the division of labor that you advocate?

Williamson: I think you can't tell in advance where the boundary between semantics and heuristics is going to come. That's something that you just have to explore in the theoretical investigation. Different theories may put the boundary between semantics and heuristics in different places. Then the question is, which of these theories gives the best overall explanation? In very broad terms, what I'm proposing is a very simple semantics and a very simple heuristic. It's simple on both the semantic and the pragmatic or heuristic side. It doesn't make very complex general assumptions. It has these very simple moving bits, and then it uses those to explain the phenomena.

If we're going to compare it with dynamic semantics, dynamic semantics may put the boundary between semantics and heuristics in a different place. I agree that it's attempting to make semantics, as you put it, do things that for me are on the side of the heuristics. I don't think that either approach is ruled out on general methodological grounds. You just have to look at the kinds of explanations that they're providing and how much complexity they're assuming. My basic assumptions are simpler than those of dynamic semantics, and they can explain the data just as well. But that's something that you have to test by looking in detail at what kind of explanations these alternatives produce. It's not something that you can settle in advance by some general methodological consideration.

6 On Epistemic Logic

Q16: Your argument against the KK principle in *Knowledge and its Limits* has been profoundly influential. By introducing the margin for error principle, you compellingly argued that cognitive indiscriminability is a non-transitive relation, thereby effectively undermining the foundation of the principle of epistemic introspection. Interestingly, this challenge to transitivity seems to reverberate across many areas of philosophy. For instance, in discussions of vagueness, the core of the Sorites Paradox

lies in the transitivity of the “looks the same as” relation. In social choice theory, the Condorcet Paradox reveals the potential non-transitivity of aggregating individual rational preferences into a collective decision. Even in logic, certain counterexamples to Hypothetical Syllogism challenge the transitivity of the conditional. In your view, is this failure of transitivity in these different domains merely a structural coincidence, or does it collectively point to some deeper, unifying feature concerning human rationality, language, or reality itself? Is there a more fundamental philosophical problem, or a unified explanatory framework, underlying all of this?

Williamson: My view is that it’s more like a structural coincidence. It’s true that in a number of different areas, failures of transitivity do help to shape the enquiry, and they’re often very rewarding things to consider. But I don’t think that’s because they all share a common structure. For example, I don’t think that the reason for the Condorcet paradox, with the potential non-transitivity in aggregating individual rational preferences, has the same origin as the issues of non-transitivity in the cases of vagueness. I don’t think anybody has succeeded in unifying all these different phenomena. If they had, if somebody could find ways in which they are unified, then great. But I doubt it, because I think a lot of the basic structure of these things has been clear for a long time, and nobody’s found a good way of unifying them. I just don’t see why all failures of transitivity should have the same origin.

7 Concepts of Modality and Possibility

Q17: There have long been different concepts of modality, referring to various types of necessity and possibility, such as metaphysical possibility, epistemic possibility, logical possibility, and so forth. How many main types of possibility do you recognize? Do you believe that the distinction between metaphysical possibility and epistemic possibility is analogous to that between objective and subjective possibility? Does this understanding presuppose a subject-object dichotomy, which early analytic philosophy sought to avoid? To which type of possibility does logical possibility belong, and how should it be characterized?

Williamson: There are many specific types of possibility because, for example, when we’re asking what somebody can do, there are all sorts of contextual constraints implied about what they can do now or what they can do if they are allowed to spend a year training to do it and all these things. There are indefinitely many kinds of possibility of a broadly objective kind like that.

I don’t think that logical possibility is a well-defined notion because, presumably, it’s meant to correspond to consistency, but consistency is a syntactic notion, so it’s sensitive to changes of word that don’t involve any changes of content. For example,

if you take two synonymous words like “furze” and “gorse”, which are just two words for the same natural kind, a kind of bush, the sentence “furze is not gorse” is logically consistent, even though it’s false. “Furze is distinct from gorse” is logically consistent just because those are distinct terms and “ a is not equal to b ”, when “ a ” and “ b ” are different terms, is consistent in pure logic. But it’s not, although “furze is not gorse” is consistent, “furze is not furze” is inconsistent. Just substituting one synonymous expression for another leads from something that’s logically consistent to something that’s logically inconsistent. There can’t really be a modality corresponding to that, because modal operators are meant to be operating on the underlying propositions expressed by sentences, not simply by the sentences themselves. Substituting one word for a synonymous word should make no difference to a genuine modality, and therefore logical possibility is not a genuine modality.

For similar reasons, I’m a bit sceptical about epistemic possibility and the way people often think about it, because a mere difference of word can make an epistemic difference, even though it shouldn’t make a semantic difference. My current view is that the most basic kind of possibility is the most general kind, just metaphysical possibility. Every other kind of possibility is just the result of restricting metaphysical possibility by various conditions. Epistemic possibility might be where we’re restricting epistemic possibility in terms of compatibility with some knowledge base.

I think the distinction between objective and subjective doesn’t correspond to “epistemic”. It would correspond more closely to “doxastic”, having to do with belief, because “epistemic” has to do with knowledge, so it’s constrained by the way in which knowledge entails truth. For example, everything that is true is epistemically possible, because it’s compatible with everything that is known, and knowledge implies truth. But it’s not the case that everything that is true is subjectively possible, because people can have false beliefs, and then something true which is inconsistent with their false beliefs is not subjectively possible for them. But because it’s true, it’s still epistemically possible. I don’t think that epistemic possibility and subjective possibility really correspond to each other.

Q17: Does this mean there is no such thing as logical possibility?

Williamson: I’m saying that logical possibility isn’t a well-defined idea, because it’s not a proper modality. I know people talk about logical possibility a lot, but I think they’ve just never thought about these issues about how logic is sensitive to mere differences of word without any underlying difference of meaning.

Q17: There is no such thing, or, there is such thing, we just can’t define it?

Williamson: No, I think there's no such thing. There's no such thing as logical possibility in the strict sense that corresponds in this way to logical consistency. The argument I gave about substitution of synonymous terms shows that there is no such thing as logical possibility that corresponds in the required way to logical consistency. There are notions of possibility which might be a bit like logical possibility, but I think none of them has the strict relation to consistency that is required to count as genuinely logical.

Q18: Do you endorse the two-dimensional semantics proposed by Chalmers and others, which distinguishes between metaphysical possibility and epistemic possibility, using the latter to explain the *a priori/a posteriori* distinction? You oppose the *a priori/a posteriori* distinction, arguing that the traditional philosophical notion of the *a priori* is merely counterfactual imagination based on empirical activities. Given your account of the relationship between counterfactuals and metaphysical modality, does this imply that the distinction between metaphysical and epistemic possibility is also undermined?

Williamson: I don't like two-dimensional semantics. In the case of Chalmers, his basic account of epistemic possibility relies on the *a priori/a posteriori* distinction, so that distinction is basic for him, and I've argued that it's a fairly superficial distinction, so that you can't do anything much with it. But also, I don't think that these epistemic differences should be playing a role in the semantics, because, for example, just to use the same example that I was using when I was talking about logical possibilities, Chalmers, I think, would want to say it's epistemically possible that furze is not gorse. But then, since furze and gorse have the same meaning, it should be possible to substitute one of them for the other, and then you get that it's epistemically possible that furze is not furze, which is false. There shouldn't be a semantic distinction between the words "furze" and "gorse", but I think the way Chalmers is setting things up, he's going to insist on some semantic distinction between them. I think that's not faithful to the semantics of natural languages. I take it that the right semantics for natural kind terms is something like a direct reference semantics that Kaplan, for example, defends. I think they're trying to put cognitive differences into the semantics, but these cognitive differences don't belong in the semantics, because they can just be the result of a difference of word without any difference of meaning, because the meaning isn't something that's completely transparent to us. I really don't like that whole approach.

8 On Modality, Essence, Grounding, and Causation

Q19: Is the metaphysical foundation of modal logic rooted in the necessitism you advocate? Necessitism appears somewhat counterintuitive on the surface, as it asserts that everything that exists does so necessarily. This brings us back to the earlier question: Should logical theory be used to defend metaphysical positions, or do metaphysical positions provide the basis for choosing a logic?

Williamson: One can say a lot about the metaphysical foundations of modal logic without assuming necessitism, because a lot of the interest of modal logic is just in propositional modal logic, which doesn't have quantifiers, so for those purposes you don't really need to worry about necessitism. Necessitism only comes in once you start doing quantified modal logic. I think it's good to do quantified modal logic, but that's not the most basic kind of modal logic, and a lot of the interesting questions about modal axioms arise at the purely propositional level.

In relation to what the argument for necessitism is, I'd agree that if we just start doing modal metaphysics and quantified modal logic in a naive way, we're probably going to reject necessitism. But what I argue in *Modal Logic as Metaphysics* is that once you go a bit deeper, you realise that necessitism is a much more elegant and systematic and simple approach than contingentism, and that the apparent counterexamples to it can be understood quite easily within a necessitist framework. Basically, on standard abductive grounds of simplicity and strength, one should prefer the necessitist view. I don't think that either one of logic and metaphysics is prior to the other. What we're doing here is making an abductive theory choice between different theories that are formulated in a quantified modal language, and we're treating the logic and the metaphysics together.

Q20: There is an essentialist position, supported by Aristotle and Kripke but opposed by Quine, which has recently been developed by Kit Fine and others, who argue that essence precedes modality. You oppose the view that essence precedes modality. Can your approach to this issue be summarized as follows: Modality provides the objective structure, while essence reflects our heuristic-driven interpretations? Does this mean that the concept of "essence" is interpreted as an epistemic concept, and that metaphysical necessity precedes epistemic necessity?

Williamson: I don't want to say that metaphysical necessity precedes epistemic necessity here. It's true that I think Kit Fine's arguments for the idea that essence precedes modality are arguments which rely on some heuristics for assessing essentialist claims, and these heuristics are really sensitive. They are sensitive to epistemic matters like the quality of explanations. I think that we get inconsistent results if we try

to project the results of that heuristic onto the metaphysics. I think that his argument for the idea that essence precedes modality is based on an unreliable heuristic, but I don't think that you have to conclude from that that essence is an epistemic concept. You could just conclude from it that essence can be explained in modal terms in the way that Kripke seems to do in *Naming and Necessity*, and that whereas it seemed that Kit Fine had refuted that explanation, I think that explanation might in fact be perfectly correct. But that explanation does not use epistemic necessity. It's just an explanation in terms of metaphysical necessity, and the epistemic side just has to do with the sources of our mistakes in assessing some of these examples.

Q21: In recent years, the concept of grounding has gained significant attention in metaphysics, with some arguing that its importance in metaphysics is comparable to that of causation in the philosophy of science. Causal relations can be characterized using counterfactual conditionals. On the other hand, as your work demonstrates, metaphysical modality can be analyzed in terms of counterfactual conditionals. Do you think there is an interesting connection here?

Williamson: I think there's some connection. I wouldn't put it exactly in those terms, and I'm not sure how central the idea of causation is in the philosophy of science, because there's a tradition that goes back to Bertrand Russell, but also includes people like Hartry Field, of denying that causation is a fundamental notion in natural science. I think Russell thought of it as a somewhat human-centred way of thinking about the world, and in the philosophy of science, really what you end up with are just things like differential equations, which don't have any special direction of causation built into them.

I think a better way of putting the connection would be that there is a strong connection between modality in metaphysics and modality in science. A few years ago, I wrote a paper called "Modal Science", which is about the modal aspect of natural science. I think that can be found, for example, if you look at phase spaces and possibility spaces and state spaces in science. These are really spaces of a modal kind. They're concerned with a space of possibilities, and possibilities in some broadly objective sense. They're not simply epistemic possibilities. I agree that metaphysics and science are unified to some extent by their interest in the modal dimension. For example, I was interested in the way that if you look at what are known as dynamical systems, which are important in lots of different branches of science, you could articulate them using the language of modal temporal logic. I think there is an interesting project there, which is to articulate the modal dimension of science. In a way, that modal dimension is present wherever you're talking about probabilities, because the mathematics of probability involves probability spaces that involve things like possible worlds, except that they're just called "outcomes" in probability textbooks, but

they're pretty much the same thing.

9 On Methodology

Q22: Your paper “On Intuitionistic Modal Epistemic Logic” is not only profound for its specific philosophical conclusions (about the relationship between knowledge and modality in the stand of anti-realism) but is also highly instructive for its methodology. Instead of directly criticizing the premises of anti-realism from an external, classical standpoint, you delve deep inside it, constructing for it the most rigorous and coherent formal model possible. You then reveal its logical consequences by exposing the internal structural properties of this optimal model, such as its non-conservative extension of the base logic. This method of Internal Exploration and Consequence Mapping, which has also been adopted in your several other papers (such as “Assertion, Denial and Some Cancellation Rules in Modal Logic” and “Truthmakers and the Converse Barcan Formula”), stands in contrast to another, more common approach: External Critique and Foundational Refutation, which tends to directly challenge the coherence of a philosophical position or the truth of its basic presuppositions. Our methodological question is this: In your view, what are the respective advantages and limitations of these two methods of using logical tools in philosophy? The internal exploration method, while more precise and profound, perhaps runs the risk of excessive charity, lending too much formal elegance and systematicity to a theory whose foundations may be unstable. Conversely, the external critique method, while direct, may struggle to foster genuinely constructive dialogue with an opponent, often risking a situation where the parties talk past each other. As a distinguished practitioner, how do you believe we should determine when adopting the internal exploration method is more philosophically fruitful than the external critique? Does this choice depend on the type of philosophical problem being addressed, or more on the philosophical goal we aim to achieve—whether that is to decisively refute a theory, or to more clearly understand the entire intellectual landscape?

Williamson: I think it definitely depends on what your project is. In the paper that you first mentioned, “On intuitionistic modal epistemic logic”, that was written because I wanted to solve a very specific problem, which is concerned with the Church-Fitch argument, that was sometimes called the knowability paradox, which is an argument for the principle that if all truths are knowable then all truths are known, which is a *reductio ad absurdum* of certain kinds of anti-realism. My first published paper was arguing that their proof didn't work in intuitionistic logic. It's not that I wanted to defend intuitionistic logic, but I just wanted to make the point that this particular argument can't always be used, because it doesn't work if you're arguing against an anti-realist who's an intuitionist. It was the American philosopher Roy

Sorensen who raised a question about my result. He said that I [TW] have shown that that particular proof of the reduction from all truths are knowable to all truths are known doesn't work in intuitionistic logic, but maybe some other proof of that conclusion would work in intuitionistic logic. I wanted to show that no proof could be given in intuitionistic logic of that result. I had to build a model of intuitionistic logic with modal and epistemic operators because they're used to formulate those principles about knowability and what's known. I had to show that there were counter-models to that argument. It was for a very specific conclusion. Although I'm not a friend of intuitionistic logic, I wasn't doing any of this to criticize intuitionistic logic but just to show that it was not vulnerable to a certain kind of critique.

In general, because of my abductive methodology, I think, if you take a certain view seriously, it's usually more productive to just see what its consequences are, and to try to unpack the view so that it can be assessed in comparison with other views on the same question by an abductive methodology. That was to some extent what I was doing, for example, in my vagueness book with different theories of vagueness. Sometimes there's no need to go in for that. If you can show that some theory has a really ridiculous consequence, that may be enough, to just give one counterexample to it. But with somewhat better developed or more robust theories, you can't just refute them by counterexamples. You have to look at how they work to understand them from the inside. But I don't think that you always have to use that method, particularly with theories that are less well developed. Sometimes theories just have ridiculous consequences and you can simply dismiss them quite quickly. But sometimes you want to take a theory more seriously and try to understand better what its limitations are, and then you have to work from the inside.

Q23: Your philosophical system presents a profound and fascinating consistency in ontology and logic, you staunchly defend classical logic and strong idealizing principles (like bivalence and necessitism) as parts of our best theory of objective reality; yet in epistemology, you forcefully critique idealized models, convincingly arguing why assumptions like the KK principle must give way to our finite, fallible cognitive reality. This sharp distinction, the “idealized world” versus the “non-idealized mind”, also seems to suggest a methodological dualism. For the former, philosophical inquiry appears to be a quasi-scientific project of construction, driven by theoretical virtues like simplicity and strength. For the latter, it seems more like a fine-grained, descriptive project aimed at accurately characterizing our cognitive capacities. Our question is do you see these as two distinct but complementary core methods of philosophical inquiry? Furthermore, what implications does this methodological division have for other areas of philosophy, such as ethics, aesthetics, or political philosophy? Are their subjects—values, norms, social structures—closer, in your view, to the objective reality that we should “construct” with idealized theories, or to the domain of

human limitation and practice that we must “describe”?

Williamson: I wouldn’t myself describe classical logic and bivalence and necessitism as idealising principles. I defend them as simply universally true, exceptionless, necessary truths and universal generalisations. I think of idealising as something that goes on where we have to introduce simplifications in order to understand the phenomena that we’re interested in and where these simplifications are, strictly speaking, false. For example, when you have a model of the solar system in which you treat planets as point masses, it’s false that planets are point masses. That is an idealisation.

I see the distinction that you’re getting at, which is one between areas in which we can expect strictly true universal generalisations versus areas in which interesting universal generalisations always have counterexamples, and the only way that we can work with universal principles is by introducing some model that is not fully correct, but that casts some light on the phenomena. I take it that in mathematics, you can have true universal generalisations, and maybe in some areas of fundamental physics, there are true universal generalisations. But already in biology, there are practically no true and really informative universal generalisations. The biological world is too messy to allow for many exceptionless generalisations, and therefore you have to use idealised models. Within philosophy, if we’re talking about really fundamental metaphysics, that’s more like the case of mathematics and very fundamental physics, where there is a real chance of exceptionless universal generalisations. Whereas when we’re talking about epistemology, we’re typically talking about human knowledge, although we could also be talking about the knowledge that non-human animals have. There we’re dealing with biological systems that are very messy and complicated, and where we can’t expect many interesting generalisations to be universally true. There may be some, like I think it’s true exceptionlessly, that if you know that something is the case, then it is the case. In other words, knowledge is *factive*. I think that’s universally true. But you can’t do much epistemology just with a few principles like that. Most of the time in philosophy, there’s only a very limited role for exceptionless universal generalisations, and we have to do something a bit more like model building.

I would also think of that as applying to subjects like ethics, aesthetics, and political philosophy., Ethics is about human action, or almost all of it is about human action. Aesthetics is about human art, and political philosophy is about human societies. Those are all very complicated, messy systems, and we shouldn’t expect much in the way of exceptionless universal generalisations. We may be better just to work with somewhat idealised, but not too idealised, models.

Q24: Looking back at the history of philosophical logic, from Frege and Russell to Lewis and Kripke, its golden ages seem to have always been intimately connected with fundamental inquiries into the foundations of mathematics, metaphysics, or the

philosophy of language. In that context, logical tools served as a powerful engine for exploring and clarifying core philosophical problems. However, a current trend in the field seems to be a concentration of work on constructing and refining formal systems. While these systems may be internally elegant, their philosophical motivation is sometimes unclear, and the work can appear more like an exercise in solving internal puzzles than a response to pressing challenges from other areas of philosophy. In your view, is this trend of “involution” or specialization an inevitable refinement of methodology that comes with a discipline’s maturation, or is it a deviation that risks sapping the field of its philosophical vitality? To reinvigorate the role of philosophical logic as an engine for philosophy, what adjustments do you believe should be made in terms of future topic selection or methodology?

Williamson: Increasing specialisation seems to be a more or less universal phenomenon in the academic world, not just in philosophy, but across all the sciences and elsewhere. When you have thousands of researchers, many of them extremely able, working away, you’re going to get some disciplines emerging where it’s as much as anybody can do to master what’s being done in that small area. I think that is how science works, and it doesn’t by itself prevent major breakthroughs. We should be willing to accept that this progress, particularly in logic, where people are solving specific problems in logic that may be quite complicated, are adding to the store of our knowledge of logic. Of course, some of this work will be better motivated than other parts. It’s possible to have work that is not well motivated and doesn’t show anything very interesting. It’s just complicated. But after all, if you think about what happens in mathematics, it’s not that every paper published in a mathematics journal is a classic. People are contributing to the development of mathematical knowledge. It’s sometimes difficult to tell when an advance in one area can help an advance in another. I don’t think one should be too critical of these trends.

It seems to me that if you look at an issue of a logic journal, typically the questions that are being answered are perfectly genuine, reasonable questions. They may not be questions of the most fundamental kind, but their solutions all contribute to progress. What we don’t want is a model where the only kind of progress that we can have in philosophy is one where some genius revolutionizes the whole field. If every article in every journal was revolutionary, we would just be in a state of perpetual revolution that is not really manageable and isn’t in the long run going to lead to anything at all useful. We just have to be prepared for this more incremental approach. If lots of things are being published in philosophy that are not terribly interesting, that’s just the same as in any other science. I don’t think we should worry too much about this. Not everybody’s going to be a genius, but one advantage of logic is that most of what’s published in logic is, at least in a small way, a positive contribution. That may well not be true in other branches of philosophy where papers may be published

where it would just have been better if they'd never been published. But at least in logic, it seems to me that this kind of specialization is pretty much inevitable.

On the question of whether logic is useful in philosophy, of course, the sorts of things that are published in logic journals are mostly not useful in philosophy. The results in logic that are most useful in philosophy tend to be the very fundamental ones, which are discoveries made early on in the progress of a discipline when there's lots of low-hanging fruit that people are gathering. Once those questions have been answered, people are often forced to ask more and more difficult and specialized questions that don't contribute so much to philosophy. But if we're talking about research in logic, that's comparable to research in mathematics. The thing about mathematics is that pure mathematicians are just motivated, in most cases, by interest in these mathematical questions. They're not doing them for the sake of applications. But often it turns out later that the results do have applications. But it's better just to let mathematicians get on with their research, because that way deeper results will be discovered that then have all kinds of implications. When people were interested in the decision problem in the 1920s and 1930s, I don't think they thought that this was something that could revolutionize human society. They thought of it as a technical mathematical problem. But the work that Alan Turing and others did turned out to be crucial in the development of computers. It played a big role in the Second World War. Our societies are completely unrecognizable as a result of the development of the computer, which is a byproduct of something that was being pursued just for mathematical interest. I think that, in principle, other developments in philosophical logic can have a similar effect. For example, I don't think Hintikka, when he developed epistemic logic, was thinking about applications in theoretical economics or computer science. But nevertheless, it turned out that epistemic logic did have such applications.

Q25: As a leading figure in the field of philosophical logic, could you recommend one of your own papers in philosophical logic that you are most satisfied with? We are especially interested in which paper best embodies the research taste you advocate for, that is, how to use logical tools to intervene in a core philosophical debate in the most profound and illuminating way.

Williamson: I don't know which of my papers is really best for this, but just to give one example, I have a paper called "Very Improbable Knowing". It's a paper in epistemic logic. The point of it is to show the possibility that you can know something, even though it's almost certain on your own evidence that you don't know it. It's not just a failure of the KK principle, it's showing that the KK principle can fail in very extreme ways. It's not just that you could know without knowing that you know, but that you could know, even though given what you know, it's almost certain that you

don't know. I constructed some models of quite a simple kind, just to show that that's possible. I showed how these seem like reasonable idealizations of quite ordinary situations. There's nothing in it that is of great technical difficulty or anything like that. But it is concerned with issues about the interaction between different levels of knowledge, what one knows and what one knows about what one knows. It has been quite influential. I introduced a class of models that often get discussed in the epistemological literature. It's more of an advance in epistemology than it is in logic, but it certainly uses logical tools.

Q26: Throughout your academic career, a remarkable characteristic has been your astonishing originality and systematic rigor. Not only are you exceptionally prolific, but what is more impressive is that these deeply influential intellectual contributions have almost all stemmed from your independent thought and writing. We are very curious, how do you sustain such discipline and vitality in intellectual creation over the long term? In the actual practice of philosophical research, how do you manage the dynamic relationship between input (reading), incubation (thinking), and output (writing), to ensure they effectively foster one another?

Williamson: I don't have any great secret. I think you have to have a certain confidence in yourself to be willing to invest in your own ideas, to invest time and energy in them. You also have to be willing to listen to criticism because if people are just very self-confident and then dismiss every criticism, they can go off in some stupid direction and just write lots of self-indulgent things. One thing that has been helpful to me is just being involved in a lot of philosophical discussion and going to conferences or discussion groups where people are presenting papers, and being involved in the discussion and not just being interested in one's own paper, but being interested in other people's papers as well. Quite often, it turns out to be a waste of time, but sometimes you hear something which makes you think about things that you hadn't thought of before.

I find sometimes ideas come to me in discussion. There was a guy called Graham Wallas who said "I don't know what I think until I hear what I say". That sometimes applies to me. Sometimes I've just been involved in philosophical discussion about somebody else's ideas. I find myself suggesting something that was relevant to the dialectic, but there's a little voice in my head saying, that's not a bad idea, maybe that will lead to something. I can also be motivated by irritation when there's some idea that I don't like and I want to find a good argument against it. I think about how I might be able to find a weak spot in some position.

But another thing is just that I really, really enjoy doing philosophy and logic. I get a lot of pleasure from them and I enjoy intellectual elegance, whoever is producing it. This is partly the to and fro of discussion, but also some aesthetic reward.

One of the things that has guided me, I think more than many philosophers, is that I really want the ideas that I'm putting forward to have a certain elegance and maybe symmetry. When I see that that is working out, that's a very strong motivation for me.

Q27: Recent studies suggest that large language models can now generate philosophical arguments with such sophistication that non-experts find it difficult to distinguish them from those written by human philosophers. This phenomenon seems to challenge our understanding of mental states like “grasping a concept” or “possessing knowledge”. Your “Knowledge First” framework posits that knowledge is a fundamental, unanalyzable mental state. From this standpoint, do you believe an AI, which operates by recognizing data patterns and making statistical predictions, can, in principle, genuinely “know” philosophical truths? Or is it doomed to be merely a sophisticated simulator of philosophical discourse? Furthermore, what core epistemic virtues or intellectual qualities essential to philosophical inquiry do you see as fundamentally inimitable by AI?

Williamson: I don't want to say anything like “There are certain things which AI can never do or that a machine can never do”. It's difficult to attribute mental states to current AI. But that's for what I think is an incidental reason, which is that mental states involve some thought typically about the external world. That reference to the external world is achieved by interaction with the external world. If you have an AI system which is only communicating with the external world via a whole lot of linguistic input, but which isn't perceiving objects or moving about the world and interacting with objects, it's hard to see that kind of AI as having mental states. It's too mediated by human operators. But that is not such a difficult obstacle because once we start talking about AI that's put into a robot, which is moving about the world and interacting with objects and does have sensors, then such a robot is in much clearer cognitive relations with objects around it. I would be very, very cautious about saying that there are certain things that a robot could never understand or anything like that. If you look at the physical basis of human intelligence, if you just looked at brain scans and things like that, or had some model of the neurons in the brain, you might think, how could anything like this be intelligent? But the fact is that that is a physical basis for human intelligence. That we are made of flesh and blood and not of circuits and that we're organic rather than inorganic, I don't see as a big issue.

If we're talking about current AI, it can do quite a good imitation of a moderately competent professional philosopher, but only moderately competent. Generally, it seems that when students try to get AI to write their essays for them, it's very easy to tell that it's been written by AI and not by a human. But AI has had such a short period in which to develop in anything like its current form that things are going to be

very different in five or ten years time. In the case of logic, a lot of the developments where computers can outdo humans are not to do with large language models, but simply to do with much more traditional programming that takes advantage of the massively greater computational power that computers have. It's already the case that a lot of proof checking and filling in gaps in arguments in mathematics can be done by a computer. I wouldn't be confident that AI could never overtake human beings in philosophy. We human beings are not very good at philosophy. It's too difficult for us. We shouldn't expect to be the best in the world at it, let alone the best in the universe. It's perfectly possible that there are lots of species on other planets that are much better at philosophy than human beings. If we get overtaken by AI, then I think we should just learn a bit of humility.

Q28: There are basically two approaches to AI. The rule-based AI uses logic, and the data-driven AI uses machine learning. Which one do you think will succeed in the end?

Williamson: My guess is that success will be through a combination of those two things because when we're talking about logical proof, we need programmes that are 100% correct, that don't have any glitches in them. It is possible to produce such programmes. They can check proofs and to some extent, it's very limited so far, but they can produce new proofs. The data-driven style is much more suitable for building an AI that can have a conversation with a human being and possibly pass the Turing test.

There's a contrast that's a little bit like that within human beings because psychologists have this distinction between system one and system two, where system one is an unreflective, very fast, heuristic-driven way of thinking that we use a lot of the time. System two is the slower but more reflective kind of calculation that we can do. Human beings use a combination of system one and system two. System one is absolutely fundamental, but we can improve our performance by bringing in our capacity for system two thinking. For example, when we're just thinking in a very casual way about probability, we're thinking in a system one way, but when we realize that we have to do a mathematical calculation, if we're okay at calculation, we can then do much better by doing the calculation reflectively. It might be that AI which has the large language element which is more like system one, but also has various modules for doing calculations, doing proofs, which is more like system two, with that combination would be better than either one by itself. If we have AI that can give an impression of understanding philosophical issues, but then can also produce actual proofs and maybe mathematical models, then that might already be operating at the level of a really good philosopher. We're not there now, but I don't think that that has to be so far away. Human beings might, in the end, have to have the humility to learn from AI.