Contributions to Symposium on *Overfitting and Heuristics in Philosophy*

Abstract

This is a contribution to a symposium on the book *Overfitting and Heuristics in Philosophy*. It consists of three parts. The first is a précis of the book. The second part is a response to Kit Fine's comments. It is divided into two halves. One half presses the book's claim that the treatment of negation and falsity in Fine's truthmaker semantics exhibits overfitting, by permitting too much independence between a sentence and its negation. The other half presses the book's critique of Russellian structured propositions against Fine's defence, which distinguishes between basic and non-basic operations on propositions; the distinction is argued to be arbitrary. The third part of the contribution is a response to Harvey Lederman's comments, which concern its treatment of propositional attitude ascriptions. It explains why rational inference cannot be understood purely at the level of content; representational form plays an ineliminable role, which is why attitudes to contents under guises have to be invoked. The guises do not figure in the underlying semantics or the model; they are more like correction terms. Ascriptions of attitudes are vague with respect to ascriptions of attitudes under guises. The simplicity of the book's intensional semantics is not bought at the expense of complexity in its psychology, for the heuristics it invokes are needed anyway, since the epistemology of attitude ascription is just that of offline mindreading. Goodman and Lederman's contextualist semantics for attitude ascriptions is resisted on the grounds that it over-generates readings, more evidence of overfitting.

Key words: overfitting, heuristics, truthmakers, hyperintensionalism, guises, contextualism

Précis of *Overfitting and Heuristics in Philosophy*

The book originates from the 2022 Rutgers Lectures in Philosophy. It is an interconnected study of two methodological features of contemporary philosophy and their implications for some first-order debates. These features have been generally ignored, to the detriment of philosophical practice. The book aims to raise awareness of the methodological issues and to intervene thereby in the first-order debates.

The first methodological feature is reliance on heuristics in making judgments on actual or hypothetical cases. A heuristic is a quick and easy way of answering questions of some kind. Psychologists are well aware of extensive reliance on heuristics in human cognition. These heuristics are typically reliable enough to be useful, but not perfectly reliable. Heuristics can be unconscious, for instance in vision. Philosophical paradoxes can be explained as exploiting limitations of humanly universal unconscious heuristics to generate contradictions; a heuristic that mostly gives correct answers may also generate occasional inconsistencies. The resulting unreflective judgments often feel primitively compelling to a subject unaware of their origin.

Several such heuristics are discussed in detail. (1) The *persistence heuristic* is to ignore small differences (including differences in time and between possibilities). It is scarcely avoidable in maintaining a database, since not all information can be continuously actively updated. The persistence heuristic generates sorites paradoxes. (2) We use the *suppositional heuristic* to assess a conditional by assessing its consequent on the supposition of its antecedent. It generates the paradoxes of material implication, the illusion that the probability of a conditional is the conditional probability of its consequent on its antecedent, and some outright contradictions. (3) We use a family of *disquotational heuristics* for indirect speech reports, belief ascriptions, and ascriptions of truth-values. They are complicit in both semantic paradoxes and Kripke's puzzle about belief. (4) We use the *weighing heuristic* to combine reasons on the metaphor of a scale, which enforces additivity and thereby generates some incorrect results, including in cases of intersectionality.

Philosophers' unconscious reliance on imperfectly reliable heuristics does not warrant full-blown scepticism about philosophy. Perception also involves unconscious reliance on imperfectly reliable heuristics, but that does not warrant full-blown scepticism about perception; we have lots of perceptual knowledge. Instead, philosophers need to become more sophisticated in their handling of data, as natural scientists are, so erroneous data can be weeded out rather than treated as counterexamples to otherwise promising theories.

The second main methodological feature of contemporary philosophy discussed in the book is *overfitting*. This is a pathology familiar through bitter experience to natural and social scientists, especially those concerned with interpreting large data sets. Overfitting typically consists in complicating a model to fit the data more closely, often by adding new parameters to the model. This may sound laudable, but such models tend to do poorly in predicting future data, provoking another iteration of the process, and so on, without convergence. The underlying problem is that the data set contains a few errors or outliers, so fitting the model tightly to the data involves fitting it to erroneous or misleading data points, thereby sending it

off on the wrong track. Scientists have found that a better approach is to keep the model comparatively simple and settle for a looser fit to the data. In the long run, that tends to yield closer predictions of new data and the identification of deeper patterns in the data.

The heuristics on which philosophers rely for data-gathering will generate some erroneous data points, and thereby facilitate overfitting. Unfortunately, few philosophers have been taught about the danger of overfitting. The disciplinary culture encourages the process of complicating the theory in response to apparent counterexamples. In effect, the *de facto* professional incentives have rewarded overfitting. This is manifest in research programmes of 'conceptual analysis', with their non-convergent sequences of proposed analysis, counterexample, more complicated analysis, and so on. In philosophical logic, non-classical logics are typically motivated by philosophical paradoxes, such as semantic paradoxes and sorites paradoxes, generated by heuristics whose outputs are typically treated as given data. The underlying heuristics are not identified, still less questioned. Instead, classical logic is replaced by something more complicated, to fit the data. Similar processes occur in formal semantics, for instance in the semantics of conditionals. Complicating a theory is treated as a cost-free move. More emphasis on the theoretical virtue of simplicity would make philosophy less vulnerable to overfitting.

Chapter three discusses hyperintensionalism in metaphysics as a case study in overfitting. On a popular narrative, just as the 'possible worlds revolution' of the 1960s made philosophical progress by advancing from an extensional framework to a more fine-grained intensional one, so the more recent 'hyperintensional revolution' made further progress in the same direction by advancing from the intensional framework to a still more fine-grained hyperintensionalist one. Crudely, for extensionalists, the content of a predicate is just its extension. For intensionalists, the content of a predicate is its intension, the function from possible worlds to its extensions at them. Sameness of intension entails sameness of extension but not conversely, since predicates can coincide in extension by accident. For hyperintensionalists, how is the content of a predicate individuated? There is no agreed answer. As a placeholder, we can call the content a hyperintension. Sameness of hyperintension entails sameness of intension but not conversely.

Arguably, the move from intensionalism to hyperintensionalism is much less like that from extensionalism to intensionalism than hyperintensionalists suggest. The first move was motivated by the discovery of a mathematically powerful systematic new explanatory framework, possible worlds semantics, which immediately found applications in logic, linguistics, metaphysics, and soon in computer science and theoretical economics. By contrast, the second move was not motivated by the discovery of any new explanatory framework, but instead mainly by apparent counterexamples to intensionalism. Applications outside metaphysics have been sparser, and no systematic new explanatory framework has established itself, though there have been many attempts to construct one. While the intensional revolution transformed the intellectual landscape in less than one decade, the putative hyperintensional analogue has been under way for more than three decades, still seems far short of maturity, and has been much less widely accepted.

Three forms of hyperintensionalism are explored in detail: impossible worlds semantics (which violates semantic compositionality), truthmaker semantics (which faces some of the same difficulties as Bertrand Russell's logical atomism), and Russellian

structured propositions (whose theory is inconsistent unless complicated restrictions are imposed). All these approaches exhibit signs of overfitting. Various apparent cases of hyperintensionality can be traced to a heuristic that assesses the truth-value of 'A because B' by the explanatory value of 'B' as an answer to the question 'Why A?' In judging explanatory value, we are influenced by superficial features such as clarity of presentation and absence of irrelevant terms, apparent hyperintensionality is inevitable. The features to which hyperintensional theorizing responds are on the linguistic surface, not below in the metaphysical depths.

The final two chapters discuss representational hyperintensionality. In a trivial sense, quotational contexts are hyperintensional: the names 'Cicero' and 'Tully' have the same intension, but replacing the former by the latter in the sentence 'The name "Cicero" has three syllables' switches it from true to false. The question is whether replacing one expression by another with the same intension in a *non*-metalinguistic attitude ascription can really switch its truth-value, as it appears to in Frege's puzzle of Hesperus and Phosphorus. The initial case for a positive answer seems overwhelming. Yet attempts to construct a compositional semantics for attitude ascriptions on that basis have run into trouble. Even substituting synonymous terms can switch an ascription's perceived truth-value; for instance, we may judge 'John is certain that furze is gorse' false but 'John is certain that furze is furze' true. Contextualist semantic theories introduce complicated extra apparatus and still fail to match unreflective judgments of truth-value in some cases; they smack of overfitting.

In his article 'A Puzzle about Belief', Saul Kripke describes variant Frege puzzles for which our normal practice of belief ascription breaks down. Rather than revealing some incoherence in our *concept* of belief, such cases are best interpreted as showing that our normal disquotational *heuristic* for ascribing beliefs to people on the basis of what they would say sometimes has inconsistent outputs (like many other heuristics). Of course, Frege puzzles arise for most propositional attitudes, not just for belief. The other attitudes, including knowledge, may be ascribed on the basis of verbal behaviour too, though the connection may be more indirect than it is between believing that P and saying that P. Further heuristics are needed for that, and for ascribing attitudes on the basis of non-verbal behaviour. These are all examples of what psychologists call *mindreading*. The fallibility of our capacity for mindreading others is a matter of common experience, so its dependence on fallible heuristics is hardly surprising. The key to understanding hypothetical Frege puzzles is to realize that they too involve applications of our mindreading capacity, but offline in imagination rather than online in social interaction. To try to understand Frege puzzles through the semantics of attitude ascriptions rather than the psychology of mindreading is to look in the wrong place.

By letting the psychology do the heavy lifting, we take much of the pressure off the semantics. This eases the way to a simple non-contextualist quasi-homophonic intensionalist semantics of attitude ascriptions, on which propositional attitudes are simply relations (at a time) between an agent and an intension. The coarse-grained nature of the contents (objects) of the attitudes will indeed implicate some of our attitude ascriptions in significant errors, which motivates many philosophers to seek a hyperintensionalist theory of content. However, such theories face severe problems of their own, as already noted. Arguably, the best strategy is just to bite the bullet and accept intensional contents, which are simpler, far better understood, and far less liable to generate contradictions such as the Russell-Myhill paradox.

Unlike hyperintensional frameworks, an intensional framework is structurally commensurable with the formal frameworks of both Bayesian probability theory and epistemic and doxastic logic. Thus, intensionalism is naturally better adapted than hyperintensionalism to treating evidential relations. Of course, evidential and subjective probabilities seem as sensitive as ungraded knowledge and belief to modes of presentation, and this apparent sensitivity has been shown to make trouble for current practice in fields such as welfare economics. However, there are unexpected structural obstacles to building such sensitivity into the probabilistic framework. The risk is that doing so will destroy its mathematical and explanatory power. We do better to retain the intensional-probabilistic framework at the level of represented content, while accepting that some cognitive phenomena must be understood at the level of the linguistic or non-linguistic representations themselves.

The final chapter applies the intensional framework to issues in metametaphysics. It explains how the framework seems to warrant the view that claims in metaphysics and mathematics do not really say anything, because non-contingent propositions are either tautologies or contradictions, but this is just one manifestation of a much more general phenomenon, the coarse-grained individuation of content, which has nothing special to do with disciplines that study non-contingent matters. It should be handled by whatever measures are needed to handle the more general phenomenon of cognitive differences between diverse representations of the same intension; once that is done, metaphysics and mathematics no longer look like outliers, and treating them as exceptions is unwarranted. Nor is the original problem exclusive to intensionalism; it has analogues for more fine-grained theories of content, such as theories of Russellian structured propositions. What we need is a clean cut between content and cognitive significance.

Response to Kit Fine's comments

Kit Fine's comments on *Overfitting and Heuristics in Philosophy* (Williamson 2024a, Fine 2026) divide into two halves, concerning, respectively, truthmaker semantics and Russellian structured propositions. I will divide my response accordingly. Both halves concern aspects of the hyperintensional programme in metaphysics and semantics, with which Fine is associated: on his view, some necessarily equivalent sentences differ both in their truthmakers and in the propositions they express.

Negation and falsity in truthmaker semantics

Section 3.4 of the book discusses the treatment of negation and falsity in Fine's truthmaker semantics. The 'bilateral' version of the semantics treats truth and falsity as mutually independent properties: each sentence is assigned both a set of truthmakers and a set of falsitymakers, and the recursive semantic clauses for the connectives are formulated in those terms. In particular, this facilitates a smooth treatment of negation. However, as I point out in the book, it also incurs a risk of overfitting, through adding another parameter to the semantics: one more degree of freedom (Williamson 2024a: 134). By contrast, in classical bivalent semantics, falsity is simply equivalent to non-truth for sentences of the object-language.

More than a century ago, Bertrand Russell faced a similar problem in developing logical atomism (Russell 1918/1919, 1956: 211-16). Instead of 'truthmakers', he talked of 'facts'. He was reluctant to postulate 'negative facts', the analogue of 'falsitymakers'. As an alternative, he considered postulating instead a relation of 'incompatibility' between 'positive facts', which would correspond to Fine's primitive relation of 'exclusion'. In 'unilateral' truthmaker semantics, a truthmaker for the negation of a sentence is a state or fact that excludes or is incompatible with all its truthmakers. In his present response, Fine also contemplates a hybrid bilateral semantics on which the falsitymakers for an *atomic* sentence are defined in terms of exclusion and its truthmakers, after which the usual bilateral clauses take over for complex sentences; this avoids awkward cases where a sentence and its double negation differ in their truthmakers. Although Fine's discussion is more systematic and rigorous than Russell's, they are struggling with the same underlying problems.

As emphasized in the book, using exclusion or incompatibility does not automatically avoid the risk of overfitting, since such a relation is itself a further component of the semantics, though not one that must be specified anew for each atomic sentence. In the terminology of modal logic, exclusion is an independent constituent of a *frame*, not one which must be added to a frame to get a *model* (in the logical sense). Fine suggests that the exclusion relation should count as a *constant* ('like the speed of light') rather than a *parameter*, because it is fixed metaphysically rather than semantically, so adding it should not increase degrees of freedom in the same way as adding falsitymakers.

Of course, as Fine recognizes, what is fixed metaphysically may not be fixed *epistemically*: stone age people did not know the speed of light. Varying hypotheses as to the value of a constant can still give a theory's proponents extra wiggle room. For a toy example, suppose that the structure of spacetime is metaphysically fixed but unknown, because many of its dimensions may be curled up very tight and so almost, but not quite, negligible (as in some versions of string theory). Then overfitting might take the form of postulating ever more curled up dimensions of spacetime each with its own structural properties, in order to explain awkward data. That could be just as methodologically vicious as more familiar cases of overfitting.

Questions of overfitting must be handled with care. For instance, to attack a theoretical framework of which one's own is effectively a special case might seem self-defeating. But imagine another toy example: a vitalist framework for biology in which every living thing at every time has a quantifiable *life force*, irreducible to anything else. A more standard non-vitalist framework for biology might be equivalent to the special case of the vitalist framework when all life forces are set to zero. That should not prevent non-vitalist biologists from criticizing the vitalist framework for its susceptibility to overfitting. Moving from a framework to an equivalent of one of its special cases often reduces degrees of freedom. Similarly, Fine suggests, a standard intensionalist framework such as I defend is equivalent to the special case of his truthmaker framework where states and exclusion are identified with worlds and distinctness respectively. Perhaps so; nevertheless, without self-defeat, intensionalists can still criticize the truthmaker framework for overfitting.

In comparing truthmaker semantics with intensionalist possible worlds semantics, one must also keep in mind Fine's insistence, in both his reply and his earlier papers, that the former's framework of states is *non-modal*, whereas the latter's framework of worlds is intended to be modal. Thus, possible worlds semantics framework should more properly be compared with whatever extension of truthmaker semantics is intended to handle the relevant modal constructions. That would bring out more clearly the difference in simplicity between the two approaches.

As Fine points out, bilateral truthmaker theorists are not alone in treating truth and falsity as independent of each other, give or take a few constraints: theorists who postulate truth-value gaps or gluts are in the same boat. That does not worry *me*, for I do *not* postulate such gaps or gluts. To postulate gaps or gluts *is* to take a step towards overfitting. It is not automatically a fatal step, for overfitting comes in degrees, and neither truthmaker semantics nor gappy or glutty semantics gets anywhere near the extreme—for instance, as noted in the book, they do not multiply degrees of freedom anything like as much as unconstrained impossible worlds semantics does. Still, the default is something simpler, like intensionalism, without gaps or gluts. That the putative explanatory successes of truthmaker semantics really justify overturning the default is far from clear.

Russellian structured propositions and basic operations

Section 3.5 of the book discusses accounts of propositions in a tradition inspired by Russell, on which reference approximates an isomorphism between, on one side, the sentence, its

syntactic constituents, and their syntactic relations and, on the other side, the proposition expressed, its structural constituents, and their structural relations—syntax is projected onto the world. Thus, sentence operators, which build complex sentences out of simpler sentences, correspond to proposition-building operations, which build complex propositions out of simpler propositions. On this fine-grained hyperintensional conception of propositions, two sentences express the same proposition if and only if they consist of coreferential atomic expression put together in exactly the same way. In the book, I show how this conception is inconsistent with natural assumptions about how proposition-building operations work. For instance, on a pure fine-grained conception, applying a proposition-building operation O to a proposition p should give a proposition of the form O(p), which determines O and p uniquely in the sense that O(p) = O*(p*) only if O = O* and p = p* (see, for instance, Menzel 2024, on which Williamson 2024b comments). We also expect two proposition-building operations O_1 and O_2 to determine a composite proposition-building operation $O_{1,2}$ such that applying $O_{1,2}$ to a proposition is equivalent to applying O_1 to it and then applying O_2 to the result. Thus, $O_{1,2}(p) = O_2(O_1(p))$, which by the fine-graining principle implies that $O_{1,2} = O_2$ and $p = O_1(p)$. But that is hopeless. Since O_1 is an arbitrary proposition-building operation and p an arbitrary proposition, it implies that no proposition-building operation ever yields a new proposition. By contrast, no such problem arises on a coarse-grained intensional conception of propositions.

Fine's response to the problem is to distinguish *basic* from *non-basic* proposition-building operations. Composing proposition-building operations, basic or non-basic, *never* yields a basic proposition-building operation. The fine-grained principle that O(p) = O*(p*) only if O = O* and p = p* must be restricted to the case where O and O* are both *basic* proposition-building operations. As Fine recognizes, this leaves the hyperintensionalist with an awkward question: *which* proposition-building operations are basic? How are we supposed to go about answering that question?

Take conjunction and disjunction. Each is normally assumed to be definable in terms of the other and negation, by De Morgan's laws. They are both amongst the best candidates to be basic proposition-building operations. We may reasonably assume that at least one of them is basic. If only one of them is basic, which one is it? They seem equally joint-carving. The extensive formal duality between them suggests that either choice would be arbitrary: neither is more basic than the other. Suppose, then, that they are equally basic. If so, the basic operation of conjunction must be distinguished from the necessarily equivalent non-basic conjunction-like operation composed in the usual De Morgan way out of negation and disjunction; similarly, the basic operation of disjunction must be distinguished from the necessarily equivalent non-basic disjunction-like operation composed in the usual De Morgan way out of negation and conjunction. Operations are proliferating alarmingly.

How bad is such proliferation? It is another manifestation of a tendency noted in the book, for hyperintensionalist arguments to over-generate and pull propositions ever closer to sentences, thereby undermining the reasons for postulating propositions distinct from sentences in the first place. Even slight structural differences between languages—for instance, in whether they have definite and indefinite articles (like English and unlike Slavic languages)—will make a difference to what propositions they express, if the syntax of sentences is reflected clearly enough in the structure of the propositions they express. In the

case of definitions, when the definiens differs syntactically from the definiendum, substituting one for the other will change the proposition expressed. This strengthens the suspicion that the structure of language is being illicitly projected onto the structure of the world.

Imagine constructing a formal language. We must decide what the primitive operators are to be. The usual attitude is that it does not matter much which we choose, as long as we give ourselves enough expressive power: for instance, if we take negation and conjunction as undefined, we still have the power to express every proposition, property, and relation we could have expressed had we started with negation and disjunction instead. But if disjunction is a basic operation, that is not so. A sentence of natural language of the form 'A or B', where 'or' expresses the basic operation of disjunction, may express a proposition expressed by no sentence of our formal language, with no symbol for the basic operation of disjunction. On this view, whether we choose negation and conjunction, or negation and disjunction, or negation and both conjunction and disjunction, makes a difference to the expressive power of the language, to what propositions, propositions, and relations it can express. If we are in the dark as to which the basic operations are, we are ill-placed to judge the metaphysical implications of our semantic stipulations.

At the end of his comments, Fine makes some tentative suggestions as to which operations are basic. As far as I can tell, he is being guided by felt psychological naturalness, but why should we expect that to be a good guide to metaphysical basicness? We might use it in selecting our initial hypotheses, but how are we then to *test* those hypotheses? What mechanisms are in place for recognizing cases of psychological naturalness without metaphysical basicness, or metaphysical basicness without psychological naturalness, if there are indeed such cases? One danger of an overfitting methodology is that it enables us to remain in denial of such cases, by allowing us cost-free to treat the contents of psychologically natural judgments of metaphysical basicness or non-basicness as simply data.

Despite the Russell-Myhill paradox and the problems raised in section 3.5 of the book, the general fine-grained approach to propositions will never succumb to inconsistency, for its proponents can model propositions and their structure ever more closely on sentences and their syntax; a true theory of syntax is consistent. The distinction between basic and non-basic operations is a case in point: it is modelled on the distinction between words, such as 'and' and 'or', and syntactically complex expressions, such as 'not-both' and 'neither-nor'. But the further one goes in that direction, the more hyperintensionalist propositions look just like glorified sentences. The effect is to void the theory of propositions of its explanatory power.

Response to Harvey Lederman's comments

Harvey Lederman's comments on *Overfitting and Heuristics in Philosophy* focus on its treatment of intentional attitudes to coarse-grained contents, our ascriptions of such attitudes to others, and the use of guises to make fine-grained cognitive distinctions (Lederman 2026, Williamson 2024a). Having raised three questions for my approach, Lederman compares it unfavourably to the contextualist treatment of attitude ascriptions he has developed with Jeremy Goodman (Goodman and Lederman 2021). My response to his comments is divided accordingly.

Lederman's first question: what are guises?

Philosophers have tended to locate rational and cognitive relations at the level of *content*. For example, they treat the array of words, symbols, and sometimes diagrams in a mathematical proof as just a means by which mathematicians grasp the *real* proof, made up of the contents expressed through the medium of those representations (Boghossian 2024). Lederman charges that I 'cannot provide a content-based explanation of the felt rational pressure' he describes in cases of deductively valid reasoning.

Compare the following two arguments (where the demonstrative 'that' is constant in reference throughout):

Argument A	Premise 1	If that is gorse, that is prickly.
	Premise 2	That is gorse.
	Conclusion	That is prickly.
Argument A*	Premise 1	If that is gorse, that is prickly.
	Premise 2*	That is furze.
	Conclusion	That is prickly.

Argument A is clearly a valid instance of modus ponens. By contrast, argument A* looks like a *non sequitur*, so not an instance of modus ponens, which is a valid argument form. If so, argument A and argument A* differ in the rational relations between their premises and their conclusion. But 'furze' and 'gorse' are *synonyms*—not just on my semantics, but by normal linguistic standards. Thus, premise 2 and premise 2* have the same content, as Lederman will accept. Since premise 1 and the conclusion are word-for-word the same in arguments A and A*, and there is no relevant difference in context, A and A* are the very same argument at the level of content. At that level, the impression that A and A* differ in validity is an illusion.

A reflex response on behalf of the content-based view is that the cognitive difference between the two verbal arguments arises only for someone incompetent with 'furze', or 'gorse', or both. But that is a mistake. As the book explains (pp. 166-77), a rational speaker

can understand both ‘furze’ and ‘gorse’ by normal linguistic standards yet still doubt that they co-refer. Thus, arguments A and A* differ in how much rational pressure they exert even on speakers who understand their premises and conclusion. At the level relevant to such rational relations, the arguments are distinct, so the relevant level is that of linguistic form, not that of content.

A clue in plain sight that normal mathematical proofs cannot be understood at the level of pure content is the central role in them of *open formulas*, with free variables. An equation like $z^2 = x^2 + y^2$ does not abbreviate its own (false) universal generalization; it is asserted in the proof only on prior assumptions involving those variables, though they are no part of its meaning. The open formula does not express a proposition, for the values of the variables are deliberately left unfixed. The introduction and elimination rules for the quantifiers in a standard system of natural deduction are stated with restrictions on the occurrences of variables: they essentially involve linguistic form, not just pure content. This applies not only to fully formalized proofs (which are rare in mathematics), but to informal proofs of the kind mathematicians find most perspicuous (Williamson 2024c).

Of course, an open formula may express a proposition under an assignment of values to variables, but a function from assignments to propositions is not itself a proposition, and does not abstract from syntax in the envisaged way, since the assignments of values are to syntactic entities, variables.

In short, our clearest paradigms of rational reasoning make no sense if we lose track of the identity of the syntactic symbols in play. The ‘real proof’ made of pure content is a myth, based on a misunderstanding of what rigour involves.

None of this justifies lurching to the opposite extreme, formalism, and treating mathematical proofs as mere sequences of moves in a formal game. Mathematical proofs are normally written in a hybrid of a natural language with formal mathematical notation; such a hybrid is a more or less interpreted language, well understood by mathematicians. The semantics is not redundant, for it explains why the proof-theoretic rules are truth-preserving in the relevant sense (itself geared to assignments). This structural correspondence between compositional syntax and compositional semantics is the basis of standard soundness theorems in metalogic, such as for first-order logic.

For a proper understanding of proof, in its informal knowledge-directed sense, we must track how its rules co-ordinate syntax and semantics; to marginalize either side at the expense of the other is to miss the point of the enterprise. In stating a mathematical theorem, we use a sentence to express a proposition; we need the sentence just as much as we need the proposition, though a slightly different sentence might do equally well. We can put this relationship by saying that we prove the proposition *under the guise* of the sentence. Such relationships are by no means confined to mathematics. They occur in reasoning quite generally. In the gorse example, the concluding proposition follows from the premise propositions under the respective guises of the sentences in argument A but *not* under the respective guises of the sentences in argument A*. Similar examples can easily be constructed for inductive reasoning about furze/gorse. Far from undermining the idea of rational pressure, sentential guises of propositions are needed to understand it properly.

A familiar point, explained in the book and emphasized by Lederman, is that sentential guises are not a one-size-fits-all solution to all problems in the vicinity. This can be

illustrated with the demonstrative 'that', when its co-reference over distinct occurrences is no longer guaranteed by stipulation, as it was for arguments A and A*. Thus, in a given extra-linguistic context, we may use the demonstrative phrase 'that bird' intending to refer to the bird we see in the tree, or the bird we hear singing, or the bird we remember from yesterday, whether they are in fact one bird, two, or three. To distinguish good from bad reasoning, the sameness of the linguistic type 'that bird' no longer suffices; the utterance must be connected to the appropriate act of attention, whether in vision, hearing, or memory. Even for a single sense modality or cognitive faculty, there may easily be, or appear to be, several candidates for attention.

Nor is the problem restricted to indexicals. In Kripke's example, Peter assumes that Paderewski the statesman and Paderewski the pianist are different people, and uses the name accordingly. Knowing that it is the same person, we may still try to track what is going on in Peter's mind by pretending to disambiguate the unambiguous name 'Paderewski' into 'Paderewski$_1$' and 'Paderewski$_2$' when we ascribe mental states to Peter. By normal standards, it is still the same name all along, so we are in effect individuating guises more finely than linguistic types.

Some guises have no linguistic element at all. One non-human animal may know something about another through visual attention, and something else about it through auditory attention, and as a result fail to integrate the two pieces of knowledge properly. What linguistic and perceptual guises have in common is that they concern the form of representations, not their content, but nevertheless help explain what further contents will, or will not, be inferred. The level of content is not fully autonomous.

Lederman finds my treatment of guises in the book frustratingly uninformative. Having raised various questions about them, he clarifies:

> These questions are not primarily about the *nature* of guises; I am not asking for an *analysis* of them. The question is instead about structural constraints on the notion of guises, which might help to show how this notion operates within the theory, and thus give us a better purchase on it.

For linguistic guises, such as names and sentences, we already have rich structural accounts of their co-ordinated syntactic and semantic structure and its implications for deductive reasoning: much of metalogic is about that. Perceptual representations seem to be structured very differently: trying to give a unified structural theory of both linguistic and perceptual representations may not be a very fruitful enterprise. Still, our own cognitive systems face a similar task, since our linguistic and perceptual representations interact.

An example: you look at a tree, and on that basis assert 'This tree is dead'. Somehow, despite the radical difference in format, the evidence encoded in your visual state engaged with the proposition contextually encoded in the sentence you utter; in the good case, your assertion is knowledgeable. Here, the guise of the content of your evidence is your visual state, and the guise of the content of your assertion is the sentence in your context.

If we theorize the *content* of your evidence as perceptually structured, and the *content* of your assertion as syntactically structured, we assign them incommensurable formats, and stymie our chances of understanding the evidential relation between them. We need a more abstract theoretical framework, one neutral between perceptual and syntactic formats.

Luckily, such a framework is available, in the state spaces used by the mathematical theory of probability. Each space has an underlying set Ω of possible *outcomes*, which are mutually exclusive, jointly exhaustive, and maximally specific in relevant ways. Any set of outcomes, a subset of Ω, is an *event*, which obtains if and only if one of its member outcomes does. Probabilities are assigned to events, constrained by satisfy the usual Kolmogorov axioms. Such a state space is very like an intensional Kripke frame for modal or epistemic logic, despite differences in terminology. A Kripke frame has an underlying set W (in place of 'Ω') of possible *worlds* (in place of 'outcomes'), which are mutually exclusive, jointly exhaustive, and maximally specific in relevant ways. Any set of worlds, a subset of W, is a *proposition* (in place of 'event'), which obtains if and only if one of its member worlds does. What we have been calling 'contents' can be identified with events or propositions in this sense. In the example above, the content of your visual state is the set of worlds (outcomes) for which the state is veridical, and the content of your assertion is the set of worlds (outcomes) for which the assertion is true. These two contents (propositions, events) stand in logical or probabilistic relations. If all the worlds (outcomes) in the content of the visual state are also in the content of the assertion, the former content entails the latter content. If a weighted majority of the worlds (outcomes) in the content of the visual state are also in the content of the assertion, the former content makes the latter content probable—the probability of the latter conditional on the former is high. More generally, integrating the contents of cognitive states into probability spaces permits us to apply the mathematical power of probability theory to understanding their epistemic relations, while integrating the contents into Kripke frames permits us to apply the logical power of epistemic logic to understanding their logical relations. Since probability spaces and Kripke frames have the same underlying structure, we have an elegant theoretical solution. By treating contents in this perspicuous coarse-grained intensional framework, we abstract away from the differences in format of the guises under which they were presented, better to understand the probabilistic and broadly logical relations amongst them (Williamson 2026). For brevity, call this 'the state-space model of content'.

Like all good models, the state-space model of content has a price. It rides roughshod over the cognitive friction implicit in the variety of ways in which contents are represented. Such friction can often be ignored—Frege puzzles are comparatively rare in cognition, and not all reasoning is cross-modal. But sometimes the friction makes a significant difference to what we are interested in, and we must take it into account. Talk of guises serves that purpose. It is comparable to the use of *correction factors* or *correction terms* to adjust a model's predictions, a widespread and hardly dispensable practice in science. Its point is to handle not random or unexpected deviations from a model, but instead its systematic and expected errors from a known source, which cannot be smoothly integrated into the model itself. Thus, relativization to guises is a correction factor for the state-space model of content, introduced only when necessary. A general relativization of the model to guises would render it quite intractable—every proposition or event with at least one sentential guise has infinitely many—and destroy its mathematical power. As far as possible, we should work with the plain state-space model of content, as standard probability theory does, and relativize to guises when we must.

What exactly is being corrected? According to standard probability theory, no conjunction is more probable than one of its conjuncts. The suggestion is *not* that, in real life, there are exceptions to such a law—at least, not on the technically informed sense of 'probable'. Likewise, I am not suggesting that, in real life, there are exceptions to guise-free intensional semantics for the literal meaning of attitude ascriptions. Instead, the corrections are to the resultant predictions about the thought processes and behaviour of broadly rational agents, given their attitudes. They may not reason as we would expect, and we may need to track their guises to understand why not.

Lederman says that his question about guises 'would be answered, for instance, by having a model which illuminates their application to cases of interest'. But, as just explained, the primary role of guises within this approach is not as constituents of a model but as correction factors when we interpret agents using a guise-free model of content.

Still, in special cases, we *can* model guises themselves. For instance, first-order logic with a standard classical system of proofs by natural deduction, interpreted over a standard set-theoretic structure, constitutes a formal model of one kind of cognition, with expressions of the object-language as guises for their corresponding semantic values in that structure (to avoid ambiguity, I use 'structure' here where logicians would say 'model'). A formal model of perceptual content and perceptual guises might look very different.

Lederman's second question: what if any is the connection between having an attitude and having that attitude-under-a-guise?

Lederman asks whether someone who takes an attitude to a content under a guise thereby takes that attitude to that content (*simpliciter*). He suggests, with examples, that a positive answer 'would follow the standard logic of "under" in English'. His implicit general principle is that if *A X*s under *B* then *A X*s. However, that principle fails for dispositional predicates. That a machine is reliable under special conditions does not entail that it is reliable. A non-nervous person may be nervous under extreme pressure or under water. Knowing, believing, and many other intentional attitudes are themselves dispositional states. Believing a proposition under a guise is something like relating to it in a belief-like way under the condition that it is presented by that guise, for instance by using it as a premise in practical reasoning. Obviously, that does not imply relating to the proposition in a belief-like way under other conditions.

Lederman suggests the principle that 'a person *X*s that *p* if and only if they *X* that *p* under some guise', whether or not it follows from the logic of English. It promises a welcome simplification of the area. Unfortunately, it generates awkward complications of its own, especially for graded attitudes. For instance, many people have high confidence in the proposition that George Orwell wrote *1984* under the guise of the sentence 'George Orwell wrote *1984*' but low confidence in the same proposition under the guise of the sentence 'Eric Blair wrote *1984*'. Taking '*X*s' as 'has high confidence' and 'has low confidence' respectively, the suggested principle yields the results that those people simultaneously have both high confidence that George Orwell wrote *1984* and low confidence that George Orwell wrote *1984*, which undermines standard ways of reasoning about degrees of confidence.

More formally, on a given evidence base, the probability of that proposition under the guise of the sentence 'George Orwell wrote *1984*' may be over 90%, while its probability under the guise of the sentence 'Eric Blair wrote *1984*' may be under 10%. By the suggested principle, the probability of the proposition (*simpliciter*) is both over 90% and under 10%, which is mathematically impossible. Thus, the suggested principle cannot hold in full generality.

Treating graded attitudes as exceptional cases would only make for more complications. We had better hope that whatever works for graded attitudes will generalize to ungraded attitudes too.

The underlying problem is metasemantic. We are trying to describe the messy complexity of agents' cognitive lives in terms of a simple framework of attitudes to guise-free contents, on the state-space model. Our coarse-grained intentional distinctions must supervene somehow or other on a vast array of finer-grained lower-level distinctions, but we have only the faintest idea how. Contemporary philosophy of language is very far from having plausible predictive principles about how the supervenience goes. Some interpretive principle of charity seems to be at work. I have argued elsewhere that it is best understood as a principle of knowledge-maximization, closely related to a default principle in our metacognitive apparatus for mindreading, which treats knowledge rather than ignorance as the defeasible default (Williamson 2007, 2024d). Thus, there is metasemantic pressure for our ascriptions of intentional states to each other to come out as knowledgeable, and therefore true, subject to all the other metasemantic pressures. For that reason, I am not too pessimistic about the epistemic standing of our mindreading practices.

Metasemantic charity is unlikely to vindicate Lederman's suggested principle that 'a person *X*s that *p* if and only if they *X* that *p* under some guise'. For charity applies to the practice as a whole, including *denials* that someone has some attitude. Thus, if an agent *X*s that *p* under some little-used guise, and thereby counts as *X*ing that *p* by the suggested principle, but speakers deny that the agent *X*s that *p* since the agent fails to *X* that p under all its other, more commonly used guises, the suggested principle will count all those denials as false. A more nuanced picture is needed.

In effect, attitude ascriptions are *vague*. When I apply the words 'bald' and 'hairy' in particular cases, I do so as an ordinary speaker of English. I have no appetite for trying to leverage my theoretical views in the philosophy of language to narrow down the boundaries of 'bald' and 'hairy'. Similarly, when I ascribe intentional states in particular cases, I do so as an ordinary speaker of English. I have no appetite for trying to leverage my theoretical views in the philosophy of language to narrow down the boundaries of intentional-state-ascribing predicates. Both enterprises look unpromising.

Lederman's third question: how does semantics relate to psychological processing on my view?

Lederman raises the suspicion that I am buying simplicity in semantics at the cost of complexity in psychology, while ignoring the latter in the overall abductive comparison of my view with others. That would of course be cheating. Shifting complexity from one side of

the semantics-psychology boundary to the other is not in itself an improvement. At that broad methodological level, Lederman and I are in agreement.

One might therefore hope that the more complicated semantic theories that Lederman defends will pay their way by telling a simpler psychological story about attitude ascription. Instead, they tell *no* psychological story at all. By omission, they treat the semantics as more or less transparent to speakers. To that end, they fit the semantics as closely as they can to all the complications of the psychological data—what sentences are acceptable in what contexts, and so on. In its way, this transparency is a strikingly simple psychological hypothesis. Unsurprisingly, it is false.

Near the end of his paper, Lederman admits of his and Goodman's contextualist semantics for attitude ascriptions: '[W]e must attribute error to speakers in some cases. For instance we deny that there are relevantly informative true readings "Lois doesn't know that Superman is Clark Kent"'. These are not performance errors, which a speaker corrects once they are pointed out. These errors are much more systematic and robust than that. They call out for psychological explanation. Similarly, in 'A Puzzle About Belief', Kripke rightly emphasizes the psychological aspect of his puzzle, how pre-theoretically it leaves us not knowing what to say about the case, and strongly tempted to say inconsistent things about it (Kripke 1979). The semantic status of the relevant sentences is far from transparent to us. What explains our ignorance? Lederman does not say.

Bringing the semantics closer to the psychology does not automatically make the residual gaps of ignorance and error between them easier to explain: instead, it may raise new questions as to why such gaps occur in some places but not in others.

The psychological mechanisms I postulate in the book are anyway independently motivated. After all, accepting or rejecting ascriptions of mental states to others is just an application of our metacognitive capacity for *mindreading*, which is generally acknowledged to be a cognitively demanding task. In testing theories about the semantics of attitude ascriptions, we assess sample ascriptions of mental states in hypothetical contexts, but that still requires our capacity for mindreading, though applied offline, by contrast with everyday online mindreading when we interact with others in real time. That is not pure semantics. The relevant heuristics I postulate are the simplest plausible ways of attributing specific propositional attitudes to others given what they say, something most people do on a daily basis. These heuristics have not been made more complex to compensate for the simplicity of the semantics. As far as I can see, Lederman makes no attempt to show that the heuristics I postulate could be simplified or dispensed with altogether if only one complicated the semantics, in his way or any other. Thus, for his charge that I have merely moved the complexity elsewhere, there is as yet no case to answer.

Lederman has a further general worry: that I have separated semantics so much from psychology that semantics 'appears completely psychologically inert' and 'effectively idle'. Such remarks quite mistake my view. We need knowledge of our environment on which to act; as humans, we hold much of our knowledge in verbal form. We preserve it for the future in memory and communicate it to others through testimony. Semantics concerns the basic connection between the verbal form and its worldly content—what we know and apply in action. Without semantics, we cannot make sense of all that; the point of the enterprise becomes invisible. But the enterprise would be impossible without its psychological

underpinnings. The semantics and the psychology are not different modules at the same level; semantics operates at a more abstract level. Psychology no more makes semantics inert, idle, or redundant than anatomy makes evolutionary biology inert, idle, or redundant.

Lederman has another more specific concern about simplicity in assessing semantic theories, where it can be assigned a double role. First, in any discipline, theories' comparative simplicity figures in their abductive assessment. Semantics is no exception: overfitting is just as possible there as elsewhere. Lederman does not contest this point. But comparative simplicity may also be assigned a second role in disciplines such as semantics. For one may worry that a complicated semantic clause for an atomic expression of natural language assigns it a meaning too complicated and gerrymandered for normal human speakers to grasp or manipulate in the usual way. That is where Lederman pushes back. He writes: 'formal semantics does not directly seek to provide characterizations of how, cognitively, information is stored or processed' (I agree); instead, 'it aims to describe truth-conditions and entailment relations, without taking a stand on the question of how information about these facts is stored or processed' (again, I agree). Lederman concludes that 'we can (and should) mostly leave questions about processing to one side while doing semantics'.

Agreed, we should not expect a formal semantic clause for an atomic expression of natural language to be isomorphic to a lexical entry for that expression written in normal speakers' brains, or to lay out pathways for cognitive processing. But that is not the issue. Take artificial words with gerrymandered meanings, such as Goodman's 'grue' and 'bleen' and Kripke's 'quus'. Philosophers have learned what these words mean, but not in the usual way, by exposure to their use. Students have to be told their definitions, otherwise they would not understand the words. If we invented other words and gave them even more complicated and irreducibly unnatural definitions, they would be both virtually unusable and unlearnable in the usual way. In areas such as mathematics and the law, technical terms are introduced by complex stipulative definitions; although experts do learn to apply them accurately, they often have to do so reflectively, checking their application clause by clause; this expertise is hard to acquire. In practice, looking at a definition or semantic clause phrased in terms drawn from natural language often gives us good evidence of how hard or easy it would be to apply a word with that meaning in the fluent, unreflective, moderately accurate way characteristic of native speakers.

Of course, a complicated and intractable definition may turn out to be logically equivalent to a much simpler and more tractable one. But that is untypical of actually proposed definitions, for the obvious reason that the proposer of a definition usually wants it to catch on, and so is strongly motivated to put it in its simplest and most tractable equivalent form.

In this slightly less direct way, the complexity of a proposed semantic clause for an ordinary word of a natural language can, and sometimes does, provide evidence that the clause is incorrect. Thus, questions about processing are more relevant to the semantics of natural language than Lederman suggests. 'Know', 'think', and 'want' are amongst the most ordinary words of English; likewise for common attitude verbs in other natural languages. They are used in the fluent, unreflective, and—we hope—moderately accurate way characteristic of native speakers. That puts some limit on what kind of meaning they can have.

Lederman's contextualism

In the final section of his comments, Lederman argues that his and Goodman's contextualist account of attitude ascriptions is more like my anti-contextualist account than might appear, just slightly more complex, and a much better fit to our practice of attitude ascription. On their view, an attitude verb φ as used in a context *c* expresses the relation that an agent *x* has to a proposition *p* just when some mentalese sentence *s* salient in *c* expresses *p* and occurs in *x*'s φ-box. Here mentalese is *x*'s language of thought, and a φ-box is the analogue of a belief-box for φ in place of 'believe'. Lederman suggests that my guises play the same role as their mentalese sentences, except that the mentalese sentences figure in their contextualist semantics while my guises do not figure in my non-contextualist semantics.

The talk of φ-boxes is not to be taken too seriously. On the Goodman-Lederman semantics, for agents to have knowledge as well as belief, they will need to have a 'know'-box as well as a 'believe'-box, but the 'know'-box cannot be anything like a special storage area of the brain, otherwise a mistaken agent who overestimates their own knowledge might store some contextually salient false mentalese sentences there and thereby count as knowing the false propositions they express. Nor is an impersonal epistemic construction such as 'On the Babylonians' evidence, it was probable that Hesperus is Phosphorus' to be understood in terms of the presence in any mental box of a salient mentalese sentence expressing the proposition that Hesperus is Phosphorus, even though such constructions raise many of the same issues as attitude ascriptions. The talk of mentalese sentences is also not to be taken too seriously, otherwise the account would imply that creatures with no language of thought know no truths about the world, because they have no mentalese sentences to store; we should not expect semantic analysis to tell us whether thought without a language of thought is possible.

Once the psychologistic flourishes are eliminated, what remains of the contextualist account has roughly this form: an attitude verb φ as used in a context *c* expresses the relation that an agent *x* has to a proposition *p* just when *x* is φ-related to some representation *s* salient in *c* that expresses *p* (where 'φ-related' needs explaining). This resembles a guise-theoretic contextualist account.

Lederman plays down the cognitive problems generated by contextualism, without identifying them clearly. Here is a simple case. In context *c*, Linda tells Mona: 'John believes that I am rich'. Linda is trustworthy and trusted. As a result, Mona comes to know in effect that John believes the proposition that Linda is rich under some guise salient in *c*, but (we may assume) Mona does not know which guise it is—for all she knows, John knows Linda by sight but not by name. In a later context *c**, with Linda absent, Mona wants to impart her knowledge about John's beliefs to Nora. But if Mona says in *c** 'John believes that Linda is rich', what she says is true if and only if John believes the proposition that Linda is rich under some guise salient in *c**. But the guise under which John believes that Linda is rich, which was salient in *c*, may easily not be salient in *c**; then it would be false for Mona to say in *c** 'John believes that Linda is rich'. Of course, Mona could clarify what she was told (if she remembers the details)—though she may prefer not to reveal her source to Nora. Such

complications are not fatal to the contextualist account, but they seem quite artificial—indeed, like artefacts of an over-elaborate theory.

The case just discussed is an instance of a generic problem for contextualist hypotheses in semantics: they tend to undermine the preservation and communication of knowledge. This raises the theoretical cost of contextualist semantic hypotheses, even if the cost must sometimes be paid—as Lederman notes, contextualism about quantifier restriction and some other cases is widely accepted, including by me. Much remains to be understood about how communication works in such cases, and what constraints it imposes.

Perhaps a more telling semantic problem for contextualism about attitude ascriptions is that it over-generates readings, as I suggest in the book (pp. 163-4). Here is an example:

(1) Someone thinks that I am a physicist.

Considering (1) makes one guise for the proposition that I am a physicist more salient than any other: the sentence 'I am a physicist' itself (or its translation into mentalese, if you prefer). Thus, given a contextualist semantics, one would expect a reading on which only that guise is relevant, so (1) is true if and only if someone thinks the proposition that I am a physicist under the guise of the sentence 'I am a physicist'. But only I can think that proposition under that guise, by the semantics of 'I'. Thus, the contextualist semantics should predict a reading of (1) on which it is true only if *I* think that I am a physicist (under the guise of that very sentence); since I do not so think, (1) is false on that reading—even in a context where 'someone' ranges over people some of whom accept 'Timothy Williamson is a physicist' with reference to me. Such a reading strikes me as utterly unnatural, forced, and convoluted, not just implausible but non-literal. A good semantics for attitude ascriptions should not go anywhere near it. There is just no such contextual restriction on who thinks that I am a physicist. In such cases, contextualism about attitude ascriptions over-generates readings. By contrast, a simple quasi-homophonic intensional semantics for attitude ascriptions, with no mention of guises, has no such problem with (1).

Overfitting in semantics multiplies degrees of freedom, and thereby tends to over-generate readings. Contextualism about attitude ascriptions looks like a case in point. In that respect, it fits native-speaker judgments about cases worse than the corresponding anti-contextualist account. For a theory advocated primarily on grounds of its fit with the data, that is a serious blow.

Acknowledgments

Many thanks to both Kit Fine and Harvey Lederman for their searching and careful comments on the book, which have pushed me to clarify and develop some of its key arguments and conclusions. Thanks also to the audience at the original event at the 2026 APA Eastern Division meeting in Baltimore for helpful questions.

References

Boghossian, Paul. 2024: 'Reply to Williamson', in Blake Roeber, Ernest Sosa, Matthias Steup, and John Turri (eds.), *Contemporary Debates in Epistemology*, 3rd ed.: 194-198. Hoboken, NJ: Wiley.

Fine, Kit. 2026: 'Comments on Timothy Williamson's "Overfitting and Heuristics in Philosophy", THIS VOLUME.

Goodman, Jeremy, and Harvey Lederman. 2021: 'Perspectivism', *Noûs*, 55: 623-648.

Kripke, Saul. 1979: 'A puzzle about belief', in Avishai Margalit (ed.), *Meaning and Use*, 239-283. Dordrecht: Reidel.

Lederman, Harvey. 2026: 'Comments on *Overfitting and Heuristics*', THIS VOLUME.

Menzel, Christopher. 2024: 'Pure logic and higher-order metaphysics', in Peter Fritz and Nicholas Jones (eds.), *Higher-Order Metaphysics*, 421-459. Oxford: Oxford University Press.

Russell, Bertrand. 1918-19: 'The philosophy of logical atomism', *The Monist*, 28: 495-527, 29: 32-63, 190-222, 345-380. Reprinted as Russell 1956.

Russell, Bertrand. 1956: 'The philosophy of logical atomism', in his *Logic and Knowledge: Essays 1901-1950*, ed. Robert Marsh, 175-281. London: Allen and Unwin.

Williamson, Timothy. 2007: *The Philosophy of Philosophy*. Oxford: Wiley Blackwell. Expanded ed. 2021.

Williamson, Timothy. 2024a: *Overfitting and Heuristics in Philosophy*. New York: Oxford University Press.

Williamson, Timothy. 2024b: 'Menzel on pure logic and higher-order metaphysics', in Peter Fritz and Nicholas Jones (eds.), *Higher-Order Metaphysics*, 460-471. Oxford: Oxford University Press.

Williamson, Timothy. 2024c: 'Is the a priori/a posteriori distinction superficial?', in Blake Roeber, Ernest Sosa, Matthias Steup, and John Turri (eds.), *Contemporary Debates in Epistemology*, 3rd ed.: 175-183. Hoboken, NJ: Wiley. To be reprinted with a reply to Boghossian 2024 in Timothy Williamson, *Essays on the Philosophy of Logic*. New York: Oxford University Press, in preparation.

Williamson, Timothy. 2024d: 'Where did it come from? Where will it go?', in Arturs Logins and Jacques-Henri Vollet (eds.) *Putting Knowledge to Work: New Directions for Knowledge-First Epistemology*, 21-70. Oxford: Oxford University Press.

Williamson, Timothy. 2026: 'Evidence in disguise', in Eva Schmidt and Martin Grajner (eds.), *Epistemic Dilemmas and Epistemic Normativity*, 134-151. London: Routledge.