

This is an electronic version of a Paper published in *Proceedings of the Aristotelian Society*
Supplementary Volume lxxxiii (2009)

THE NORMATIVE ROLE OF LOGIC

HARTRY FIELD AND PETER MILNE

I—HARTRY FIELD

WHAT IS THE NORMATIVE ROLE OF LOGIC?

The paper tries to spell out a connection between deductive logic and rationality, against Harman's arguments that there is no such connection, and also against the thought that any such connection would preclude rational change in logic. One might not need to connect logic to rationality if one could view logic as the science of what preserves truth by a certain kind of necessity (or by necessity plus logical form); but the paper points out a serious obstacle to any such view.

What is the connection between (deductive) logic and rationality? Answers to this vary markedly.

At one extreme is the view that a law of deductive logic is a law of rational thought. Frege seems to have advocated this: 'Laws of logic ... are the most general laws, which prescribe universally the way in which one ought to think if one is to think at all' (Frege 1893, p. 12). The quotation may suggest that something is a law of logic *if and only if* it is a law of rational thought, but my interest is in the weaker claim that *a requirement on* being a law of deductive logic is being a law of rational thought. But even this seems problematic, if rational change of logic is possible: can it really be that in a debate over logic, the party who advocates the incorrect logic is automatically irrational, however compelling her case may be and however poor the currently available arguments on the opposite side? The connection between logic and rationality seems more subtle than this.

At the other extreme, Gil Harman has cited a large number of obstacles to there being a close connection between logic and rationality, and has argued that logic has no more of a connection to rationality than any other important discipline does (Harman 1986, ch. 2). In Harman's view, logic is a science on par with all others: its goal is to discover laws of a certain kind, viz., about what forms of argument must preserve truth. Rational people will try to have the right views about this, but they will try to have the right views about physics and sociology too, so there is no more essential tie be-

tween logic and rationality than between physics or sociology and rationality.

This view has the advantage of easily accommodating rational change in logic. We can have a rational change of logic whenever we have a rational change in our beliefs about what forms of argument must preserve truth, and there seems to be no more reason to doubt that these beliefs can rationally change than to doubt that our beliefs about physics can rationally change.

I will, however, be defending a view that connects logic to rationality. Part 1 is mostly concerned with overcoming Harman's obstacles to such a connection, but I will also address the question of how to make the connection loose enough to allow for rational debate about logic, and for rational change in logic resulting from such debate. (See Problem 4b below.)

Part 2 argues against Harman's alternative view: it gives (perhaps surprising) grounds for the conclusion that logic *can't* be the science of what forms of inference necessarily preserve truth—even if the necessity in question is restricted to *logical* necessity, or *necessity by virtue of logical form*. And that makes it hard to see what logic could possibly be, if not somehow connected to laws of rational thought.

I

Harman has raised a number of problems about the connection between logic and rational belief. I'm not sure how seriously he takes them all, but rather than try to discuss his views I will just discuss the problems he's raised. (My thinking about this has been influenced by MacFarlane (unpublished), though I think that the views I arrive at are mostly different from his.)

Here are the four main problems Harman raises:

1. Reasoning (change of view) doesn't follow the pattern of logical consequence. When one has beliefs A_1, \dots, A_n , and realizes that they together entail B , sometimes the best thing to do isn't to believe B but to drop one of the beliefs A_1, \dots, A_n .
2. We shouldn't clutter up our minds with irrelevancies, but we'd have to if whenever we believed A and recognized that B was a consequence of it we believed B .

3. It is sometimes rational to have beliefs even while knowing they are jointly inconsistent, if one doesn't know how the inconsistency should be avoided.
4. No one can recognize all the consequences of his or her beliefs. Because of this, it is absurd to demand that one's beliefs be closed under consequence. For similar reasons, one can't always recognize inconsistencies in one's beliefs, so even putting aside point 3 it is absurd to demand that one's beliefs be consistent.

(This fourth problem really splits into two rather different problems, as we'll see.) The third and (both aspects of) the fourth are the ones of most interest, I think, but it is important to discuss them all since there are interactions among them that lead to some further problems.

Problem 1: 'When one has beliefs A_1, \dots, A_n , and realizes that they together entail B , sometimes the best thing to do isn't to believe B but to drop one of the beliefs A_1, \dots, A_n .' This shows that the following is not a correct principle:

If one realizes that A_1, \dots, A_n together entail B , then if one believes A_1, \dots, A_n , one ought to believe B .

But the obvious solution is to give the 'ought' wider scope:

If one realizes that A_1, \dots, A_n together entail B , then one ought not to believe A_1, \dots, A_n without believing B .

This would give a strong connection between reasoning and logic, even if reasoning doesn't 'follow the pattern of logical consequence'.

Problem 4a (the 'computational aspect' of Problem 4): Another issue about principles like these is whether one should strengthen them by weakening the antecedent from 'If one realizes that A_1, \dots, A_n together entail B ' to just 'If A_1, \dots, A_n together entail B '.

There is a clear rationale for wanting the strengthened forms. John MacFarlane has remarked that if the only normative claims that logic imposes are from *known* implications, then 'the more ignorant we are of what follows logically from what, the freer we are to believe whatever we please—however logically incoherent it is. But this looks backward. We seek logical knowledge so that we know how we ought to revise our beliefs: not just how we *will* be obligated to revise them when we acquire this logical knowledge,

but how we are obligated to revise them even now, in our state of ignorance' (MacFarlane unpublished, p. 12).

On the other hand, there are obvious problems with the strengthened forms. Believing all the logical consequences of one's beliefs is simply not humanly possible, so failure to do so can hardly be declared irrational. For similar reasons, the idea that it is always irrational to be inconsistent seems absurd. Indeed, it is natural to suppose that *any* rational person would have believed it impossible to construct a continuous function mapping the unit interval onto the unit square, until Peano came up with a remarkable demonstration of how to do it. The belief that no such function could exist (in the context of certain set-theoretic background beliefs) was eminently rational, but inconsistent.

Is there a way between? For the interim, let's resolve this by a different alteration of the antecedent: let's take our principle to be

- (*) If A_1, \dots, A_n together *obviously* entail B , then one shouldn't believe A_1, \dots, A_n without believing B .

This may seem a bit of a cheat, and I'll come back to it.

Problem 3: 'It is sometimes rational to have beliefs even while knowing they are inconsistent, if one doesn't know how the inconsistency should be avoided.' A famous example is the Paradox of the Preface: one says in the preface that probably one has made an error somewhere in the book, even though this amounts to the disjunction of negations of claims in the book. More interesting examples involve well-known cases where physical theories (such as classical electrodynamics taken together with accepted background assumptions) lead to absurdities, but one doesn't know the best way to fix them and for each claim in them thinks it's probably right. This seems a rational attitude, and it is licensed by Bayesian views: one can have a high degree of belief in each of A_1 through A_n , but not in their conjunction or in some other claims entailed by their conjunction. (Take belief to be just degree of belief over some high contextually determined threshold.)

So while examples like this do create a problem for (*), it seems at first blush obvious how the problem should be fixed: replace 'if one believes A_1, \dots, A_n ' by 'if one believes $A_1 \wedge \dots \wedge A_n$ ':

- (w) If A_1, \dots, A_n together obviously entail B , then one shouldn't believe $A_1 \wedge \dots \wedge A_n$ without believing B .

Or slightly more generally:

- (w^+) If A_1, \dots, A_n together obviously entail B , then one's degree of belief in B should be at least as high as one's degree of belief in $A_1 \wedge \dots \wedge A_n$.

But even in the stronger form (w^+) this is excessively weak, for two reasons.

First, the force of \wedge -Introduction on degrees of belief is completely lost. \wedge -Introduction should be a substantive constraint on our degrees of belief: if one believes A_1 to degree \mathfrak{I} and A_2 to degree \mathfrak{I} , one ought to believe $A_1 \wedge A_2$ to degree \mathfrak{I} ; and if one believes A_1 to degree 0.95 and A_2 to degree 0.95 , one ought to believe $A_1 \wedge A_2$ to degree at least 0.9 . But (w^+) tells us only that the degree of belief in $A_1 \wedge A_2$ should be at least as high as itself!

The second problem is that people don't have degrees of belief for everything, so a principle governing a person's degrees of belief ought to be understood as having the tacit assumption that the person has all the degrees of belief in question. But so understood, (w) and (w^+) impose no constraint whatever on a person's degree of belief in B when she has high degrees of belief in A and in $A \rightarrow B$ but none in their conjunction.

We can handle both problems simultaneously as follows:

- (D) If A_1, \dots, A_n together obviously entail B , then one's degrees of belief in A_1, \dots, A_n and B (which I denote $P(A_1), \dots, P(A_n), P(B)$) should be related as follows: $P(B) \geq P(A_1) + \dots + P(A_n) - (n - \mathfrak{I})$.

The $n = \mathfrak{I}$ case just says that if A obviously entails B , one's degree of belief in B should be at least that of A . And the $n = 0$ case just says that if B is an obvious logical truth, $P(B)$ should be \mathfrak{I} . (D) seems the proper generalization. (For any n , (D) entails that if all the $P(A_i)$ are \mathfrak{I} , $P(B)$ should be too.)¹

(D) doesn't directly yield (w): for $P(A_1 \wedge A_2)$ may be higher than $P(A_1) + P(A_2) - \mathfrak{I}$, and if B is a consequence of A_1 and A_2 together, (D) yields only that $P(B) \geq P(A_1) + P(A_2) - \mathfrak{I}$ rather than the tighter

¹ We really ought to strengthen (D), to a principle about conditional degree of belief:

If A_1, \dots, A_n together obviously entail B , then for any C , it should be the case that $P(B|C) \geq P(A_1|C) + \dots + P(A_n|C) - (n - \mathfrak{I})$.

Similarly for other principles under discussion in this paper; I stick to unconditional belief only to simplify the discussion.

bound $P(B) \geq P(A_1 \wedge A_2)$. But this isn't a problem: if the logic includes conjunction-elimination, then when B is an obvious consequence of A_1 and A_2 together it is also an obvious consequence of $A_1 \wedge A_2$, so applying (D) to this we'll get the tighter bound $P(B) \geq P(A_1 \wedge A_2)$.

It should be noted that Principle (D) is quite neutral to the underlying logic (and thus to the full principles of Bayesianism, which require that the underlying logic be classical). Whatever logic is assumed correct, it seems to me that

- (i) if B is obviously entailed by A in that logic, a proponent of that logic should believe B to at least as high degree as A ;
- (ii) if B is obviously a theorem of the logic, it should be believed to degree 1;

and so forth. Some other features of degrees of belief in Bayesian theories fall out of Principle (D) *together with the assumption of classical logic*. Indeed, given any non-paraconsistent logic (that is, any logic in which contradictions entail everything), (D) yields that everything should be believed to degree at least $P(A) + P(\neg A) - 1$, whatever the A ; so if some things are legitimately believed to degree 0, this yields that $P(A) + P(\neg A)$ can never be more than 1. (In the context of paraconsistent logics, $P(A) + P(\neg A)$ is typically allowed to be more than 1. In the contexts of logics without excluded middle, it is allowed to be less than 1. (D) is general enough to apply to any of these logics.)

Actually a stronger generalization than (D) is available in standard Bayesian theories; rather than writing it out in full, I write it for the $n = 2$ case only:

- (D⁺) If A_1 and A_2 together obviously entail B , then it should be the case that $P(B)$ is at least $P(A_1) + P(A_2) - P(A_1 \vee A_2)$.

This is a tighter bound than the $P(A_1) + P(A_2) - 1$ that is delivered by (D). But this tighter bound is a special feature of Bayesian theories that doesn't hold in some generalizations of it such as the Dempster-Shafer theory (Shafer 1976): it isn't simply due to the impact of logical implication on our degrees of belief. Moreover, the point made above about people not having degrees of belief in every proposition applies here too: the $n = 2$ case of (D) has the tacit condition that the person has degrees of belief in A_1 , A_2 and B , but (D⁺) has the tacit

condition that the person has degrees of belief not only in these but also in $A_1 \vee A_2$; since one can have degrees of belief in the former without having them in the latter, (D) gives information in cases where (D⁺) doesn't.

Problem 2: 'Even though "The earth is round" entails "Either the earth is round or there are now Martian elephants in Times Square", it would be a bad thing to clutter up one's brain with such irrelevancies.'

The obvious solution, as Harman himself notes, is to distinguish explicit belief from implicit belief. Explicit beliefs are ones that are directly stored; one implicitly believes something when one is disposed to explicitly believe it should the question arise. So we should change (*) to something like

(***) If A_1, \dots, A_n together obviously entail B , then one shouldn't explicitly believe A_1, \dots, A_n without at least implicitly believing B .

But how does one fit this with a degree of belief model, used in (D)?

I've already mentioned the idea of generalizing standard Bayesian theories, so that an agent needn't have a degree of belief in every sentence of her language. An obvious addition to this is to make an explicit-implicit distinction among one's actual degrees of belief: an explicit degree of belief is a degree of belief that is represented explicitly in the agent, and an implicit degree of belief is a disposition to have that degree of belief explicitly.

However, the notion of implicit degree of belief is not general enough for our needs: Principle (D) cannot be suitably generalized using it alone. For instance, given that one's degree of belief in a coin's coming up heads is $\frac{1}{2}$, we needn't have even an implicit degree of belief in the claim that *either it will come up heads or there will be war in Iran next year*. What's true is only that we are implicitly precluded from having a degree of belief less than $\frac{1}{2}$ in this. We need an account of this 'implicit precluding'. I'll come back to this.

Problem 4a revisited: I've tried to avoid some of the problems of excessive demand by restricting to obvious entailments, but there may be a question as to whether I've brought back the problems by my talk of degrees of belief. Standard discussions of degrees of belief totally ignore computational limitations. A rather minimal computational limitation is Turing-computability; but no Bayesian probability function on a rich language is computable, at least if it satisfies

a very minimal condition of adequacy. For every Bayesian probability function must assign the value 1 to every logical truth. By Church's theorem on the undecidability of classical logic, this tells us that any computable probability function would have to assign value 1 to things other than logical truths as well as to the logical truths. One could live with that; however, it is easy to extend the proof of Church's theorem, to show that any computable function on an arithmetic language that assigns value 1 to all logical truths must also assign value 1 to something inconsistent with Robinson arithmetic (a very weak arithmetic theory). So any computable probability function would have to assign probability 0 to a very weak fragment of arithmetic! I suppose someone might take this as an argument for nominalism (arithmetic is false!), but I wouldn't recommend it. What I think it shows is that one shouldn't focus on probability functions.

I say this as a Bayesian, of sorts. But in my view the focus shouldn't in the end be on probability functions, but on certain *probabilistic constraints*: constraints such as that the degree of belief in some specific claim A is at least $\frac{1}{2}$; or that the conditional degree of belief in A given B is no greater than that of C given D ; or that (for specific A , B and C) the conditional degree of belief in $A \wedge B$ given C is equal to the product of the conditional degrees of belief in A given C and in B given C (conditional independence). It is constraints such as these that we explicitly represent. These constraints evolve, both by something akin to the Bayesian process of conditionalization and also by thinking. The process of thinking can impose new explicit constraints; for instance, a new theorem will henceforth be explicitly constrained to get value 1. Before, it may have been constrained by logic to get value 1, but only by a very unobvious proof. *Obvious* logical consequence will however impose *implicit* probabilistic constraints; for instance, given that one's degree of belief in a coin's coming up heads is $\frac{1}{2}$, we will have an implicit constraint not to believe to degree less than $\frac{1}{2}$ any disjunction that includes it. The right way to think of the explicit–implicit distinction in this context is as a distinction among *constraints on* degrees of belief, not among degrees of belief themselves.

Given this, the natural idea for handling all four problems together is to modify (D) (from the solution to Problem 3) in something like the following manner:

(D*) If it's obvious that A_1, \dots, A_n together entail B , then one ought to impose the constraint that $P(B)$ is to be at least $P(A_1) + \dots + P(A_n) - (n-1)$, in any circumstance where A_1, \dots, A_n and B are in question.

(For instance, if one's explicit constraints obviously entail lower bounds of at least p_1, \dots, p_n on A_1, \dots, A_n respectively, then one is to either weaken these constraints or impose a lower bound of at least $\sum p_i - (n-1)$ on B , should the question of B arise.)

These remarks on probabilistic constraints fall far short of a serious theory: it is a mere gesture toward one, and doesn't go much beyond common sense. But my goal wasn't to deliver a theory, but to say why I think there's no problem in supposing that logic imposes a rationality constraint on our degrees of belief. The story I've told avoids the excessive demands of logical closure. It also avoids excessive demands of logical consistency: the constraints may be probabilistically inconsistent; a proper story of the updating procedure should be such that when an inconsistency is discovered, adjustments are made to try to eliminate it.

And it avoids these problems without confining the normative requirements to cases where the logical relations are *known by the agent*: the requirements are there whenever the entailments are obvious, even if the agent doesn't know them.

Avoiding Obviousness? I've been deferring a worry: what counts as obvious? In my view, there is no general answer to this, it depends on both who is being assessed and who is doing the assessing; but this is not obviously a problem for using the notion in describing normative requirements, for normative requirements are relative in both these ways. (I'll say *a bit* more about this in connection with Problem 4b.)

However, there is an alternative to using the notion of obviousness: we could list specific rules that we count as obvious and insist that they impose obligations; this would give the veneer of objectivity, for better or worse. That is, we'd get

(D*_{alt}) If B follows from A_1, \dots, A_n by a single application of a rule on the list, then one ought to impose the constraint that $P(B)$ is to be at least $P(A_1) + \dots + P(A_n) - (n-1)$, in any circumstance where A_1, \dots, A_n and B are in question.

This imposes obligations only for simple inferences. Even if complicated inferences can be obtained by putting together simple inferences of the sort covered in the antecedent of (D^*_{alt}) , there is no obligation to have one's beliefs accord with the complex inference; there is only an obligation to take the first step, a potential obligation to take the step after that once one has fulfilled that obligation, a still more potential obligation to take the third step, and so forth. For long complicated proofs, we have *at most* a long chain of potential obligations; this is far short of an obligation to believe the conclusion if one believes the premisses.

And for typical proofs, we don't even have a long chain of potential obligations.² For there is a distinction to be made—probably one of degree rather than a sharp dichotomy—between two kinds of obvious inference. In some cases, like the inference from $A \wedge B$ to A , it's hard not to explicitly think of the conclusion when one thinks of the premiss. So in these cases, when one has an explicit (constraint on) degree of belief in the premiss *and also attends to it*, it's very likely that one will have an *explicit* constraint on one's degree of belief in the conclusion. In other cases, like the inference from $\forall xA(x)$ to $A(t)$ for specific A and t , the inference is totally obvious once the premiss *and conclusion* are before the mind; nonetheless *explicit* belief in the conclusion based on explicit belief in the premiss is atypical because one needs the specific t to be brought to one's attention.³ (There is no conflict here with (D^*_{alt}) , because of its last clause.) Famous proofs like Russell's disproof of naïve comprehension remained unobvious for so long, even though the derivation involved there is so quick, because of this. Perhaps there are hard-to-see proofs that use only premisses and rules of inference of the first sort. If so, then in those cases it is mere length of proof that makes the proofs unobvious. But that is certainly atypical of hard proofs. In the case of most hard proofs, then, there doesn't seem to be even the long chain of potential obligations contemplated in the previous paragraph. I think this fully handles the problem of computational limitations.

There is another aspect to Problem 4 that does not concern computational limitations. *Problem 4b* revolves around a question:

² Here I'm indebted to a discussion with Sinan Dogramaci and Ted Sider.

³ The point can't be avoided by taking *modus ponens* as the only rule of inference, for then the problem arises for the axioms: for instance, instances of the schema $\forall xA(x) \rightarrow A(t)$ are all obvious, but it can be hard to come up with the appropriate one to use in a proof.

should the facts of logical implication impose an obligation on those who don't accept the logic, especially those who have serious (even though not ultimately correct) reasons for not accepting it?

On what is probably the most natural interpretation of (D^*_{alt}) , the 'simple rules' it talks about are simple rules *of the correct logic*. In that case, (D^*_{alt}) answers 'yes'. But there is a case to be made that this consequence of (D^*_{alt}) is incorrect. Suppose that classical logic is in fact correct, but that Bob has made a very substantial case for weakening it—we may even suppose that no advocate of classical logic has yet to give an adequate answer to his case. Suppose that usually Bob reasons in accordance with the non-classical logic he advocates, but that occasionally he slips into classical reasoning that is not licensed by his own theory. Isn't it *when he slips and reasons classically* that he is violating rational norms? But (D^*_{alt}) (on the natural interpretation) says that it is on the *other* occasions, when he follows the logic he believes in, that he is violating norms. That's Problem 4b.

There are two obvious solutions. One is to switch to another interpretation of (D^*_{alt}) , on which the 'simple rules' are simple rules *of the logic the agent accepts*. (Or alternatively, *of the logic that the agent has most reason to accept*.) The effect of such agent-relativism is to remove the normative pull of reasoning in accord with *the correct* logic, when that logic is at odds with the logic that one accepts or that one has most reason to accept.

The above quote from MacFarlane suggests a discomfort about the relativist response. Paraphrasing a bit for the present context, we get 'This looks backward. We seek logical knowledge so that we know how we ought to revise our beliefs: not just how we *will* be obligated to revise them when we have the correct logical theory, but how we are obligated to revise them even now, in our state of logical error.'

MacFarlane himself advocated a very different solution: that there is an obligation to reason in accordance with the correct logic, but that there can also be competing obligations. In the case of those with serious reasons for doubting what is in fact the correct logic, these competing obligations are quite strong, so that there is simply no way to satisfy all of one's obligations until one corrects one's mistaken views about the logic. If that's right, we can have a basically Fregean view, in which there is a close tie between the correct logic and *one facet of* rationality, while allowing argument about logic to

be governed by another facet of rationality as well. However, the suggestion as it stands says little about that other facet of rationality.

My own view is that each of these responses to Problem 4b contains a considerable element of truth—especially the second response. This is, to a large extent, predicted by (D^*) , which is thus in some ways preferable to (D^*_{alt}) .⁴ On one reading of ‘obviously’ (obviously *to the agent*), (D^*) gives a response like the first, and on another (obviously *to someone with the correct logic*, or *to someone with our logic*) we would get a response like the second.

Really what’s at issue isn’t an *ambiguity* in ‘obvious’, it’s that ‘obvious’ is normative: an obvious entailment is one that an agent *ought* to see. Primarily we evaluate using our own norms (or if you like, using what we take to be the correct norms). But secondarily, we sometimes evaluate with respect to the agent’s norms. (D^*) thus yields both evaluations. (MacFarlane’s solution *allowed* for both evaluations, but only one was actually given by the main normative principles governing logic.)

Normativity Demoted? In this part of the paper I’ve been writing in ways that may seem to presuppose normative realism (‘What are the objective normative constraints that logic imposes?’). There is, however, an alternative, which I prefer: its core is that

- (1) The way to characterize what it is for a person *to employ* a logic is in terms of *norms the person follows*, norms that govern the person’s degrees of belief by directing that those degrees of belief accord with the rules licensed by that logic.

More specifically, we recast (D^*) or (D^*_{alt}) into something much less normative, roughly as follows:

- (E) Employing a logic L involves it being one’s practice that when simple inferences $A_1, \dots, A_n \vdash B$ licensed by the logic are brought to one’s attention, one will normally impose the constraint that $P(B)$ is to be at least $P(A_1) + \dots + P(A_n) - (n-1)$.

We get a certain kind of normativity derivatively, by the following obvious principle:

⁴ But (E) and (2) below preserve the advantages of (D^*) , in a format closer to (D^*_{alt}) .

- (2) In externally evaluating someone's beliefs and inferences, we go not just by what norms the person follows, but also by what norms *we take to be good ones*: we will use the logic *we* advocate in one facet of the evaluation, though we may use the agent's logic in another.

(2) doesn't connect up *actual oughts* with *the actually correct logic*, but connects *ought judgements* with *what we take to be good logic*. But my suggestion would be that there are no 'actual oughts' that this leaves out: normative language is to be construed expressivistically. So construed, a normative principle like (D*) will turn out to be correct, but will be seen as something like an epiphenomenon of (E) together with the evaluative practices alluded to in (2). And these evaluative practices allow the consideration of both our own logic and the other person's in evaluating the other person's beliefs; as I've said, this seems the best resolution of the issues under Problem 4b.

II

Harman's case against logic having a significant normative role rested largely on his dissatisfaction with all attempts to formulate that role. But it also rested in part on his view that there is an alternative role for logic: as, roughly, the science of what forms of argument necessarily preserve truth. I will now argue against any such alternative characterization. If right, this will substantially increase the plausibility of the idea that logic is to be characterized in part by its normative role.⁵ For if logic is not the science of what necessarily preserve truth, it is hard to see what the subject of logic could possibly be, if it isn't somehow connected to norms of thought.

My view is perhaps a surprising one: barring a small qualification, it is that we must reject the claim that all logically valid inferences preserve truth.

To motivate this, consider Gödel's second incompleteness theorem, which says that no remotely adequate mathematical theory can prove its own consistency (or even, its own non-triviality).⁶ This can seem puzzling: why can't we prove the consistency (or at least, non-

⁵ More fully, the idea (1) that *a person's* logic is to be characterized in part by the role that it plays in *that person's* norms for degrees of belief; and (2) that a *good* logic is to be characterized in part by the role that it plays in *good* norms for degrees of belief.

triviality) of a mathematical theory T within T by

- (A) inductively proving within T that T is sound, that is, that all its theorems are true,

and

- (B) arguing from the soundness of T to the claim that T is consistent (or at least, non-trivial)?

Except in the case of quite uninteresting theories of truth, the problem must lie in (A). But why can't we argue

- (Ai) that all the axioms are true;

- (Aii) that all the rules of inference preserve truth;

and conclude by induction that all the theorems are true? (Strictly, this can only work in a formalization of logic in which all reasoning is done at the level of sentences rather than sub-sentential formulas; but we can either pick such a formalization or else modify (Ai) and (Aii) by speaking of satisfaction rather than truth.)

In standard mathematical theories, the resolution of this is clear: Tarski showed that there can be no general truth predicate in classical logic that obeys the laws we'd expect, so standard mathematical theories do without a general truth predicate. In such theories one can't even formulate (Ai) and (Aii), let alone prove them. But then in such theories, we can't identify the valid inferences with the necessarily truth-preserving ones: that would require a general notion of truth. Tarski overcame this by identifying the valid inferences with those that preserve truth *in all classical models*. Truth in a classical model *is* definable, since it is very different from truth; but those very differences mean that this account of validity doesn't have the philosophical punch that necessary truth-preservation has. Even non-classical logicians will agree that classical inferences preserve truth in all classical models; but they will not agree that they preserve truth, for they think that classical models misrepresent reality. In fact, even classical logicians think that classical models misrepresent reality: classical models have domains restricted in size, where-

⁶ A theory is trivial if it can prove everything; this is equivalent to inconsistency (proving contradictions) for classical theories, but the point of the parenthetical remark in the text is to generalize to paraconsistent theories which allow for the acceptance of localized contradictions.

as set-theoretic reality doesn't. (That's the reason that truth-in-a-model can be defined when truth can't be.) So proving that the rules preserve truth-in- M for each M is not proving (Aii), and in standard approaches (Aii) is rejected as meaningless because of its unrestricted truth predicate. And if the 'true' in it were taken as a restricted truth predicate, (Ai) and (Aii) would be false.

There are, though, non-standard theories that do have a general truth predicate: both classical logic theories that give that predicate unusual laws, and non-classical theories that keep the usual laws of truth while weakening the logic. But it turns out that in every such theory of any interest, it is either inconsistent to suppose that all the axioms are true or inconsistent to suppose that all the rules preserve truth.

For instance, classical 'truth-value gap' theories contain specific axioms that the theory asserts while also asserting to be untrue. Typically, the axiom will be of form 'True($\langle A \rangle$) \rightarrow A '. Belief in the axiom is licensed—the axiom is taken to be valid, in the normative sense—but it is declared untrue! Axioms are degenerate cases of rules, so this is a degenerate case of a theory accepting a rule while declaring it invalid. Of course, one could simply define 'valid' to mean 'necessarily truth-preserving' (or in the case of axioms, 'necessarily true'). In that case the theory keeps the connection between validity and truth, but at the cost of taking its own axioms not to be valid. This move is uninteresting: the point is that the theory gives a special status to claims of the form 'True($\langle A \rangle$) \rightarrow A ', in that it takes them to be axioms; and this special status is not truth.

The above feature of gap theories is, I think, a gross defect in them: they seem somehow 'self-undermining'. Most other theories with a general truth predicate imply the truth of all their own axioms, but not the truth-preservingness of their own rules; indeed, such a theory will *reject* the claim that certain of the rules that it employs preserve truth, because adding such a claim would result in inconsistency.⁷ (Usually the rule that the theory employs but doesn't take to be generally truth-preserving is either *modus ponens* or a rule governing truth.) While this might at first seem just as counter-intuitive as rejecting the truth of some of one's axioms, I don't think this is so. The reason is that with most such theories, there is no rea-

⁷ Some of these theories don't accept that the rules *don't* preserve truth: they allow for rejection without acceptance of the negation.

son to doubt that all the rules preserve truth *when it matters*. That is, the rejection of the claim that a rule like *modus ponens* (or like the inference from $\text{True}(\langle A \rangle)$ to A) preserves truth generally arises because of a rejection of the claim that it preserves truth in certain degenerate instances (that is, with certain degenerate choices of premisses for the rule)—instances involving ‘ungrounded’ occurrences of predicates like ‘True’. But these are all instances in which, if the theory is consistent, the premisses of the rule could never be established or be rationally believable. In that case, the rejection of the claim that the rule preserves truth generally doesn’t seem to me to undermine the use of the rule.

So this doesn’t undermine the legitimate use of the rule, but it does show that the legitimate use of the rule is compatible with rejection of the claim that the rule generally preserves truth.

I’m inclined to state my conclusion by saying that the validity of a rule does not require that it generally preserve truth. However, some may think that this simply violates the meaning of the term ‘valid’: ‘valid’, they may say, simply means ‘necessarily preserves truth’, or ‘necessarily preserves truth in virtue of logical form’, or some such thing. I don’t think it does mean this—more on that in a moment—but I don’t want to fight about semantics: if one insists on using ‘valid’ to mean that, then my point is that every serious theory of truth employs rules whose ‘validity’ (in this sense) it rejects (or else can’t express). This seems initially surprising, but becomes less so when one reflects that the rule might still preserve truth ‘when it matters’.

Perhaps we should redefine validity, not as (necessarily) preserving truth in general but as (necessarily) doing so ‘when it matters’? Maybe, but this would require giving a clearer content to the quoted phrase than I know how to give. (Above, I basically said that a rule ‘preserves truth when it matters’ if it preserves truth *when applied to premisses that can be established or are rationally believable*. This seems too vague for a definition of validity.) I should note that even if the idea could be clarified, there would be no hope of *proving* that the inferences we employ are valid in this sense: our theory could prove that its rules ‘preserve truth when it matters’ only if it could prove its own consistency. (Indeed, this can be turned into an argument that such a claim would be not merely unprovable, but inconsistent with the theory, if one makes an assumption about the theory that most theories of this type meet: that it contains the inference rule $A \vdash \text{True}(\langle A \rangle)$.)

If validity isn't defined in terms of necessary truth-preservation (whether general or restricted to 'when it matters'), how is it to be understood? In my view, the best approach is to take it as a primitive notion that governs our inferential or epistemic practices. (That was the suggestion earlier in the paper: for instance, when we discover that the inference from A and B to C is valid, then we should ensure that our degree of belief in C is no lower than the sum of our degrees of belief in A and in B , minus 1.)

From this viewpoint, we can easily explain the naturalness of thinking that validity coincides with necessary truth-preservation. It is natural because the following argument is natural:

The validity of the inference from A_1, \dots, A_n to B

is equivalent to

the validity of the inference from $True(\langle A_1 \rangle), \dots, True(\langle A_n \rangle)$ to $True(\langle B \rangle)$,

by the usual truth rules. That in turn is equivalent to

the validity of the inference from $True(\langle A_1 \rangle)$ and ... and $True(\langle A_n \rangle)$ to $True(\langle B \rangle)$,

by the usual rules for conjunction. And that in turn is equivalent to

the validity of the sentence *If $True(\langle A_1 \rangle)$ and ... and $True(\langle A_n \rangle)$, then $True(\langle B \rangle)$,*

by the usual rules for the conditional. But validity of a sentence is necessary truth (by virtue of form), so this last is just the claim that the inference necessarily preserves truth (by virtue of form).

This argument looks very persuasive. However, it turns on principles that can't be jointly accepted! In particular, we can't subscribe both to the truth rules employed in the first step of the argument and to the rules for the conditional employed in the last step, on pain of triviality: that is the upshot of the Curry paradox. (See, for instance, Field 2008, §19.1.) There are different views on how the Curry paradox is to be resolved, but every one of them undermines one or another step in the argument that validity is to be identified with necessary truth-preservation.

As I've said, one could still stipulate that 'valid' is to mean 'necessarily preserves truth'. But this doesn't undermine the main point, which is that *that* notion of validity isn't what underwrites our no-

tion of goodness in deductive argument—validity in that sense isn't even extensionally equivalent to goodness of deductive argument. Our notion of good argument is an essentially normative notion, not capturable even extensionally in terms of truth-preservation. In this sense, logic is essentially normative.⁸

Department of Philosophy
New York University
5 Washington Place
New York, NY 10003
 USA
hf18@nyu.edu

REFERENCES

- Field, Hartry 2008: *Saving Truth from Paradox*. Oxford: Oxford University Press.
- Frege, Gottlob 1893: *Grundgesetze der Arithmetik*. Translated by Montgomery Furth as *Basic Laws of Arithmetic*. Berkeley, CA: University of California Press, 1967. Page references to this translation.
- Harman, Gilbert 1986: *Change in View*. Cambridge, MA: MIT Press.
- MacFarlane, John (unpublished): 'In what sense (if any) is logic normative for thought'. Delivered at the American Philosophical Association Central Division meeting, 2004.
- Shafer, Glen 1976: *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press.

⁸ Thanks to Gail Leckie, Jim Pryor, Josh Schechter and Tim Williamson for comments on an earlier version.